CODEN: JASMAN

The Journal of the Acoustical Society of America

ISSN: 0001-4966

Vol. 127, No. 1 January 2010

(Continued)

CONTENTS—*Continued from preceding page*

(Continued)

CONTENTS—*Continued from preceding page*

(Continued)

CONTENTS—*Continued from preceding page*

(Continued)

CONTENTS—*Continued from preceding page*

Voice fundamental frequency modulates vocal response to pitch perturbations during English speech

Hanjun Liu
Department of Rehabilitation Medicine, The First Affiliated Hospital, Sun Yat-sen University, Guangzhou 510080, People's Republic of China
lhanjun@mail.sysu.edu.cn

James Auger and Charles R. Larson
Department of Communication Sciences and Disorders, Northwestern University, 2240 Campus Drive, Evanston, Illinois 60208
jamesauger2007@u.northwestern.edu, clarson@northwestern.edu

Abstract: Previous research has demonstrated task-dependent vocal responses to pitch perturbations during speech production. The present study investigated the effect of voice fundamental frequency (F_0) on the modulation of vocal responses during English speech. Randomized pitch shifts of ±100 or 200 cents during speaking were presented to English speakers. Results indicated larger vocal responses and shorter latencies at a high voice F_0 than at a low voice F_0, but no significance differences were observed for stimulus magnitude or direction. These findings suggest that the pitch-shift reflex during speech can be modulated as a function of voice F_0.

PACS numbers: 43.70.Aj, 43.70.Bk, 43.70.Gr, 43.70.Jt [AL]
Date Received: March 10, 2009 **Date Accepted:** October 15, 2009

1. Introduction

Control of voice fundamental frequency (F_0) plays an important role in vocal communication. Through the regulation of voice F_0, humans can convey general information about the affective state, speech inflection, artistic purpose, etc. Auditory feedback has been demonstrated to be important for the on-line control of voice F_0 during sustained vowels (Hain *et al.*, 2000), in which subjects compensated for the pitch feedback perturbation by changing their voice F_0 in the direction opposite to the stimulus. Such compensatory mechanisms were also observed during speech production (Chen *et al.*, 2007; Liu *et al.*, 2009). These findings indicate that the audio-vocal system can be modulated in a task-dependent manner to correct the discrepancy between auditory feedback and vocal output.

During most of the previous pitch-shift studies, subjects were asked to vocalize a vowel sound or speak nonsense or meaningful syllables at their habitual pitch (Xu *et al.*, 2004; Chen *et al.*, 2007). A recent pitch-shift study during vowel phonation, however, demonstrated larger response magnitudes and shorter latencies when subjects vocalized at a high F_0 compared to a low F_0 (Liu and Larson, 2007). This study suggested that the sensitivity of the audio-vocal system to voice feedback perturbation might vary as a function of voice F_0 during sustained vowels. However, no research has been conducted on whether the same is true for speech production. Hypothetically, a reflexive input to motor neurons that are discharging at a high rate may lead to a greater level of muscle contraction than an equal input to neurons discharging at a lower rate. The greater degree of muscle contraction associated with vocalizing at a higher F_0 (Hirano *et al.*, 1970) may cause an increase in magnitudes and a decrease in latencies of voice F_0 responses to voice pitch-shifted feedback.

Therefore, the purpose of the present study was to investigate if vocal responses to pitch perturbation in auditory feedback can be modulated as a function of voice F_0 during English speech. We hypothesized that, similar to the sustained vowels, the response magnitudes

to pitch-shifted voice feedback would be larger and response latencies would be shorter at a high F_0 compared to a low F_0 during speech production.

2. Methods

2.1 Subjects

Fifteen female Northwestern University students (18–22 years old) participated in the experiment and produced data that met our criteria of acceptable responses. All of the subjects passed a hearing screening for a 25 dB hearing level bilaterally at 250, 500, 1000, 2000, and 4000 Hz. None of the subjects reported a history of neurological or communication disorders and they all signed informed consent approved by the Northwestern University Institutional Review Board.

2.2 Apparatus

During the testing, subjects were seated in a sound-treated room and wore Sennheiser headphones with attached microphone (model HMD 280). The vocal signal from the microphone was amplified with a Mackie mixer (model 1202) and shifted in pitch with an Eventide Eclipse Harmonizer, and then amplified with a Crown D75 amplifier and HP 350 dB attenuators at 80 dB sound pressure level (SPL). MIDI software (MAX/MSP V.4.6 by Cycling 74) was used to control the harmonizer. A Brüel and Kjar sound level meter (model 2250) and in-ear microphone and headphones were used for calibration to make sure that there was a gain in amplitude of 10 dB SPL between the subject's voice amplitude and the feedback loudness. The voice output, feedback, and transistor transistor logic (TTL) control pulses were digitized at 10 kHz, low-pass filtered at 5 kHz, and recorded using CHART software (AD Instruments, Castle Hill, New South Wales, Australia).

2.3 Procedure

Prior to the experiment, a phrase "You know Nina?" was spoken by one experimenter at a high and a low voice F_0 and recorded. Then these two voice samples were processed in PRAAT (Boersma, 2001) so that the average F_0 values during the voice period before the rise of the final syllable (i.e., "You know Ni") were maintained at around 300 and 200 Hz, which were regarded as reasonably high and low voice F_0 for female subjects according to pilot testing results. Those subjects who were unable to produce the phrase at either of these two voice F_0s were excluded. Subjects were then instructed that they would hear the phrase sample at either a high or a low voice F_0, and then they should repeat the phrase within 1 s in exactly the same manner as that of the sample. After a brief period of training, subjects were asked to repeat the phrase immediately following its presentation over the headphones 60 times in each of four conditions: high voice F_0, low voice F_0, and 100 or 200 cent pitch-shift magnitude. For each vocalization, the pitch feedback was increased, decreased, or held constant (no stimulus) in a randomized sequence for a total of 60 trials. The duration of each stimulus was 200 ms and the magnitude was held constant at ±100 or ±200 cents. The pitch-shift stimulus was presented 200 ms following onset of vocalization. Data were analyzed using event-related averaging techniques (Chen *et al.*, 2007) in Igor Pro (Wavemetrics, Inc., Lake Oswego, OR). A statistical method was used to determine if responses to pitch-shifted feedback differed significantly from control trials (see Xu *et al.*, 2004). SPSS (v. 16.0) was used to test for significant differences in the response magnitude and latency across all the conditions.

3. Results

The period between vocal onset and stimulus onset was extracted to measure the mean F_0 values for each condition. Statistical results showed that on average, subjects spoke at 275 Hz for the high voice F_0 condition and 209 Hz for the low voice F_0 condition [$F(1,58)=35.437$, $p<0.0001$]. From 15 subjects across two voice F_0s, two stimulus magnitudes, and two stimulus directions, there were 120 responses ($15\times2\times2\times2$). 93 responses opposed the stimulus direction, and 22 responses followed the direction of stimulus. 5 of 120 responses did not meet our criteria of validity and were declared to be non-responses. A chi-square test revealed a statistically greater

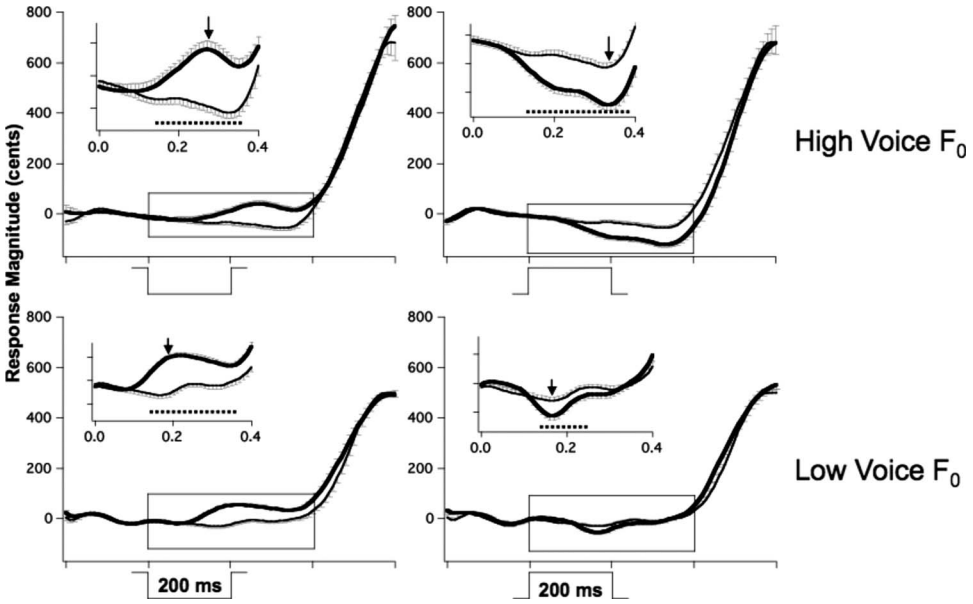


Fig. 1. Representative vocal responses to 200 cent pitch-shifted stimuli as a function of stimulus direction at a high (top) and a low (bottom) voice F_0, respectively. Thick lines represent the averaged F_0 contours of responses to pitch-shifted feedback, and thin lines represent contours for control trials. The solid vertical arrow indicates time where the response magnitude was measured. The dashed horizontal line represents the onset and offset of the response, and response latency is indicated by the start of this line. The inset shows an expanded portion of average waves. Error bars attached to the contours represent the standard error of the mean for a single direction. Boxes at the bottom indicate the time and the direction of the stimulus.

number of non-response and "following" responses in the low voice F_0 compared to the high voice F_0 condition ($\chi^2=6.259$, d.f.$=1$, and $p=0.012$). The distribution of opposing, following, and non-responses was even across stimulus direction and stimulus magnitude.

Figure 1 shows the representative vocal responses to pitch perturbations as a function of stimulus direction at a high and a low voice F_0. As shown in this figure, larger vocal responses occurred when subjects spoke the phrase at a high voice F_0 compared to a low voice F_0. Table 1 presents the average and standard deviations (SDs) of the response magnitudes and latencies across all conditions. Although these data lend themselves to repeated-measures ANOVAs (analysis of variance), factorial ANOVAs without repeated-measures were used in the present study to account for missing data and unequal cell size. A three-way ANOVA performed on the response magnitude indicated a significant main effect for voice F_0 [$F(1,81)=11.817$, $p=0.001$] but not stimulus magnitude [$F(1,81)=0.916$, $p=0.341$] or stimulus direc-

Table 1. Averaged response magnitude SD in cents and response latency SD in ms as a function of voice F_0, stimulus magnitude, and stimulus direction.

		Response magnitude		Response latency	
		High voice F_0	Low voice F_0	High voice F_0	Low voice F_0
100 cent stimuli	Up	27 (18)	19 (13)	138 (61)	165 (77)
	Down	37 (24)	28 (18)	117 (48)	152 (66)
200 cent stimuli	Up	40 (25)	22 (11)	106 (38)	124 (67)
	Down	48 (33)	23 (22)	115 (58)	140 (63)

tion [$F(1,81)=2.062$, $p=0.155$]. The high voice F_0 produced significantly larger response magnitudes (38 ± 26 cents) than the low voice F_0 (23 ± 16 cents). No significant interactions were found across all conditions.

A three-way factorial ANOVA on the response latency revealed a significant main effect for voice F_0 [$F(1,81)=4.251$, $p=0.042$] but not for stimulus magnitude [$F(1,81)=2.967$, $p=0.89$] or stimulus direction [$F(1,81)=0.029$, $p=0.866$]. Faster response latencies were generated at the high voice F_0 (119 ± 52 ms) than the low voice F_0 (146 ± 68 ms). No significant interactions were found in the response latency across all conditions.

4. Summary and discussion

Liu and Larson (2007) reported that voice F_0 has effects on the vocal responses to pitch feedback perturbations during sustained vowels. The present study showed that vocal responses during speech production were also modulated as a function of voice F_0. As we hypothesized, the higher voice F_0 led to larger response magnitudes and shorter latencies. The finding that larger response magnitudes occurred at the high voice F_0 compared to the low voice F_0 once again demonstrates that vocal response can be modulated in a task-dependent manner, which is consistent with previous research during speech production (Xu *et al.*, 2004; Liu *et al.*, 2009). One possible explanation for the greater magnitudes with higher F_0 is that speaking a sentence at a high voice F_0 may require a greater reliance on auditory feedback than at a low voice F_0. Thus, auditory feedback is closely monitored and pitch-shifted errors can be corrected with a greater degree of accuracy at a high voice F_0 than at a low voice F_0.

It was also found that response latencies during speech production varied as a function of voice F_0, which is consistent with Liu and Larson's (2007) study. Furthermore, this study provides supportive evidence that the timing of vocal responses to the pitch-shifted feedback can be adjusted according to the variations in speech contexts (i.e., voice F_0). Previous research has demonstrated that increases in voice F_0 are accompanied by increases in the activity of the tensor muscles of the larynx as reflected by the greater magnitudes of electromyographic signals (Hirano *et al.*, 1970; Gay *et al.*, 1972). The greater laryngeal motor responsiveness may also lead to the reduction in vocal response latencies when subjects were vocalizing at a high voice F_0.

The paradigm of the present study was similar to Chen *et al.* (2007), and the only difference is that we instructed the subjects to say "You know Nina?" at either a high or a low voice F_0 while Chen *et al.* (2007) just used a habitual pitch. Nevertheless, the two studies are comparable in that mean response magnitudes were 38 and 23 cents for the high and low voice F_0 in the present study and 31.5 cents for Chen *et al.*'s (2007) study. On the other hand, Chen *et al.* (2007) reported greater response magnitudes for downward stimulus direction than upward direction, which was not observed in the present study. As suggested by Chen *et al.* (2007), the directional effect was observed because downward stimuli may sound to subjects that their voice F_0 was changing in the wrong direction during the production of the phrase with a rise of F_0 on the final syllable. In contrast, the manipulation of voice F_0 in the present study required precise control of laryngeal muscles and may have weakened this interaction between the suprasegmental production and stimulus direction, leading to the absence of the directional effect on the response.

Another difference between the present and the Chen *et al.* (2007) study was that more following responses were found in the present study compared to the Chen *et al.* (2007) study (20% vs 7.5%). It should be noted that 67% of following responses occurred in the low pitch condition and 33% in the high pitch condition, which is consistent with Liu and Larson's (2007) finding during vowel phonation at a low and high F_0. One possible reason for this difference is that an increase in voice F_0 resulted in greater accuracy of detecting the correct direction of the pitch-shift stimulus and a greater percentage of compensating responses during the high F_0 condition in the present study. In a previous study, it was suggested that accuracy in perception of the direction of a pitch-shift stimulus may be a factor contributing to the direction of voice F_0 responses (Larson *et al.*, 2007). In addition, it was reported that some subjects were better than

others at perceiving the direction of the pitch changes (Semal and Demany, 2006). So individual differences might also contribute to the greater number of following responses and less sensitivity to the stimulus direction in the present study compared to Chen *et al.* (2007).

Acknowledgments

This work was supported by NIH Grant No. 1R01DC006243. The authors thank Chun Liang Chan for programming assistance.

References and links

Boersma, P. (**2001**). "Praat, a system for doing phonetics by computer," Glot International, **5**, 341–345.
Chen, S. H., Liu, H., Xu, Y., and Larson, C. R. (**2007**). "Voice F0 responses to pitch-shifted voice feedback during English speech," J. Acoust. Soc. Am. **121**, 1157–1163.
Gay, T., Hirose, H., Strome, M., and Sawashima, M. (**1972**). "Electromyography of the intrinsic laryngeal muscles during phonation," Ann. Otol. Rhinol. Laryngol. **81**, 401–409.
Hain, T. C., Burnett, T. A., Kiran, S., Larson, C. R., Singh, S., and Kenney, M. K. (**2000**). "Instructing subjects to make a voluntary response reveals the presence of two components to the audio-vocal reflex," Exp. Brain Res. **130**, 133–141.
Hirano, M., Vennard, W., and Ohala, J. (**1970**). "Regulation of register, pitch and intensity of voice. An electromyographic investigation of intrinsic laryngeal muscles," Folia Phoniatr Logop **22**, 1–20.
Larson, C. R., Sun, J., and Hain, T. C. (**2007**). "Effects of simultaneous perturbations of voice pitch and loudness feedback on voice F0 and amplitude control," J. Acoust. Soc. Am. **121**, 2862–2872.
Liu, H., and Larson, C. R. (**2007**). "Effects of perturbation magnitude and voice F0 level on the pitch-shift reflex," J. Acoust. Soc. Am. **122**, 3671–3677.
Liu, H., Xu, Y., and Larson, C. R. (**2009**). "Attenuation of vocal responses to pitch perturbations during Mandarin speech," J. Acoust. Soc. Am. **125**, 2299–2306.
Semal, C., and Demany, L. (**2006**). "Individual differences in the sensitivity to pitch direction," J. Acoust. Soc. Am. **120**, 3907–3915.
Xu, Y., Larson, C., Bauer, J., and Hain, T. (**2004**). "Compensation for pitch-shifted auditory feedback during the production of Mandarin tone sequences," J. Acoust. Soc. Am. **116**, 1168–1178.

A method for finding constrictions in high front vowels

Michel T.-T. Jackson and Richard S. McGowan
CReSS LLC, 1 Seaborn Place, Lexington, Massachusetts 02420
ladmttj@ix.netcom.com

Abstract: The purpose of this study was to devise a consistent and robust method for defining vocal tract constrictions in high front vowels. A procedure was devised to find the length and position of the articulatory constriction in high front vowels that is not sensitive to local fluctuations in vocal tract shape and to the constriction-defining parameters. A method based on a visual examination of plots for constriction length and position as functions of the constriction-defining parameters was found to provide stable constriction definitions.

PACS numbers: 43.70.Bk, 43.70.Kv [AL]
Date Received: July 3, 2009 **Date Accepted:** October 21, 2009

1. Introduction

Constriction positions and lengths are important parameters in determining the acoustic output of the vocal tract for vowels and other speech sounds (Stevens, 1998; Fant, 1960). However, not many methods for automatically determining these constriction parameters from x-ray or other data have been published. As Iskarous (2005) observed, "imaging modalities such as X-ray cinefluorography, MRI, and ultrasound produce images … but these images are difficult to quantify and compare." Even determining the direction of tongue movement in such images (e.g., Wood, 1982; Iskarous, 2005) does not determine the actual position and size of the resulting constriction, since a small tongue blade movement toward a nearby portion of the hard palate may produce an acoustically relevant constriction, whereas even a large tongue body movement toward a relatively distant portion of the dorsal wall of the pharynx may not. This study presents a method for determining constriction parameters in high front vowels when midsagittal images are available from any of the afore-mentioned measurement techniques. The method is robust in three ways: its results are not sensitive to local speaker-specific variation in vocal tract shape, its results are not sensitive to the selection of specific values of the constriction-finding method's own parameters, and it is applicable across a number of speakers of different languages.

Figure 1 illustrates a common kind of local speaker-specific variation in vocal tract shape, from a cinéradiographic tracing of the vowel /i/ as produced by the speaker described in Perkell, 1969. The midsagittal cross-dimensions in this token are measured along gridlines constructed according to the procedure described in Jackson and McGowan, 2008. These midsagittal cross-dimensions, plotted as a function of the gridline number, are shown in Fig. 2. In this vowel token, the small undulation in the shape of the palate creates a local minimum [indicated by the arrows in Figs. 1 and 2(a)] in the vocal tract's midsagittal cross-dimension function. Yet neither the acoustically relevant vocal tract constriction nor articulatory goal of the constriction-making gesture is likely to be limited to this local minimum alone.

It is possible to eliminate the effects of this kind of local minimum in at least two ways. One is to take the criterion for being "within the constriction" as being some multiple k, with $k > 1$, of the minimum midsagittal cross-dimension. Another way to eliminate the effects of a local minimum is to construct a criterion for being within the constriction based on the mean midsagittal cross-dimension d within some window around the location of the minimum. This reduces the effects of single gridlines having small cross-dimensions.

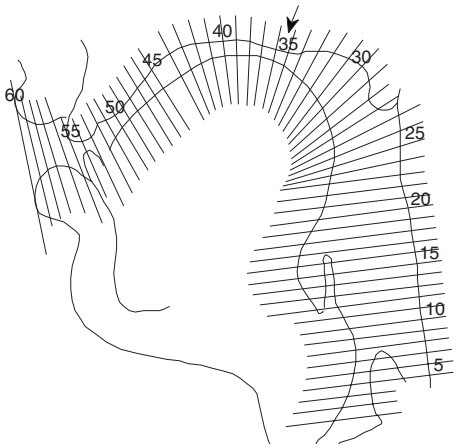


Fig. 1. Measurement gridlines for the speaker described in Perkell, 1969, shown on a token of /i/. The arrow near gridline 35 shows a local minimum in the vocal tract midsagittal cross-dimensions measured along the gridlines.

This study investigates the effect determining a criterion cross-dimension $c=k\cdot d(n)$ while varying both k and the number n of gridlines posterior to the location of the minimum cross-dimension, in order to determine a combination of k and n that is both useful in determining constriction length and position in high front vowels, and is not unduly sensitive to individual cross-dimension measurements on particular gridlines. This method will be applied to articulatory data from a variety of languages.

2. Procedure

2.1 Materials

Various images of the vocal tract during high, front vowel production from 13 different speakers were analyzed in this study. The vowels were from North American English, French, Spanish, and Mandarin Chinese. The digital image of each vowel token was enhanced and aligned using the MATLAB Image Processing Toolkit. Hard structures in and near the vocal tract, especially teeth, fillings, and the hard palate, were used as landmarks for rotation and translation of the images in the alignment procedure as described in Jackson and McGowan, 2008. The images analyzed as part of a larger study are summarized in Table 1.

In North American English, tracings of cinéradiographic images of three speakers' productions of the vowels /i, ɪ/ were analyzed in this study. The three speakers included the speaker from Perkell (1969), and two speakers from the ATR X-ray Film Database (Munhall *et al.*, 1994) originally recorded by Rochette (1973, 1977).

Original tracings from Perkell (1969), provided courtesy of the author, were scanned and digitized at 300 dpi. These tracings were based on cinéradiography at 45 frames/s. The cinéradiographic recordings of running speech in the ATR X-ray Film Database (Munhall *et al.* 1994) have been digitized and are available on DVD. The original films were shot at 50 frames/s. Single frames from multiple tokens of the vowels /i, ɪ/ produced by the speakers L73/74 and L78/79 were selected and analyzed. Each frame was taken at or near the extreme of mandibular motion or a period of negligible tongue movement during the vowel.

In French, tracings of cinéradiographic images from running speech of three speakers' productions of vowels from Bothorel *et al.*, 1986 were analyzed. The original x-rays were filmed at 50 frames/s. The tracings were digitally scanned at 300 dpi, and the vowels /i, y/ were used in this study.

In Spanish, tracings of x-ray images of three speakers' static productions of Spanish

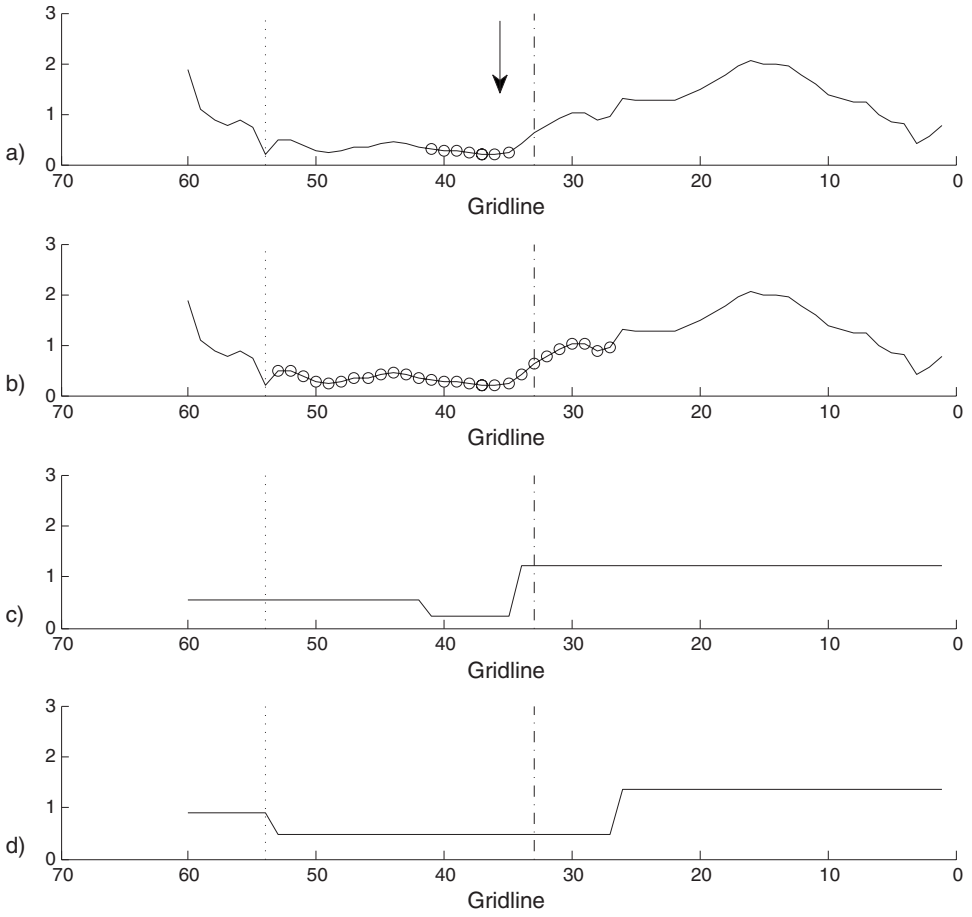


Fig. 2. The midsagittal cross-dimension function for the vowel token shown in Fig. 1, together with typical results of the parametrized constriction-finding procedure and derived three-tube models. The dashed line at gridline 33 represents the posterior boundary of the hard palate; the dotted line at gridline 53 represents the anterior boundary. Gridlines marked with "o" are in the constriction found by the procedure. The arrow near gridline 35 shows the same local minimum as Fig. 1. In (a), the o's show the constriction found using $k=1.5$ and $n=2$. In (b), the o's show the constriction for $k=4.0$ and $n=8$. (c) shows the three-tube model derived from (a); (d) shows the model from (b).

vowels from Navarro-Tomás, 1916; Parmenter and Treviño, 1932; Russell, 1929 were analyzed. Each tracing was scanned at 300 dpi, in some cases after photocopying with enlargement. The vowel /i/ was used in this study.

In Chinese, tracings of x-ray images of four speakers' productions of Chinese vowels from Ohnesorg and Svarný, 1955; Abramson *et al.*, 1962 were analyzed. The speakers (speakers A and B from Ohnesorg and Svarný, 1955; and speakers 1 and 4 from Abramson *et al.*, 1962), spoke either Mandarin or a closely related dialect.

The tracings of x-ray images of static productions of vowels from Ohnesorg and Svarný, 1955 were scanned at 300 dpi after photocopying with enlargement. The vowels /i, y/ were used in this study. The original 16 mm film described in Abramson *et al.*, 1962, which was shot at three times the normal rate for 16 mm film (i.e., 3×24 frames/s $=72$ frames/s) was transferred to Betamax- and VHS-format videotape in earlier work. The VHS-format videotape was then digitized and transferred to DVD. Single frames from running speech taken near the articulatory peak of the vowels /i, y/ in the digitized video were selected and analyzed in this study.

Table 1. Summary of languages, speakers, vowels, and number of tokens used to construct measurement gridlines in this study.

Language	Speaker and source	Number of tokens													
		i	ɪ	e	ɛ	æ	y	ø	œ	ʌ/ə	ɑ	ɔ	o	ʊ	u
English	Perkell	1	1	—	3	1	—	—	—	—	1	—	—	1	1
	L73/74	13	12	—	8	7	—	—	—	—	8	—	—	3	8
	L78/79	8	6	—	7	8	—	—	—	3	6	—	—	1	9
French	S1	2	—	2	2	—	2	2	1	—	2	2	2	—	2
	S2	2	—	2	2	—	2	2	1	—	2	2	2	—	2
	S3	2	—	3	2	—	2	2	1	—	2	2	2	—	2
Spanish	NT	1	—	1	1	—	—	—	—	—	1	1	1	—	1
	PT	1	—	1	1	—	—	—	—	2	1	1	—	1	1
	R	1	—	1	—	—	—	—	—	—	1	—	1	—	1
Chinese	OSA	2	—	—	—	—	1	—	—	1	3	—	—	—	1
	OSB	1	—	—	—	—	1	—	—	2	—	—	2	—	1
	A1	4	—	—	3	—	3	—	—	3	5	—	2	—	5
	A4	2	—	—	2	—	1	—	—	1	1	—	1	—	2

2.2 Gridlines

The measurements used in this study were taken along gridlines based on the method described in Jackson and McGowan, 2008, supplemented where necessary by gridlines spaced evenly from the alveolar ridge to the line tangent to the anterior surfaces of the upper and lower lips in the midsagittal plane. As in Jackson and McGowan, 2008, all the tokens produced by each speaker were traced to ensure that all of each speaker's gridlines were above the highest position of the glottis. In order to make the measurements comparable from speaker to speaker, 53 gridlines were placed between the upper incisors and the highest position of the glottis.

2.3 Parametrized constriction-finding procedure

Since the overall goal of this study was to analyze high front vowels, which typically have a constriction in the palatal region of the vocal tract, the initial step was to manually define a region of interest for each speaker. The region of interest was taken as all the gridlines intersecting the hard palate. In Fig. 1, the posterior boundary of the hard palate is approximately gridline 33, marked with a dashed line in Fig. 2. The anterior of the hard palate is gridline 53, marked with a dotted line in Fig. 2. The gridline with the minimum midsagittal cross-dimension in the region of interest was then identified. The mean cross-dimension $d(n)$ for the minimum plus the cross-dimensions from the n gridlines posterior to that gridline was calculated. Gridlines posterior to the minimum were used in order to avoid including the incisors and lip constriction in the window. Thus $n=0$ is the minimum cross-dimension alone, $n=1$ is the window with the minimum cross-dimension plus the next gridline posterior to the minimum. The number of gridlines in the window was $(n+1)$. The criterion distance $c=k\cdot d(n)$ was calculated and the constriction was then defined as the set of all contiguous gridlines around the minimum gridline with cross-dimensions less than the criterion distance. The multiplier k was varied in steps of 0.25 from 1.25 to 8.0 and n was varied from 0 to 10.

2.4 Tube models

In order to provide some insight into the acoustic plausibility of the constriction found by the procedure, three-tube models based on the constriction found for each value of k and n were constructed. The posterior, constriction, and anterior tube diameters were taken as the means of

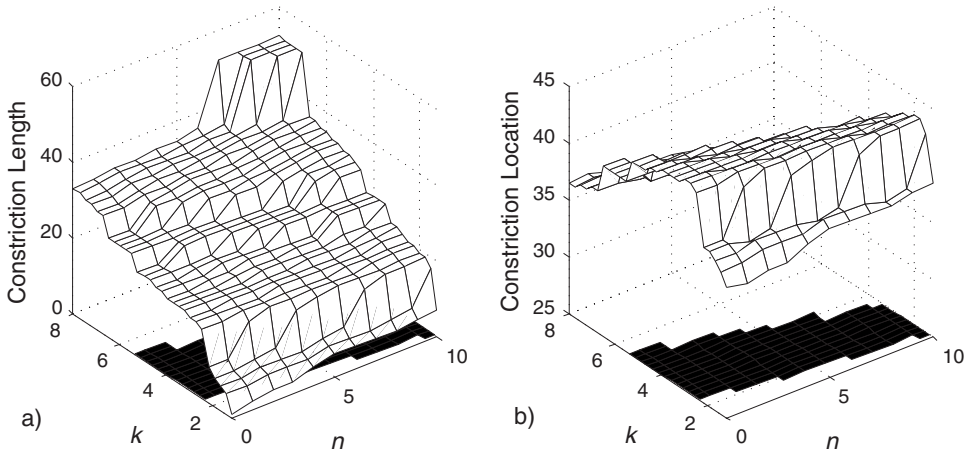


Fig. 3. The results of varying k and n over the ranges 1.25–8.0 and 0–10, respectively. (a) shows the length of the constriction (in gridlines) determined by the constriction-finding procedure. (b) shows the location of the center of the constriction determined by the procedure. The black shadow under each plot shows the range of (k,n) values that yield stable constriction length or location estimates.

the cross-dimensions of all the gridlines posterior to, within, and anterior to the constriction, respectively. The cross-dimensions were converted to areas using the function given in Wood, 1982 (p. 35) $A=1.93\ (d^{1.5})$, where A is the estimated cross-sectional area and d is the mean midsagittal cross-dimension of the relevant section of the vocal tract. For comparability with other work, the cross-sectional area of the constriction was normalized to 0.3 cm^2 (see Stevens, 1998) (p. 277 ff. and Fig. 6.1) and the areas of the other sections were scaled up or down according to the same factor. The overall length of the tube model was set to 16 cm. The area function of each tube model was then used in the EASY speech synthesizer (McGowan and Wilhelms-Tricarico, 2005) in order to determine its transfer function and the corresponding formant frequencies.

3. Results

Figure 2 shows typical results of the parameterized constriction-finding procedure. Figure 2(a) shows the results for small values of k and n, and Fig. 2(b) shows the results for larger values. It can be seen that as k and n increase, the constriction identified by this procedure grows. Figure 2(c) shows the tube model corresponding to the constriction as identified in Fig. 2(a); Fig. 2(d) shows the tube model corresponding to Fig. 2(b). The EASY synthesizer calculated formant frequencies $F_1=338$ Hz, $F_2=1819$ Hz, and $F_3=2140$ Hz for the tube configuration in Fig. 2(c); and 393, 1912, and 2643 Hz for Fig. 2(d).

Figure 3 summarizes the results of varying k and n through the entire range. Figure 3(a) shows the length of the constriction (in gridlines) determined by the parametrized constriction-finding procedure. For values of roughly $2.5<k<4.5$, the plot shows a "shelf" on which the constriction length is not very sensitive to changes in n. As k increases beyond about 5, depending on n, the constriction found by this procedure expands to eventually include the entire vocal tract, a length of 53 gridlines. A black shadow under Fig. 3(a) shows the range of (k,n) under the shelf. Smaller values of k and n can be rejected on acoustic grounds (see below).

Figure 3(b) shows the location of the center of the constriction. As k increases, the constriction location again reaches a stable shelf around gridlines 40–45, but when k is greater than about 5, the constriction moves to a more posterior location than would be expected. On the other hand, the location of the center of the constriction determined by this procedure is relatively insensitive to the value of n. In Fig. 3(b), the shadow shows the range of parameter values under the shelf. Again, smaller k and n can be rejected on acoustic grounds (see below).

Figure 4 shows the results of synthesis of each candidate tube model resulting from the (k,n) combinations. The values of F_1, F_2, and F_3 identified by the EASY synthesizer for each

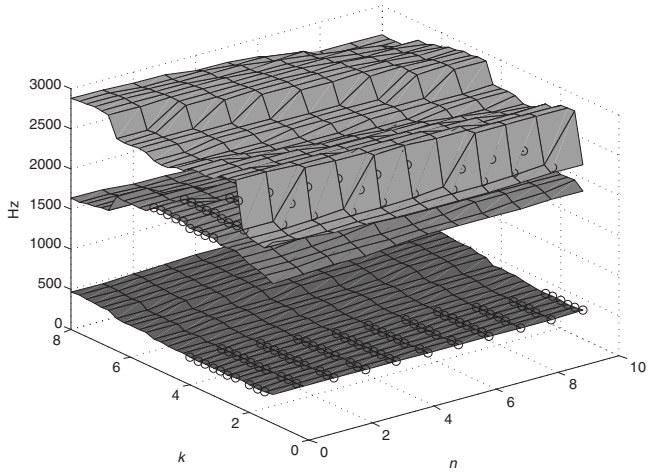


Fig. 4. Results of EASY synthesis of each three-tube model derived from the constriction-finding procedure. F_1, F_2, and F_3 are plotted as a function of k and n. On the F_1 and F_2 surfaces, points highlighted with an o emphasize the regions around the minimum F_1 and maximum F_2 values. The regions are values of (k,n) where the value of the formant was within 2.5% of the minimum F_1 or maximum F_2.

three-tube model (i.e., each combination of k and n) are plotted. F_1 varies from 330 to 620 Hz, F_2, varies from 1560 to 2000 Hz, and F_3, varies from 2136 to 2910 Hz. The figure is annotated with symbols showing the range of k and n values which produced tube models that satisfied various acoustic constraints appropriate to /i/. On the F_1 and F_2 surfaces, points highlighted with an "o" are (k,n) combinations where the value of F_1 was within 2.5% of the minimum F_1, or maximum F_2, respectively. The percentage was chosen solely for illustrative purposes to emphasize the regions enclosing the ranges of points with the minimum F_1 and maximum F_2.

In Fig. 4, as in Fig. 3, it can be seen that the results are not very sensitive to n, but vary more in the k direction. Generally, realistic values of F_1 in the "basin" around the minimum value of F_1 are limited to $k \leq 3$. Similarly, the F_3 surface shows that values of $k < 2.5$ give unrealistically low values of F_3. Utilizing these constraints on the value of k, it turns out that the plateau of high F_2 values crosses the $k=3$ line at $n > 2$. However, since the acoustic-to-articulatory mapping itself is well-known to be indeterminate (Atal *et al.*, 1978), it should not be surprising that constraints on the values of F_1, F_2, and F_3 are not sufficient to uniquely determine the parameters of the procedure for finding the constriction in a given x-ray profile.

The articulatory analysis was repeated for every high front vowel token. The results were then combined by counting the number of times each combination of k and n fell within the "plateau" or "shelf" for the specific token, as in Fig. 3. The results, weighted equally by speaker, are shown in Fig. 5. Since vowel tokens from 13 speakers were analyzed in this study, combinations of k and n in Fig. 5 that are larger, and approach $N=13$, are better for more speakers. In both Figs. 5(a) and 5(b), it can be seen again that the results are not very sensitive to the n, as long as $n \geq 3$. In the range $n \geq 3$, values of k that include the most tokens appear to fall in the range $2.5 < k < 3.5$.

4. Discussion

The method for defining vocal tract constrictions in this study is based on finding a contiguous set of gridlines around the gridline with the minimum cross-dimension. It has been shown that the method gives consistent results over certain ranges of two parameters: the number of cross-dimensions averaged with the minimum cross-dimension and the multiplier. It is not dependent on specific values chosen. Because the threshold cross-dimension used to define the constriction is based on a window of gridlines, the method is not sensitive to local variation in vocal tract shape. It has been tested across a number of speakers of different languages.

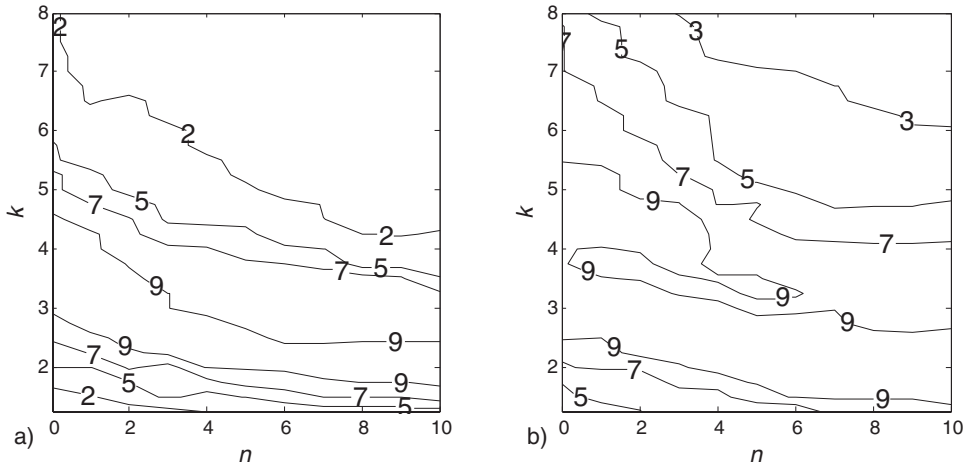


Fig. 5. The number of times each combination of k and n yielded appropriate constriction estimates. (a) shows the results for constriction length. (b) shows the results for constriction location. Results from each speaker weighted equally, so that the maximum value (not actually attained) could have been $N=13$.

Pragmatically, it seems like this method is easiest to understand with a small window size. Values of $n=3$ and $k=3.0$ seem suitable. These values also fall within the range of parameters that yield acoustically plausible three-tube models of the vowels in question.

Acknowledgment

This work was supported by Grant No. NIDCD-001247 to CReSS LLC.

References and links

Abramson, A. S., Martin, S., Schlaeger, R., and Zeichner, D. (**1962**). *Mandarin Chinese X-Ray Film in Slow Motion with Stretched Sound* (Columbia University, Columbia-Presbyterian Medical Center and Haskins Laboratories, New York).

Atal, B. A., Chang, J. J., Mathews, M. V., and Tukey, J. W. (**1978**). "Inversion of articulatory-to-acoustic transformation in the vocal tract by a computer-sorting technique," J. Acoust. Soc. Am. **63**, 1535–1555.

Bothorel, A., Simon, P., Wioland, F., and Zerling, J.-P. (**1986**). *Cinéradiographie des voyelles et consonnes du Français (Cineradiography of French vowels and consonants)* (Institut de Phonétique de Strasbourg, Strasbourg).

Fant, G. (**1960**). *Acoustic Theory of Speech Production* (Mouton, The Hague, Netherlands).

Iskarous, K. (**2005**). "Patterns of tongue movement," J. Phonetics **33**, 363–381.

Jackson, M. T.-T., and McGowan, R. S. (**2008**). "Predicting midsagittal pharyngeal dimensions from measures of anterior tongue position in Swedish vowels: Statistical considerations," J. Acoust. Soc. Am. **123**, 336–346.

McGowan, R. S., and Wilhelms-Tricarico, R. (**2005**). "An educational articulatory synthesizer, EASY," J. Acoust. Soc. Am. **117**, 2543.

Munhall, K. G., Vatikiotis-Bateson, E., and Tohkura, Y. (**1994**). *X-Ray Film Database for Speech Research* (ATR Human Information Processing Research Laboratories, Kyoto, Japan).

Navarro-Tomás, T. (**1916**). "Siete vocales Españolas (Six Spanish vowels)," Revista de Filologia Española (Review of Spanish Philology) **3**, 51–62.

Ohnesorg, K. and Svarný, O. (**1955**). *Études Expérimentales des Articulations Chinoises. (Experimental Studies on Chinese Articulations.)* (Czech Academy, Prague), Vol. **65**, Issue No. 5.

Parmenter, C. E. and Treviño, E. (**1932**). "An X-ray study of Spanish vowels," Hispania **15**, 483–496.

Perkell, J. S. (**1969**). *Physiology of Speech Production: Results and Implications of a Quantitative Cineradiographic Study. Cambridge* (MIT, Cambridge, MA).

Rochette, C. (**1973**). *Les Groupes de Consonnes en Français (Consonant Groups in French)* (Laval University Press, Quebec).

Rochette, C. (**1977**). "Radiologie et phonétique (Radiology and phonetics)," Vie Médicale au Canada Français (Medical Life in French Canada) **6**, 55–67.

Russel, G. O. (**1929**). "The mechanism of speech," J. Acoust. Soc. Am. **1**, 83–109.

Stevens, K. N. (**1998**). *Acoustic Phonetics* (MIT, Cambridge, MA).

Wood, S. (**1982**). "X-ray and model studies of vowel articulation," University of Lund Phonetics Laboratory Working Papers **23**, 1–49.

Multiple instances of vocal sensorimotor adaptation to frequency-altered feedback within a single experimental session

Colin S. Hawco
Department of Psychology, Wilfrid Laurier University, Waterloo, Ontario N2L 2C5, Canada
hawc2550@wlu.ca

Jeffery A. Jones[a)]
Department of Psychology and Centre for Cognitive Neuroscience, Wilfrid Laurier University, Waterloo, Ontario N2L 2C5, Canada
jjones@wlu.ca

Abstract: Vocal sensory-motor adaptation is typically studied by introducing a prolonged change in auditory feedback. While it may be preferable to perform multiple blocks of adaptation within a single experiment, it is possible that a carry-over effect from previous blocks of adaptation may affect the results of subsequent blocks. Speakers were asked to vocalize an /a/ sound and match a target note during ten adaptation blocks. Each block represented a unique combination of target note and shift direction. The adaptation response was found to be similar for all blocks, indicating that there were no carry-over effects from previous blocks of adaptation.

PACS numbers: 43.70.Mn, 43.70.Bk, 43.70.Gr [AL]
Date Received: August 7, 2009 **Date Accepted:** November 17, 2009

1. Introduction

Understanding the contribution of auditory feedback to speech production is important for theories of speech motor control. The role of auditory feedback during speech production has been investigated by altering feedback during ongoing vocalizations. When a random feedback alteration is introduced during a vocalization, there is an online compensation response in the direction opposite the feedback alteration (Larson, 1998; Burnett *et al.*, 1998; Natke and Kalveram, 2001). For example, Natke and Kalveram (2001) had participants vocalize a nonsense word and shifted the fundamental frequency (F0, vocal pitch) of auditory feedback down on 20% of trials. They found an online compensation response in the direction opposite to the F0 shift. This online compensation response uses auditory feedback to correct vocal output and maintain vocal stability, and is delayed by approximately 100–150 ms after a feedback alteration is introduced (Burnett *et al.*, 1998).

When a feedback alteration is introduced and left in place for a prolonged period, sensory-motor adaptation occurs as the system adjusts to compensate for this novel feedback environment. Houde and Jordan (1998, 2002) examined sensory-motor representations for formants by shifting F1 and F2. Participants compensated for the feedback alterations by modifying their formant production in the direction opposite the shift, with these modifications persisting when auditory feedback was removed. Other studies examining vocal control of formants have found similar after-effects when a feedback alteration is removed (Purcell and Munhall, 2006; Villacorta *et al.*, 2007; Tourville *et al.*, 2008). Jones and Munhall (2000, 2002) also noted prominent after-effects after participants heard their F0 shifted over many trials.

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

These after-effects are evidence of sensory-motor adaptation; because of the sensory-motor remapping during adaptation, the motor system must re-adjust to the original, pre-shift sensory-motor mapping when the feedback alteration is removed.

The results of studies on auditory feedback indicate that feedback is important for online correction during vocalization (Burnett *et al.*, 1998; Natke and Kalveram, 2001) and the maintenance of stored motor commands for vocal production (Guenther, 2006; Jones and Munhall, 2000, 2002; Purcell and Munhall, 2006). Auditory feedback therefore has a dual role in vocal production: to stabilize vocalizations as they are occurring and to modify motor plans for future vocalizations to accommodate for novel feedback environments.

Many previous studies examining sensory-motor adaptation have used a small number of post-shift "test" trials to determine if adaptation had occurred (Houde and Jordan 1998, 2002; Jones and Munhall, 2000, 2002). In studies examining the control of F0, the F0 value at utterance onset has more recently been used as a measure of vocal sensory-motor adaptation (Keough and Jones, 2009; Hawco and Jones, 2009). This measure allows for an online analysis of the time course of adaptation without the need to examine after-effects. Since feedback-based F0 control is typically delayed by 100–150 ms (Burnett *et al.*, 1998), an examination of the F0 data at utterance onset shows changes in the initiation of vocalization, which are not affected by online auditory feedback control. Although it allows for an examination of the course of adaptation, this technique suffers the same lack of power as experiments utilizing post-shift test trials, as there is only a single instance of the critical trials in the experiment (the first shift trials, first post-shift test trials, etc.).

Although several studies have examined sensory-motor adaptation in the vocal system, there are still many questions to be investigated. Some studies have suggested that there may be differences between singers, who are highly trained in vocal control, and non-singers when they adapt to changes in F0 (Jones and Keough, 2008; Keough and Jones, 2009). There are several other special populations who are of interest in studies of sensory-motor adaptation, including Parkinson patients, amusics, schizophrenics (who may have a disruption in the signals sent to auditory cortex to identify speech as self-generated; Ford and Mathalon, 2004), and children at various stages of development. In addition, the neurological mechanisms involved in feedback control of vocalization, and how these mechanisms change during adaptation, are not well understood.

Most studies of sensory-motor adaptation during vocalization involve a single block of trials. With only a single set of critical trials, there is a limit to the statistical power of the analysis, as only a small number of trials for each participant can be used to test for adaptation. Using multiple blocks of adaptation trials would overcome this issue, but it is possible that carry-over effects may occur during repeated blocks of adaptation, resulting in changes to the sensory-motor adaptation observed in subsequent blocks. While many questions regarding sensory-motor adaptation during vocalizations could be addressed using a between-subject design, it may be preferable to be able to perform multiple adaptation blocks within a single participant. This not only reduces the number of participants that would be needed if a between-subject design were used (a particularly important issue when dealing with special populations) but it may also be critical to examining vocal sensory-motor adaptation using neuroimaging techniques [e.g., event-related potentials (ERP), magnetoencephalography (MEG), and functional magnetic resonance imaging (fMRI)]. However, if there are carry-over effects across multiple blocks of adaptation (for example, if the motor system becomes faster to adapt after repeated exposure to adaptation), it may be difficult to examine experimental differences between blocks as they would be confounded with these carry-over effects.

The present study demonstrates the feasibility of using multiple blocks of adaptation trials in a within-subject frequency-altered feedback design. In order to determine if we could repeat blocks of adaptation without carry-over effects, we attempted to minimize possible carry-over effects in two ways. First, in each block, feedback-shifted trials were presented in the middle of the block, with unshifted trials present at the beginning and end of each block. Having no feedback alterations at the end of each block should allow a "de-adaptation" response, where the participants re-adjust to unaltered feedback prior to the initiation of the next block. We also

used different target notes for vocalization in each block and shifted feedback either up or down in frequency. If the adaptation response does not show a change over blocks, this would suggest that no carry-over effects from previous exposure to adaptation have occurred, and future studies can use blocked adaptation designs without confounds from repeating blocks of adaptation. Data were collected as part of a larger ERP study that will not be presented here, as the ERP findings are not relevant to the current discussion.

2. Methods

2.1 Participants

Data from 14 participants (ages 18–21, 7 males). All participants were undergraduate students at Wilfrid Laurier University participating for course credit and reported no formal training as singers. Informed consent was obtained from all participants.

2.2 Apparatus and procedure

Participants were seated in a comfortable chair in a closed room. They were given a headset with attached boom microphone (Sennheiser HMD 280-13). Vocalizations were sent from the microphone to a digital mixer (828 mkII, MOTU), which passed the voice signal to a digital signal processor (VoiceOne, TC-Helicon). The participant's auditory feedback was mixed with multispeaker babble (20 speakers simultaneously reading different passages; Auditec, St. Louis, MO) played at 90 dB sound pressure level (SPL) and returned to the participant as auditory feedback. The multispeaker babble served to mask air and bone-conducted feedback. The unaltered voice signal was digitally recorded (H4, Zoom) at a sampling rate of 44.1 kHz. Participant's vocalizations were amplified such that a 75 dB SPL production (measured approximately 5 cm from the mouth) was played at approximately 90 dB SPL. Participants vocalized above 75 dB SPL (resulting in feedback of over 90 dB SPL), and participants indicated they could clearly hear their auditory feedback over the multi-speaker babble.

At the beginning of each trial, an auditory cue was played at approximately 92 dB SPL for 1 s. The cue was a male or female voice producing the vowel sound /a/ at a specific target frequency. Target notes for female participants were G4 (392 Hz), E4 (329.63 Hz), D4 (293.66 Hz), B3 (246.94 Hz), and A3 (220 Hz). The target notes for males were G3 (192 Hz), E3 (164.83 Hz), D3 (146.83 Hz), B2 (123.47 Hz), and A2 (110 Hz). Cues were recorded from trained singers who were asked to match a specific pitch. Their productions were processed using the speech modification algorithm speech transformation and representation using the adaptive interpolation of weighted spectrum (Kawahara *et al.*, 1999) to ensure that the mean F0 of the cue was equal to the desired target. Participants were instructed to wait until the cue had ended, and to then produce an /a/ sound, matching the pitch of the cue, until a tone sounded (2 s after utterance onset), resulting in approximately 2 s of vocalization per trial.

The experiment was divided into ten blocks of 42 trials. Each block had a specific target note which was played throughout that block. A trial consisted of an auditory cue (of the target note for that block) and a vocalization. Each block contained 9 or 12 trials with normal, unshifted auditory feedback (pre-shift baseline trials), 18 or 21 trials in which feedback was shifted up or down 100 cents for the entire vocalization (shift trials), followed by 9 or 12 trials of normal, unshifted feedback (post-shift test trials). The different number of trials in each block served to make the onset of shifting in each block less predictable. At the end of each block, there was a break, during which participants could rest and drink water. A schematic diagram of the design is presented in Fig. 1, showing the target note and shift direction of each block.

Voice data were segmented into trials, and F0 was calculated for the first 1500 ms of vocalization using the autocorrelation algorithm in PRAAT (Boersma, 2001), with an F0 value calculated for every 5 ms of voice data. Data were converted into cents using the formula $100^{*}(39.86^{*}\log 10(\mathrm{F0}/\mathrm{target}))$, where "F0" was the F0 produced by the participant, and "target" was the target note. In order to assess adaptation, the median F0 (in cents) was calculated for the initial 50 ms of vocalization. We chose to examine the first 50 ms of production because it is not

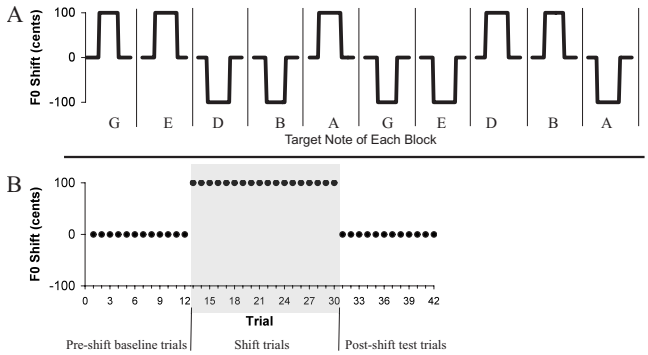


Fig. 1. (A) Schematic diagram of the experimental design, showing the F0 shift in each block (with blocks separated by vertical lines). The target note of each block is shown and was the same for all trials in that block. Target notes for female participants were G4 (392 Hz), E4 (329.63 Hz), D4 (293.66 Hz), B3 (246.94 Hz), and A3 (220 Hz). The target notes for males were G3 (192 Hz), E3 (164.83 Hz), D3 (146.83 Hz), B2 (123.47 Hz), and A2 (110 Hz). Although each target note was presented in two separate blocks, it was shifted in a different direction on each presentation. (B) Detailed view of the first block of trials. Each block consisted of 42 trials, with 9 or 12 pre-shift baseline trials, 18 or 21 shift trials, and 9 or 12 post-shift test trials. Shift trials for the presented block are highlighted in gray. The F0 shift was maintained for the entirety of all shift trials, and there was no F0 shift in any of the pre-shift baseline or post-shift test trials.

affected by volitional feedback control (Burnett *et al.*, 1998). We also calculated the median F0 for the first 1500 ms of vocalization to determine accuracy for hitting the target note.

3. Results

To test for differences in accuracy for hitting the target notes, the median F0 for the first 1500 ms of vocalization for baseline trials 7–9 (the last three baseline trials which were present in all blocks) was analyzed. A 2×5 (repetition by target note) repeated measure analysis of variance (ANOVA) was conducted to test both for any effects on vocal accuracy of repeating the target note in a subsequent block and differences in hitting each target note. No statistically significant differences were found for the target notes or for repetitions of target notes, and there was no interaction between target note and repetition (all $p > 0.1$).

Repeated measure ANOVAs were conducted to test for adaptation. Analyses of upward shifted and downward shifted blocks were first conducted separately. For each shift direction, a 5×5 (block by phase) repeated measure ANOVA was conducted on the median F0 values for the initial 50 ms of the utterances. The factor Phase represented different phases of feedback alteration within the block and included the average of the pre-shift baseline trials 7–9, the shift trials 1–3, shift trials 15–18 (the last three shift trials which were present in all blocks), the post-shift test trials 1–3, and post-shift test trials 7–9. Figure 2 shows the median 50 ms data for each phase for all blocks. For two participants, some trials were missing from the last block (downward shifts). These trials were treated as missing data points. For both upward and downward shifted trials, a main effect of block [$F(4,52) = 8.59$, $p < 0.001$ and $F(4,44) = 6.61$, $p < 0.001$, respectively] and phase [$F(4,52) = 10.96$, $p < 0.001$ and $F(4,44) = 6.72$, $p < 0.001$, respectively] was found. The main effect of block was caused by different F0 values during the first 50 ms of vocalization for different target notes (see Fig. 2). The main effect of phase indicated adaptation occurred, as utterance onset was changed in response to the shifted feedback. A block by phase interaction was not found for either shift direction, indicating that the pattern of adaptation did not significantly differ across blocks.

To further examine the main effect of phase, planned comparisons were conducted for both upward and downward shifted data. For the 50 ms data, significant differences were found in the upward shifted trials between baseline trials 7–9 and shift trials 15–18 ($p = 0.002$) and post-shift test trials 1–3 ($p = 0.028$), and in the downward shifted trials between pre-shift baseline trials 7–9 and shift trials 15–18 ($p < 0.001$) and post-shift test trials 1–3 ($p < 0.001$). For both up-

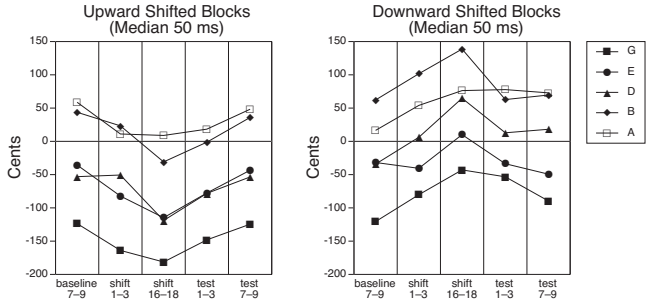


Fig. 2. Median F0 (in cents) for the first 50 ms of production for each target note shifted upward or downward for each phase of shifts used in the statistical analysis. Adaptation responses are in the direction opposite the feedback alteration in order to offset for the altered feedback.

ward and downward shifts, there was no significant difference in the 50 ms data between pre-shift baseline trials 7–9 and post-shift test trials 7–9 ($p > 0.1$), indicating that participants' initial F0 returned to pre-shift baseline values by the end of the block.

To test for the effects of order of the blocks, data were normalized such that the pre-shift baseline phase was equal to zero, and all responses were in the positive direction. First, a $2 \times 5 \times 5$ (shift direction by block by phase) ANOVA was conducted. No main effects of shift direction or block, and no interactions, were observed. There was a main effect of phase [$F(4,44) = 20.57$, $p < 0.001$]. A further 10×5 (order by phase) ANOVA was conducted, with no main effect of order and no interaction observed, but again the main effect of phase was significant [$F(4,44) = 17.88$, $p < 0.001$].

4. Discussion

The planned comparisons showed a significant difference between the last pre-shift baseline trials and the last shift trials, as well as the last pre-shift baseline trials and the first post-shift test trials, for F0 at utterance onset. This pattern of results shows that adaptation occurred. Participants altered their initial voice F0 to adapt to the feedback alteration, and this change in motor production continued when the feedback alteration was removed. No significant difference was found between the final pre-shift baseline and final post-shift test trials, demonstrating that the F0 at utterance onset returned to the pre-shift baseline level after several trials of normal feedback. This suggests that the sensory-motor mapping at the end of the post-shift test trials was identical to the sensory-motor mapping for the pre-shift baseline trials, indicating that the sensory-motor mapping had returned to the pre-adaptation state before the onset of the next block of the experiment.

We did not observe carry-over effects in adaptation across different blocks. In the current design, each block represented the unique pairing of a target note and shift direction, and therefore a unique context for adaptation to occur. We have shown that when each block is a unique context, adaptation effects do not carry-over across repeated blocks. This is important in that it demonstrates the feasibility of using a within-subject blocked design for experiments on vocal adaptation. Such designs may allow for more efficient experiments to test vocal adaptation to different conditions or different stimuli or to allow for repeated blocks of adaptation for added statistical power. Using repeated blocks can be very important when testing special populations, where a limited number of participants may be available. However, it should be noted that all the participants in this study were non-singers. Given the potential differences in vocal motor control between singers and non-singers (Keough and Jones, 2009), the results of this study should not be applied to trained singers without further testing. Also, it would be interesting to test if the lack of carry-over effects in different vocalization contexts applies to formant control as well as F0 control.

When conducting neuroimaging research (ERP, MEG, fMRI, etc.), it is necessary to have repetitions of the experimental conditions in order to conduct a statistical analysis. If only a single block of adaptation trials was used when conducting imaging studies, it would be difficult to construct a meaningful statistical analysis. Furthermore, any such analysis would be inherently confounded with time, as the pre-shift baseline trials would always precede the shifted trials. This represents a major confound in the contexts of an imaging study on vocal adaptation, where activation patterns may change as the participant adjusts to the novel environment (such as in fMRI).

The lack of a statistical effect of block when the data were normalized suggests that adaptation responses to different target notes are comparable. This, plus the fact that we found no statistical differences across blocks when the data were normalized, suggests that adaptation effects across different notes can be directly compared.

Acknowledgments

This research was supported by the National Institute of Deafness and Communicative Disorders Grant No. DC-08092 and a grant from the Natural Sciences and Engineering Research Council of Canada.

References and links

Boersma, P. (**2001**). "Praat, a system for doing phonetics by computer," Glot. Int. **5**, 341–345.
Burnett, T. A., Freedland, M. B., Larson, C. R., and Hain, T. C. (**1998**). "Voice F0 responses to manipulations in pitch feedback," J. Acoust. Soc. Am. **103**, 3153–3161.
Ford, J. M., and Mathalon, D. H. (**2004**). "Electrophysiological evidence of corollary discharge dysfunction in schizophrenia during talking and thinking," J. Psychiatr. Res. **38**, 37–46.
Guenther, F. H. (**2006**). "Cortical interactions underlying the production of speech sounds," J. Commun. Disord. **39**, 350–365.
Hawco, C. S., and Jones, J. A. (**2009**). "Control of vocalization at utterance onset and mid-utterance: Different mechanisms for different goals," Brain Res. **1276**, 131–139.
Houde, J. F., and Jordan, M. I. (**1998**). "Sensorimotor adaptation in speech production," Science **279**, 1213–1216.
Houde, J. F., and Jordan, M. I. (**2002**). "Sensorimotor adaptation of speech I: Compensation and adaptation," J. Speech Lang. Hear. Res. **45**, 295–310.
Jones, J. A., and Keough, D. (**2008**). "Auditory-motor mapping for pitch control in singers and nonsingers," Exp. Brain Res. **190**, 279–287.
Jones, J. A., and Munhall, K. G. (**2000**). "Perceptual calibration of F0 production: Evidence from feedback perturbation," J. Acoust. Soc. Am. **108**, 1246–51.
Jones, J. A., and Munhall, K. G. (**2002**). "The role of auditory feedback during phonation: Studies of Mandarin tone production," J. Phonetics **30**, 303–320.
Kawahara, H., Masuda-Katsuse, I., and de Cheveigne, A. (**1999**). "Restructuring speech representations using a pitch-adaptive time-frequency smoothing and an instantaneous-frequency-based F0 extraction: Possible role of a repetitive structure in sounds," Speech Commun. **27**, 187–207.
Keough, D., and Jones, J. A. (**2009**). "The sensitivity of auditory-motor representations to subtle changes in auditory feedback while singing," J. Acoust. Soc. Am. **126**, 837–846.
Larson, C. R. (**1998**). "Cross-modality influences in speech motor control: The use of pitch shifting for the study of F0 control," J. Commun. Disord. **31**, 489–503.
Natke, U., and Kalveram, K. T. (**2001**). "Effects of frequency-shifted auditory feedback on fundamental frequency of long stressed and unstressed syllables," J. Speech Lang. Hear. Res. **44**, 577–584.
Purcell, D. W., and Munhall, K. G. (**2006**). "Adaptive control of vowel formant frequency: Evidence from real-time formant manipulation," J. Acoust. Soc. Am. **120**, 966–977.
Tourville, J. A., Reilly, K. J., and Guenther, F. H. (**2008**). "Neural mechanisms underlying auditory feedback control of speech," Neuroimage **39**, 1429–43.
Villacorta, V. M., Perkell, J. S., and Guenther, F. H. (**2007**). "Sensorimotor adaptation to feedback perturbations of vowel acoustics and its relation to perception," J. Acoust. Soc. Am. **122**, 2306–2319.

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.

Post-transient shifts in auditory lateralization: Effects of target level (L)

Yoav Arieh[a)]
Department of Psychology, Montclair State University, Montclair, New Jersey 07043

Lawrence E. Marks
John B. Pierce Laboratory and Yale University, New Haven, Connecticut 06519

(Received 26 May 2009; revised 10 November 2009; accepted 10 November 2009)

Exposing one ear to a series of brief 80 dB sound pressure level (SPL) inducing tones reduces the tendency to lateralize subsequent target tones to that ear and shifts the point of subjective equality (PSE) toward the unexposed ear. Furthermore, targets with average SPLs of 60 and 80 dB at the two ears showed similar changes in PSE. These results support and extend earlier findings of Arieh and Marks (2007). Percept. Psychophys. **69**, 523–528 and suggest the presence of a mechanism, located at least partially before the site of binaural integration, that depresses the magnitude of intensity information from the ear in response to moderately intense transient stimuli. Assuming that lateralization depends on the ratio of the magnitude of intensity information from the two ears that reach the central integrator, the results imply that the inducer reduced the representations of magnitude of 60 and 80 dB test tones in equal proportion.
 [DOI: 10.1121/1.3270393]

PACS number(s): 43.66.Mk, 43.66.Cb, 43.66.Pn, 43.66.Qp [CJP] Pages: 1–4

I. INTRODUCTION

Recent psychoacoustic research has identified a new time-dependent phenomenon, evident in the perceived magnitude of transient tones; the loudness of a pure tone is reduced when it is preceded by a stronger, transient inducer tone at a nearby frequency (Marks, 1988, 1994; Scharf *et al.*, 2002; Wagner and Scharf, 2006). First labeled differential context effect (DCE) and later induced loudness reduction (ILR), the effect is robust; under optimal conditions, the reduction in loudness can extend to about 10 dB or 50% in sones (Arieh and Marks, 2003; see also Epstein, 2007 for review), and some residual effects may last as long as several minutes after the end of exposure (Arieh *et al.*, 2005; Epstein and Gifford, 2006). Important for the present work is the finding that ILR transfers only partially between the ears. That is, the loudness reduction is smaller when the inducer is played to one ear but loudness is measured in the contralateral ear compared to the ipsilateral ear (Marks, 1996; Nieder *et al.*, 2007). This result suggests that ILR is rooted in processes that take places at two loci; central processes, subsequent to binaural integration, suppress intensity information received from both ears, whereas peripheral processes, prior to binaural integration, suppress intensity information arising separately from each ear.

Arieh and Marks (2007) suggested that ILR may reflect the operation of a general-purpose mechanism of auditory intensity suppression that is activated by moderately loud transient acoustic signals. If true, then this general mechanism should reveal itself in other intensity-dependent processes besides the perception of loudness, such as lateralization. If exposing one ear to transient inducers produces greater suppression in the exposed ear than in the contralateral ear, then the interaural difference in the magnitude of suppression should reveal itself not only in the loudness of subsequent test tones, but also in their lateralization. That is, reduced sensitivity in the exposed ear should distort interaural intensity differences (IIDs) that are crucial for lateralization. Consistent with this hypothesis, Arieh and Marks (2007) showed that transient, 80 dB sound pressure level (SPL) inducing tones presented to one ear affected judgments of lateral position of subsequent 60 dB SPL target tones. Presented with target tones comprising unequal SPLs to the two ears, subjects were less likely to lateralize the target tones to the exposed ear. In the present report, we extend the results of Arieh and Marks (2007), which they labeled as post-transient shifts (PTSs), by examining the effect of the target level on the magnitude of the PTSs pro-

[a)]Author to whom correspondence should be addressed. Electronic mail: ariehy@mail.montclair.edu

duced by monaural inducers. Specifically, we compared the lateralization of 60 and 80 dB SPL targets after monaural exposure to an 80 dB SPL inducer.

Although nothing is yet known about the effect of target level on PTSs for lateralization, Mapes-Riordan and Yost (1999) did investigate the effect of target level on ILR. Those investigators used a loudness matching method, in which a comparison tone was varied from trial to trial until its loudness matched that of a test tone. Following a baseline match between the test and the comparison tones, an inducer was inserted before each trial. The inducer generally reduced the SPL of the matching point, presumably because it reduced the loudness of the target, relative to its baseline value. The difference between the SPLs of the matching point in the experimental and baseline conditions quantified the loudness reduction in the test tone. Importantly, their results suggest that the reduction in the target's loudness follows an inverted U-shaped intensity function of the target's SPL; inducers of 80 dB SPL reduced the loudness of subsequent test tones at 60–70 dB SPL by as much as 10 dB, but reduced the loudness of test tones at 40 dB SPL by only 4 dB, and reduced the loudness of test tones at 80 dB SPL by only 1–3 dB. If a single mechanism is responsible for induced reduction in loudness and PTSs in lateralization, then we would expect that PTSs in lateralization will be greater in targets at 60 dB than at 80 dB.

In the present study, we followed closely the procedure of Arieh and Marks (2007) and examined the intensity tuning function of PTSs in auditory lateralization. We asked subjects to judge the lateral position of pure tones whose perceived location was manipulated by adjusting the interaural difference in SPL. In the baseline condition, subjects judged the location of targets having average SPLs of either 60 or 80 dB; and in the experimental condition, they repeated the task while listening periodically to an 80 dB SPL inducer in one ear. The amount of PTSs was calculated as the difference in the likelihood of lateralizing the target tone to the exposed ear in the baseline and experimental conditions.

II. METHOD

A. Listeners

Sixteen men and women (age range 18–33 yrs) associated with the Department of Psychology at Montclair State University participated in the study. All reported normal hearing and received course credit for participating.

B. Apparatus

A MATLAB program running on a Dell Pentium-IV personal computer controlled all aspects of stimulus presentation, data collection, and on-line computations. Auditory signals were generated by a Tucker-Davis System 3 Real Time processor at a sampling frequency of 50 kHz. The signals were then appropriately attenuated in each ear's channel (Tucker-Davis PA5 module) and delivered through calibrated TDH-49 headphones mounted in MX41/AR cushions. Responses were entered via the computer's numeric keypad.

C. Stimuli

The inducing tones lasted 50 ms and had a frequency of 2500 Hz and a level of 80 dB SPL. The target tones had the same duration and frequency as the inducers, but the SPLs at the right and left ears were adjusted to vary the perceived location relative to the midline, while keeping mean SPL constant. In one session, the mean SPL of the target was 60 dB [(left SPL+right SPL)/2=60], and for a second session, the mean SPL was set to 80 dB [(left SPL+right SPL)/2 =80]. Targets at the two ears were simultaneous. For example, in one stimulus, the SPL was 61 dB at the right ear and 59 dB at the left. In this case, we expected that, at baseline, listeners would tend to judge the stimulus to be on the right more often than the left. Overall, for each session, we created six dichotic stimuli, three of which should be lateralized to the right side and three to the left side (right minus left ear dB differences of −4, −2, −1, 1, 2, and 4 dB).

D. Procedure

Each subject served in two sessions—average target level set at 60 or at 80 dB SPL—and each session contained two conditions, i.e., baseline and induction (or experimental). In the baseline condition, subjects were asked to decide, by pressing a key on the computer's keypad, whether the stimulus was perceived to the left or to the right side. After an intertrial interval of 1 s, the next stimulus was presented. The six stimuli were presented 50 times each and the resulting 300 trials were given in a different random order to each listener.

The induction condition was similar to the baseline condition; however, at the start of the induction condition and every 15 trials thereafter, 10 inducer tones were played to the right ear at a rate of 1/s. The baseline condition was always given first, followed by the induction condition, and the order of the two sessions was counterbalanced across subjects. The baseline condition lasted about 15 min and the induction condition about 20 min, and the two sessions were separated by at least by 24 h.

III. RESULTS

The percentage of trials on which each of the six target tones was lateralized to the right was computed for each listener and each condition. In Fig. 1 we show the average percentages of "right ear" responses for targets at average SPLs of 60 and 80 dB in panels (A) and (B), respectively. Visual inspection of the data suggests that (a) when the inducers were played to the right ear, the average percentage of tones lateralized to that ear decreased relative to baseline, and (b) the magnitude of the shift in lateralization is comparable with 60 and 80 dB targets. When the target level was 60 dB, the average of right ear responses across the six stimuli was 49.36% in the baseline condition and 33.72% in the induction condition—a shift of 15.64% in favor of the unexposed (left) ear. When the target level was 80 dB, the averages were 51.55% and 36.75% in the baseline and induction conditions, respectively—a shift of 14.8% in favor of the unexposed ear.

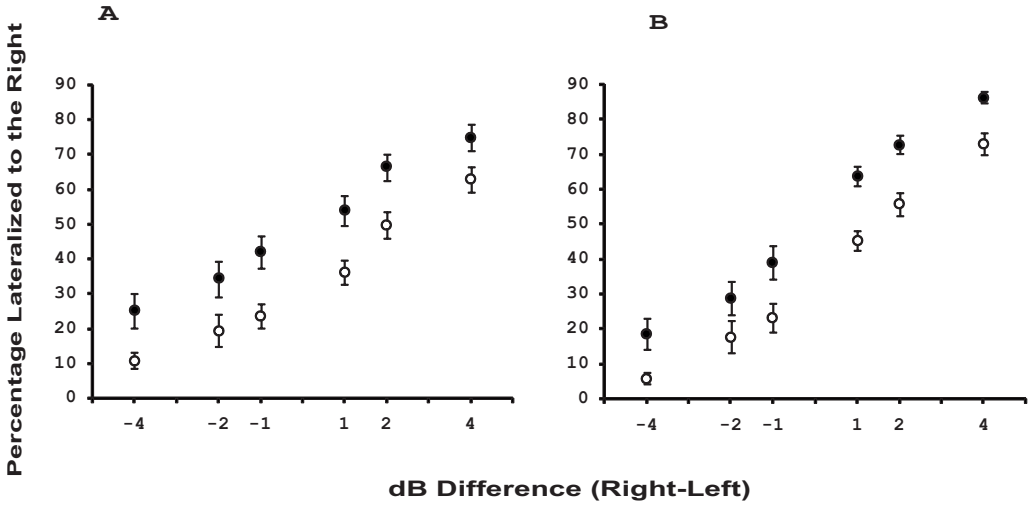


FIG. 1. The percentage of right lateralization responses to targets with average SPLs of (a) 60 and (b) 80 dB at baseline (filled circles) and after induction (empty circles). Error bars indicate one standard error of the mean.

To confirm the descriptive analysis, we computed the arcsine of the averaged proportions and submitted the transformed values to a repeated measure analysis of variance (ANOVA) with target level (80 vs 60 dB), stimulus difference (−4, −2, −1, 1, 2 and 4 dB), and condition (baseline vs induction) as within subject variables. The main effect of condition was significant [$F(1,15)=51.58$, MS$e=3.38$, and $p<0.01$] confirming our observation that the percentage of right ear responses was suppressed at the induction condition. Importantly, there was no interaction between target level and condition, indicating that the magnitude of "right ear" response suppression did not vary with the level of the target [$F(1,15)=0.037$, MS$e=0.01$, ns]. To round out the analysis, we note that the only other term that reached statistical significance was the large main effect of stimulus difference, which reflects the well-known way that intensity difference affects lateralization [$F(5,75)=103.8$, MS$e=4.7$, and $p<0.01$].

We also computed the PSE for each subject in each condition. The PSE is the difference between SPLs at the two ears producing a sound image at the midline, that is, producing 50% response at each ear. To compute the PSEs, we used linear equations to regress the average right side responses (y) against the corresponding decibel differences (x) and solved the equation for $x=50\%$. The linear fits were generally good, about 90% of the adjusted R^2 falling between 0.7 and 0.98. For the two baseline conditions, the average PSE was not statistically different from 0, −0.18, and −0.44 dB for the 60 and the 80 dB targets, respectively [$t(15)=-0.3$ and −0.8, ns for both]. Importantly, for both target levels, the PSE shifted similarly toward the unexposed ear; the sound level in the right ear needed to be raised by 2.32 dB in the 60 dB SPL condition and by 2.19 dB in the 80 dB SPL condition for the tones to be perceived at the center of the median plane (shift$=\mathrm{PSE}_{\mathrm{induction}}-\mathrm{PSE}_{\mathrm{baseline}}$). The difference between the conditions was not statistically significant [$t(15)=0.274$, ns].

IV. DISCUSSION

Exposing one ear to a series of inducers at 80 dB SPL significantly decreases the tendency to lateralize subsequent tones to that ear. The results replicate and extend the findings of Arieh and Marks (2007), which the authors labeled as PTS in auditory lateralization. PTS implies that the existence of a mechanism, located at least partly in the auditory periphery, reduces the magnitude of neural signals in response to moderately intense transient inducers. The new result that we report here is that the magnitude of the post-transient shifts in lateralization is the same with target levels of 60 and 80 dB SPLs. Let us assume that the "interaural information for lateralization" depends on the ratio of the intensity representations emanating from the two ears and reaching a central integrator. Let us assume further that the intensity of the neural signals (prior to PTSs) vary as power functions of sound intensity. Taking these assumptions together with the present finding that an inducer of 80 dB shifted target of both 60 and 80 by the equivalent of about 2.25 dB, we infer that PTSs may constitute a constant proportional decrease in the magnitude of the intensity information in the exposed ear, independent of the SPL of the target.

Arieh and Marks (2007) hypothesized that the same general-purpose intensity suppression mechanism underlies both ILR and PTSs. The present results raise two questions about that hypothesis: First, how does the magnitude of PTSs obtained here, as measured in decibels, compare with the magnitude of ILR that was obtained in other studies that used comparable stimulus parameter? Second, why should PTSs in lateralization be independent of target level, at least for the levels tested here, while ILR is not?

To answer the first question, it is most appropriate to compare the magnitude of PTSs measured here to the magnitude of the peripheral, ear-specific component of ILR. Presumably, the suppression of intensity information affects IIDs, and consequently affects lateralization, only to the extent that the suppression differentially affects the two ears. To assess the peripheral component of loudness reduction, however, it is necessary to make an assumption about the combination of peripheral and central components. A simple assumption is that that loudness reduction takes place in successive and independent stages: peripheral reduction, followed by central reduction. Thus, the total loudness reduc-

tion measured in the ipsilateral ear=peripheral reduction +central reduction. If true, then the peripheral reduction will be equal to the amount of ILR measured at the contralateral ear and the central reduction can be inferred by subtracting the peripheral reduction from the total reduction measured in the ipsilateral ear.

Using these assumptions together with the data of Marks (1996) and Nieder *et al.* (2007), we calculated the magnitude of the peripheral component to loudness reduction to be 4.2 dB in the former and 5.5 dB in the latter. Both of these values exceed the ear specific component of 2.25 dB in PTSs found here, but the discrepancy might reflect differences in testing procedure. For example, Nieder *et al.* (2007) used an adaptive procedure in which the inducers were presented on every trial, whereas in this study, a series of inducers was presented every 15 trials. Coupled with the fact that ILR tends to accumulate over repeated presentations (Arieh *et al.*, 2005), this procedural difference may have led to a greater effect in loudness than in lateralization.

The answer to the second question is even less clear. Why does ILR depend on the target's level—an 80 dB inducer reduces the loudness of 60 dB tone by about 10 dB, but the loudness of an 80 dB tone by only 1–3 dB (Mapes-Riordan and Yost, 1999)—whereas PTS in lateralization, as reported here, does not depend at all on the target's level—PTS was virtually identical with 60 and 80 dB test tones. At this point we can only offer two possibilities. One possibility follows from the distinction between peripheral and central contributions to the reduction in intensity information. PTS presumably reflects only the contribution of the peripheral, ear-specific component; otherwise, the IIDs that are crucial for lateralization will not be affected. Perhaps that contribution is constant in decibels, independent of the target's level, whereas the central contribution is maximal in targets around 60–70 dB, and smaller when targets are lower or higher in SPL. To test this possibility, it is necessary to measure directly the effect of target level on interaural transfer of ILR.

Another possibility is that PTS in lateralization and ILR are, at least in part, independent phenomena. Changes in lateralization do not always track loudness differences between the ears. For example, listeners who wore a monaural earplug over long time periods came eventually to set signals at the two ears close to equal intensity rather than equal loudness in order to center a sound image (Florentine, 1976; see also pp. 15–17 of Scharf, 1983 for extended discussion). The subjects presumably "learned" over time to compensate for the reduction in intensity, and the consequent change in loudness, at the plugged ear. These findings make at least plausible a possible dissociation between the functional roles of intensity information in loudness perception and in lateralization.

ACKNOWLEDGMENTS

The authors would like to thank Sibyl N. Arbelo for help in data collection and Christopher Plack and two anonymous reviewers for their cogent comments on an earlier version of this paper.

Arieh, Y., Kelly, K., and Marks, L. E. (**2005**). "Tracking the time to recovery after induced loudness reduction," J. Acoust. Soc. Am. **117**, 3381–3384.
Arieh, Y., and Marks, L. E. (**2003**). "Time course of loudness recalibration: Implications for loudness enhancement," J. Acoust. Soc. Am. **114**, 1550–1556.
Arieh, Y., and Marks, L. E. (**2007**). "Posttransient shifts in auditory lateralization," Percept. Psychophys. **69**, 523–528.
Epstein, M. (**2007**). "An introduction to induced loudness reduction," J. Acoust. Soc. Am. **122**, EL74–EL80.
Epstein, M., and Gifford, E. (**2006**). "A potential carry-over effect in the measurement of induced loudness reduction," J. Acoust. Soc. Am. **120**, 305–309.
Florentine, M. (**1976**). "Relation between lateralization and loudness in symmetrical hearing loss," J. Am. Audiol. Soc. **1**, 243–251.
Mapes-Riordan, D., and Yost, W. A. (**1999**). "Loudness recalibration as a function of level," J. Acoust. Soc. Am. **106**, 3506–3511.
Marks, L. E. (**1988**). "Magnitude estimation and sensory matching," Percept. Psychophys. **43**, 511–525.
Marks, L. E. (**1994**). "'Recalibrating' the auditory system: The perception of loudness," J. Exp. Psychol. Hum. Percept. Perform. **20**, 382–396.
Marks, L. E. (**1996**). "Recalibrating the perception of loudness: Interaural transfer," J. Acoust. Soc. Am. **100**, 473–480.
Nieder, B., Buus, S., Cazals, Y., and Scharf, B. (**2007**). "Loudness reduction induced by contralateral tone (L)," J. Acoust. Soc. Am. **122**, 35–37.
Scharf, B. (**1983**). "Loudness adaptation," in *Hearing Research and Theory*, edited by J. V. Tobias and D. D. Schubert (Academic, New York), Vol. **2**, pp. 1–56.
Scharf, B., Buus, S., and Nieder, B. (**2002**). "Loudness enhancement: Induced loudness reduction in disguise? (L)," J. Acoust. Soc. Am. **112**, 807–810.
Wagner, E., and Scharf, B. (**2006**). "Induced loudness reduction as a function of exposure time and signal frequency," J. Acoust. Soc. Am. **119**, 1012–1020.

The role of temporal fine structure in harmonic segregation through mistuning (L)

Brian C. J. Moore[a] and Brian R. Glasberg
Department of Experimental Psychology, University of Cambridge, Downing Street, Cambridge CB2 3EB, England

(Received 25 March 2009; revised 28 September 2009; accepted 19 October 2009)

Bernstein and Oxenham [(2008). J. Acoust. Soc. Am. **124**, 1653–1667] measured thresholds for discriminating the fundamental frequency, F0, of a complex tone that was passed through a fixed bandpass filter. They found that performance worsened when the F0 was decreased so that only harmonics above the tenth were audible. However, performance in this case was improved by mistuning the odd harmonics by 3%. Bernstein and Oxenham considered whether the results could be explained in terms of temporal fine structure information available at the output of a single auditory filter and concluded that their results did not appear to be consistent with such an explanation. Here, it is argued that such cues could have led to the improvement in performance produced by mistuning the odd harmonics.
 [DOI: 10.1121/1.3268509]

PACS number(s): 43.66.Hg, 43.66.Fe, 43.66.Ba [RLF] Pages: 5–8

I. INTRODUCTION

Thresholds for detecting a change in the fundamental frequency (F0) of a harmonic complex tone, which will be denoted F0DLs, are usually small when the tone contains low-numbered harmonics and larger when the tone contains only high-numbered harmonics (Moore and Glasberg, 1988; Carlyon and Shackleton, 1994). If a complex tone with many harmonics is passed through a fixed bandpass filter, F0DLs are small when the F0 is relatively high (so that the filter passes some resolved harmonics) but become larger when the F0 is decreased, such that the lowest harmonic in the passband is increased from about the 6th or 7th to about the 14th (Houtsma and Smurzynski, 1990; Bernstein and Oxenham, 2003, 2005, 2008; Moore *et al.*, 2006). The progressive worsening in performance as the number of the lowest audible harmonic is increased from 7 to 14 has been attributed to the progressive loss of resolution of harmonics (Carlyon and Shackleton, 1994), to the progressive loss of ability to use temporal fine structure (TFS) information (Moore and Moore, 2003; Moore *et al.*, 2006, 2009), and/or to a center-frequency-specific limitation in the ability of the auditory system to measure neural inter-spike intervals for the determination of pitch (Moore, 1982; Bernstein and Oxenham, 2005; de Cheveigné and Pressnitzer, 2006). According to this last idea, the accuracy with which time intervals can be measured accurately for a given "channel" [corresponding to a specific center frequency or characteristic frequency (CF)] decreases as the intervals increase above about 8/CF. For example, for a channel with a CF of 2000 Hz, the accuracy decreases for intervals above about 4 ms.

In a recent study, Bernstein and Oxenham (2008) measured F0DLs for complex tones that were passed through a bandpass filter with lower and upper cutoff frequencies of 1.5 and 3.5 kHz, respectively. The slope of the filter outside the passband was 50 dB/octave, and a background noise was used to limit the audibility of components falling outside the passband of the filter. They measured F0DLs for F0s of 50, 75, 100, 125, 150, 175, and 200 Hz. The phases of the components were selected randomly for each stimulus. In one set of conditions (harmonic conditions), all components had the same F0. In another set of conditions (F0-shifted conditions), the F0 for the odd components was 3% higher than the F0 for the even components (but all components were changed by the same percentage in the F0-discrimination task). Both diotic and dichotic listening conditions were included, but we focus here on the diotic conditions, where all components were presented to both ears. For the harmonic condition, the F0DLs were relatively small (less than 2%) for F0s of 200 and 175 Hz, but increased to about 8% as the F0 was decreased to 125 Hz, and then remained roughly constant with further decreases in F0. For the F0-shifted condition, the F0DLs were relatively small (2% or less) for all F0s from 200 to 100 Hz, but increased when the F0 was decreased to 75 or 50 Hz. Thus, for F0s around 125 Hz, F0DLs were markedly smaller for the F0-shifted than for the harmonic condition.

Bernstein and Oxenham (2008) suggested that the shift in the F0 of the odd components allowed them to be perceptually segregated from the even harmonics. Supplementary experiments showed that the shift did not reliably affect the ability of listeners to hear out individual harmonics. Based on this and other results, they reached the important conclusion that peripherally resolved harmonics are neither necessary nor sufficient to support accurate pitch perception.

This letter is concerned with the possible contribution of TFS information to the lower F0DLs found by Bernstein and Oxenham (2008) in the F0-shifted condition. It has been proposed that, for complex tones containing harmonics that are too high to be resolved, but with some harmonics below about the 14th, the pitch may be extracted from the time intervals between prominent peaks in the TFS of the waveform evoked on the basilar membrane (de Boer, 1956;

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

Schouten *et al.*, 1962; Moore and Moore, 2003; Moore *et al.*, 2006, 2009; Moore and Sek, 2009a, 2009b). For harmonic complex tones, there is usually one major envelope peak per period, and most nerve spikes are evoked by peaks in the TFS close to the envelope maxima. Thus, the period of the sound is represented in the inter-spike intervals. This is described in more detail later on (see Fig. 1). One way of quantifying the available time-interval information is by calculation of the autocorrelation function (ACF); usually the largest peak in the ACF (apart from the one at a lag of zero) corresponds to 1/F0. The inter-spike intervals evoked in a small group of neurons with similar CFs convey ambiguous information, since intervals other than 1/F0 are present (Moore and Sek, 2009a); this is revealed by the presence of multiple peaks in the ACF. Correspondingly, sounds with only a few high harmonics have an ambiguous pitch (de Boer, 1956; Schouten *et al.*, 1962). However, for broadband sounds, the ambiguity may be resolved by combining information across neurons with different CFs; the only time interval that is common across CFs is the one corresponding to 1/F0 (Moore, 1982). One way of combining information is by summing ACFs across CFs, giving the summary ACF or SACF (Meddis and O'Mard, 1997).

Bernstein and Oxenham (2008, Fig. 8) simulated the waveform evoked at a place on the basilar membrane with a CF of 1500 Hz by passing the stimuli through a gammatone filter (Patterson *et al.*, 1995) centered at 1500 Hz, using an F0 of 125 Hz. They pointed out that the presence of multiple TFS peaks of similar amplitude close to an envelope peak should lead to a less precise estimate of pitch than when there are only one or a very few TFS peaks close to each envelope peak; this is linked to the inherent ambiguity of the TFS information (Moore and Sek, 2009a). The waveform samples in their Fig. 8 showed a greater number of TFS peaks near the envelope peak for the F0-shifted than for the harmonic stimulus. This led Bernstein and Oxenham (2008) to conclude that the waveform samples did not appear to be consistent with an explanation based on TFS, when considering the output of a single auditory filter. Here, we describe how, nevertheless, their results are compatible with an explanation based on TFS, even when considering the output of a single auditory filter.

II. THE POSSIBLE ROLE OF TFS IN THE BENEFIT OBTAINED FROM THE F0 SHIFT

To illustrate the information that was potentially available in the TFS, we used a simple functional model of peripheral auditory processing. The stimuli used by Bernstein and Oxenham (2008) were generated. The overall duration of each stimulus was 300 ms, including 30-ms raised-cosine rise/fall ramps. The stimuli were passed through three fourth-order gammatone filters, which were distributed across the passband of the stimuli, and were centered at 1750, 2200, and 2800 Hz. The sampling rate used was 100 kHz, chosen to allow high precision in defining the times of peaks in the TFS. We consider here only the case of F0=125 Hz, which led to a large difference in F0DLs for the harmonic and F0-shifted stimuli. It was assumed that each peak in the TFS at the output of a given filter had the potential for evoking a nerve spike, but that the probability of evoking a spike increased with increasing amplitude of the TFS peak. Specifically, we assumed that the probability of obtaining an inter-spike interval corresponding to the time between two given TFS peaks was proportional to the product of the amplitudes of those two peaks.

For each gammatone filter, histograms were constructed of the inter-spike intervals simulated in this way, using a bin width of 0.1 ms. The analysis was restricted to times between 50 and 250 ms following the onset of the stimuli to reduce transient effects relating to the on and off ramps. All time intervals below 10 ms were included. Each histogram was based on an average of the inter-spike intervals evoked by five different samples of the stimulus (each with a different selection of random component starting phases). The histograms do not reflect the jitter that occurs in the timing of the spikes. Also, it should be noted that this method of simulation results in an over-representation of short inter-spike intervals (less than 2 ms) since refractory effects are not taken into account. The shorter intervals reflect the times between successive or alternate cycles of the temporal fine structure produced by the interaction of harmonics at the output of the auditory filters. These shorter intervals depend strongly on the center frequency of the filter, and do not convey very effective information about F0. In what follows, we focus on intervals close to 4 or 8 ms, since these showed changes when F0 was changed that were the same (in percentage terms) as the changes in F0.

The left-hand column of Fig. 1 shows the histogram derived from the output of each of the gammatone filters and the summed histogram across filters (bottom) for the harmonic stimulus; the information displayed here is similar to that obtained from a SACF. There was a prominent inter-spike interval at 8 ms for all gammatone filter CFs, as expected. However, there were also inter-spike intervals at nearby values. For example, for the gammatone filter centered at 1750 Hz, there were intervals close to 7.4 and 8.6 ms. The detrimental effect of these nearby intervals probably depends on two factors:

(1) the relative heights of neighboring peaks in the TFS, which depend partly on the bandwidth of the auditory filter and of the stimulus, and
(2) the spacing in time of adjacent peaks in the TFS at the output of the auditory filter relative to the period of the sound, which decreases as CF/F0 increases.

The effect of factor 1 is reflected in the relative heights of the peaks in the histograms in Fig. 1; as the CF gets higher, the peaks adjacent to the peak at 8 ms become closer in magnitude to the peak at 8 ms. The effect of factor 2 can be seen by the peaks adjacent to 8 ms becoming closer in time to the peak at 8 ms as the CF increases. For the filter centered at 1750 Hz, the "competing" intervals differ from the "target" inter-spike interval by factors of 1.08 and 0.93. The auditory system may have difficulty distinguishing inter-spike intervals that are very close because of jitter in the timing of nerve spikes and limitations in the precision with which central mechanisms can measure the time intervals; it seems plausible to assume that the typical error ε in the

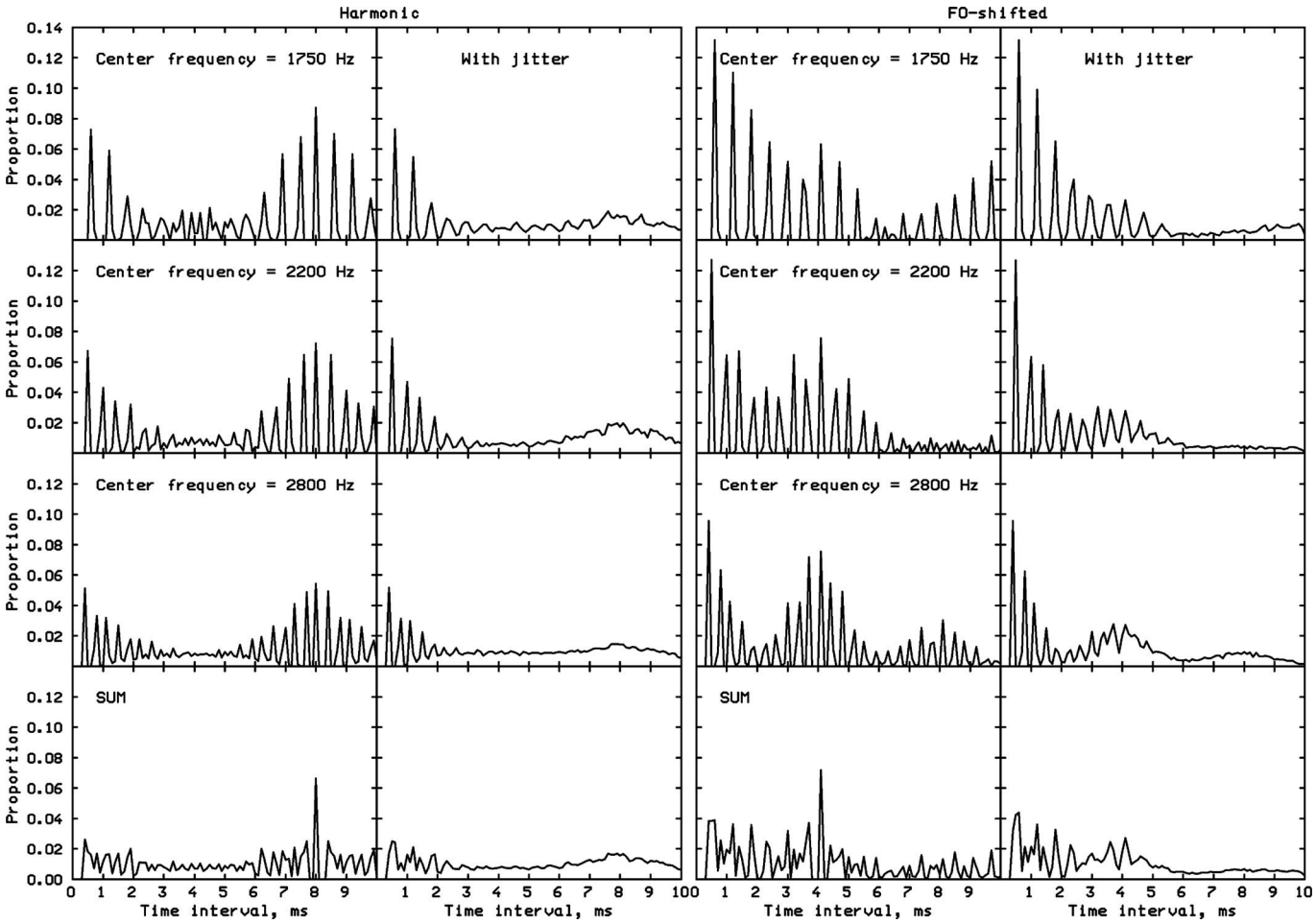


FIG. 1. The first and third columns show histograms of the intervals between simulated nerve spikes evoked by peaks in the TFS of the outputs of gammatone filters centered (from the top down) at 1750, 2200, and 2800 Hz, for the harmonic stimulus with F0=125 Hz (column 1) and the F0-shifted stimulus (column 3). The bottom panel shows the summed histogram. All histograms are normalized so that the summed probability of all intervals between 0 and 10 ms is equal to 1. The second and fourth columns show the effect of introducing a temporal jitter to the intervals with standard deviation proportional to the size of each interval, with constant of proportionality=0.025.

estimation of an interval by central mechanisms is proportional to the size of the interval I (Moore *et al.*, 2006, 2009). If ε were, say, $0.03I$, then the intervals of 7.4, 8, and 8.6 ms could not be distinguished reliably. However, the summed histogram (bottom panel) shows a distinct peak at 8 ms because the summed histogram does not reflect limitations of within-channel interval analysis (or limitations imposed by jitter in the timing of spike initiation). The auditory system might also have difficulty in measuring the interval of 8 ms accurately because of some CF-specific limitation, as described earlier (de Cheveigné and Pressnitzer, 2006) and as assumed in the model of Bernstein and Oxenham (2005).

To simulate the effect of limited precision in the estimation of inter-spike intervals, the intervals used to construct column 1 of Fig. 1 were subjected to a random perturbation with a Gaussian distribution and a standard deviation ε proportional to the size of the interval. A new histogram was then constructed from the summed perturbed intervals. Several values of ε were tried, and the second column of Fig. 1 illustrates the results with $\varepsilon=0.025I$. With this value of ε, there is no longer a sharp peak in the histograms at 8 ms. Instead the individual peaks in the histograms are smeared, leading to a broad peak around 8 ms. In the summed histogram, there are no very distinct peaks. This could account for the relatively poor F0 discrimination observed for the harmonic stimulus.

The third column of Fig. 1 shows the histogram derived from the output of each of the gammatone filters and the summed histogram across filters (bottom) for the F0-shifted stimulus. Crucially, the histograms did not show a prominent inter-spike interval at 8 ms. However, for each gammatone filter, there was a prominent interval close to 4 ms (actually at about 4.1 ms). This is consistent with the finding of Bernstein and Oxenham (2008) that the pitch of the F0-shifted stimulus with F0=125 Hz was shifted upwards by about 1 octave, relative to the pitch of the harmonic stimulus. As for the harmonic stimulus, there were other prominent intervals at nearby values. For example, for the gammatone filter centered at 1750 Hz, there were intervals close to 3.4 and 4.7 ms. However, expressed in relative terms, i.e., as the ratio of the intervals to be discriminated, the competing intervals were further removed from the target interval than was the case for the harmonic stimuli. The ratios of the target and competing intervals were 1.20 and 0.87. Using the same example as above, if ε were $0.03I$, then the intervals of 3.4, 4.1, and 4.7 ms could be distinguished reliably.

The fourth column of Fig. 1 shows the effect of introducing a jitter to the inter-spike intervals to simulate the effects of limited precision in measuring the intervals, in the same way as was done earlier. The value of ε was again $0.025I$. The individual peaks in the histogram became somewhat broadened, as expected, but there were still distinct peaks corresponding to intervals in the temporal fine structure. Thus, the central auditory system may be able to distinguish the target and competing intervals in this case. This could account for the smaller F0DLs found by Bernstein and Oxenham (2008) in their F0-shifted condition with F0 = 125 Hz. Such an explanation would not require the combination of information across channels; the F0-discrimination task could be performed via discrimination of the time intervals in the TFS regardless of which intervals were compared. For example, if the difference in F0 between the two-halves of a trial was 2%, so that the F0 was 125 Hz in one interval of a trial and 127.5 Hz in the other, then the task might be performed by discriminating the interval of 3.4 ms (for F0 = 125 Hz) from the interval of 3.33 ms (for F0 = 127.5 Hz) or the interval of 4.1 ms from the interval of 3.92 ms.

The results shown in Fig. 1 are broadly consistent with the modeling results presented by Bernstein and Oxenham (2008), which were based on the autocorrelation model of Meddis and O'Mard (1997), as modified by Bernstein and Oxenham (2005). Bernstein and Oxenham (2008) incorporated an *ad hoc* weighting function in each channel to limit the range of lags that could be used in that channel. This was similar in its functional consequences to the temporal jitter used in this letter. The weighted SACFs presented by Bernstein and Oxenham (2008) showed a small peak at 1/F0 for the harmonic stimuli, and a larger peak close to 0.5/F0 for the F0-shifted stimuli. Presumably, these peaks occurred because of information carried in the TFS of the outputs of the simulated auditory filters used in the model. Whatever approach is used, it may be important to consider the usefulness of the TFS information that can be extracted within each auditory channel, rather than simply considering the summary autocorrelation. The 4-ms inter-spike interval may be estimated more accurately than the 8-ms interval (for CFs in the range 1750–2800 Hz) because of limitations intrinsic to each channel. Since intervals close to 4 ms occurred for the F0-shifted stimulus but not for the harmonic stimulus, such within-channel limitations could explain the better F0 discrimination for the former without the need for across-channel combination of information. Nevertheless, as described earlier, combination of information across channels can help to resolve ambiguities about which inter-spike interval corresponds to the F0, since that is the only interval that is common across channels. Hence, in the summed histogram for the F0-shifted stimulus (bottom panel of column 3 of Fig. 1), the interval close to 4 ms (corresponding approximately to the F0 of the nonshifted even harmonics) is clearly represented more often than any other interval.

In conclusion, the simulations presented here suggest that the smaller F0DLs for the F0-shifted than for the harmonic stimuli, for F0s around 125 Hz, can be explained in terms of changes in the TFS evoked by the stimuli on the basilar membrane, in conjunction with the idea that TFS information is used less effectively when the inter-spike intervals within a channel are closely spaced or when the interval corresponding to the F0 is long relative to the reciprocal of the CF of the channel.

ACKNOWLEDGMENTS

This work was supported by the MRC (UK). We thank Kathryn Hopkins, Joshua Bernstein, Andrew Oxenham, and an anonymous reviewer for helpful comments on an earlier version of this paper.

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.

Bernstein, J. G., and Oxenham, A. J. (**2005**). "An autocorrelation model with place dependence to account for the effect of harmonic number on fundamental frequency discrimination," J. Acoust. Soc. Am. **117**, 3816–3831.

Bernstein, J. G., and Oxenham, A. J. (**2008**). "Harmonic segregation through mistuning can improve fundamental frequency discrimination," J. Acoust. Soc. Am. **124**, 1653–1667.

Carlyon, R. P., and Shackleton, T. M. (**1994**). "Comparing the fundamental frequencies of resolved and unresolved harmonics: Evidence for two pitch mechanisms?," J. Acoust. Soc. Am. **95**, 3541–3554.

de Boer, E. (**1956**). "Pitch of inharmonic signals," Nature (London) **178**, 535–536.

de Cheveigné, A., and Pressnitzer, D. (**2006**). "The case of the missing delay lines: Synthetic delays obtained by cross-channel phase interaction," J. Acoust. Soc. Am. **119**, 3908–3918.

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.

Meddis, R., and O'Mard, L. (**1997**). "A unitary model of pitch perception," J. Acoust. Soc. Am. **102**, 1811–1820.

Moore, B. C. J. (**1982**). *An Introduction to the Psychology of Hearing*, 2nd ed. (Academic, London).

Moore, B. C. J., and Glasberg, B. R. (**1988**). "Effects of the relative phase of the components on the pitch discrimination of complex tones by subjects with unilateral cochlear impairments," in *Basic Issues in Hearing*, edited by H. Duifhuis, H. Wit, and J. Horst (Academic, London).

Moore, B. C. J., Glasberg, B. R., Flanagan, H. J., and Adams, J. (**2006**). "Frequency discrimination of complex tones; assessing the role of component resolvability and temporal fine structure," J. Acoust. Soc. Am. **119**, 480–490.

Moore, B. C. J., Hopkins, K., and Cuthbertson, S. J. (**2009**). "Discrimination of complex tones with unresolved components using temporal fine structure information," J. Acoust. Soc. Am. **125**, 3214–3222.

Moore, B. C. J., and Sek, A. (**2009a**). "Development of a fast method for determining sensitivity to temporal fine structure," Int. J. Audiol. **48**, 161–171.

Moore, B. C. J., and Sek, A. (**2009b**). "Sensitivity of the human auditory system to temporal fine structure at high frequencies," J. Acoust. Soc. Am. **125**, 3186–3193.

Moore, G. A., and Moore, B. C. J. (**2003**). "Perception of the low pitch of frequency-shifted complexes," J. Acoust. Soc. Am. **113**, 977–985.

Patterson, R. D., Allerhand, M. H., and Giguère, C. (**1995**). "Time-domain modeling of peripheral auditory processing: A modular architecture and a software platform," J. Acoust. Soc. Am. **98**, 1890–1894.

Schouten, J. F., Ritsma, R. J., and Cardozo, B. L. (**1962**). "Pitch of the residue," J. Acoust. Soc. Am. **34**, 1418–1424.

The effect of hearing-aid compression on judgments of relative distance (L)[a)]

Michael A. Akeroyd[b)]
MRC Institute of Hearing Research (Scottish Section), Glasgow Royal Infirmary, Alexandra Parade, Glasgow G31 2ER, United Kingdom

(Received 19 December 2008; revised 15 October 2009; accepted 17 October 2009)

The overall level of a sound is an important auditory cue to distance in rooms, but this cue might be affected adversely by the amplitude compression found in most modern hearing aids because this explicitly changes levels. This prediction was tested using a synthetic-distance design to measure the just-noticeable difference (JND) in distance from distances of 2 or 5 m. Twenty-six aided listeners participated. The results did not show any effect of compression ratio upon JNDs. A possible interpretation is that the listeners had acclimatized to the effect their aids have on level.
 [DOI: 10.1121/1.3268505]

PACS number(s): 43.66.Ts, 43.66.Sr, 43.66.Pn [BCM] Pages: 9–12

I. INTRODUCTION

The auditory perception of the distance of sound sources in rooms is based mainly on level [for reviews, see Coleman (1963), Blauert (1997), and Zahorik *et al.* (2005)]. One cue is the overall level of a sound: the farther away a source is, the less intense is the sound that arrives at a listener. In an anechoic environment, the rate of reduction in level is 6 dB per doubling in distance (i.e., an inverse-square law), but for other environments the rate is somewhat less. A second cue, but one which can only be used when there is a substantial amount of reflections or reverberation (i.e., echoic environments), is the ratio of the level of the first-arriving sound to the level of the remainder; the larger this is, the closer is the sound source [see Zahorik (2002b) for a discussion of how this ratio could be measured]. There are two other acoustic cues to distance, although of less relevance to sources in most rooms as they only affect particularly distant or close sources. If a source is farther-than about 15 m, then absorption by the air affects high frequencies more than low frequencies (Blauert, 1997), while if the source is closer-than about 1 m, then the interaural time and level differences that indicate direction can be affected (Brungart *et al.*, 1999).

As self-report data have shown links between difficulties in distance perception and auditory disability (Gatehouse and Noble, 2004), Akeroyd *et al.* (2007) developed an experimental paradigm to study judgments of relative distance by hearing-impaired listeners. The paradigm used the "room-image" method (Allen and Berkeley, 1979) to calculate the angle of arrival, time of arrival, and level at arrival of the direct sound and the first 74 reflections at a listener placed in a virtual 7×9 m room. Each of these 75 sounds was then presented using a circular array of loudspeakers, at the required level, required time, and from the loudspeaker closest to the required angle (cf. Hafter and Seeber, 2004). The result is a recreation of what would have been heard in the room. The system gives a vivid percept of distance, with the sound appearing to come from farther away than the actual loudspeakers.[1] Akeroyd *et al.* (2007) measured just-noticeable differences (JNDs) in distance in four tasks, using two reference distances (2 and 5 m) crossed with two comparison directions (either closer-than or farther-than the reference). These JNDs were measured under conditions termed "normal level," in which the overall level variation due to the inverse-square law was included, and "equalized level," in which this variation was removed. The stimuli were single sentences. For the "normal-level" conditions, it was found that the JNDs were largest for the farther-than-5-m task and smallest for the closer-than-2-m task, consistent with the typical result that thresholds for distance discrimination generally increase with farther reference distances (Zahorik *et al.*, 2005). In general, the results for the hearing-impaired listeners did not differ from those of the normal-hearing listeners, excepting (1) in the farther-than-5-m task the group of hearing-impaired listeners performed more poorly than an age-matched normal-hearing group and (2) everyone with a hearing-loss of more than about 25 dB performed at chance in the equalized-level conditions.

Level is crucial to distance perception, but the deliberate purpose of a hearing-aid compressor is to control level (for a review, see Moore, 2008). An ideal fast-acting compressor will reduce the instantaneous fluctuations in level inherent to speech or dynamically changing backgrounds, and so can improve the audibility of the weaker parts of a speech signal (e.g., Dillon, 2001). An ideal slow-acting compressor will instead act as an automatic volume control, adjusting the gain when the listener moves from one environment to another. Compression creates a nonlinear relationship between the actual level of a sound and that received by the listener via the hearing aid, thus alleviating some of the effects of loudness recruitment that would be not compensated by a linear (i.e., noncompressive) hearing aid. But compression should therefore adversely affect distance perception, as it

[a)]Portions of this work were presented in "The effect of hearing aids on distance perception," 155th Meeting of the Acoustical Society of America, Paris, June–July 2008.
[b)]Author to whom correspondence should be addressed. Electronic mail: maa@ihr.gla.ac.uk

should interfere with the levels of sounds upon which distance perception is based. Thus, it was felt valuable to study experimentally the effects of hearing-aid compression on judgments of relative distance. The experiment reported here repeated the normal-level conditions of Akeroyd *et al.* (2007), using a sample of 26 aided listeners from our research panel. Each listener wore their own aids: the sample included listeners with two linear aids, or two compressive aids, or one of each, and with a variety of gains and compression settings.

II. EXPERIMENTAL METHOD

The method was the same as that used by Akeroyd *et al.* (2007). Briefly, the room-image method (Allen and Berkeley, 1979) was used to calculate the azimuths, time delays, and level differences of the direct path and the first 74 echoes in a virtual 7×9 m room. The virtual listener was placed near one corner and the sources were set to be at distances from it of 1–8 m, along a diagonal of the room. These calculations were then used to determine the choice of actual loudspeaker, time delay, and attenuation used to present a stimulus and 74 copies of it. The loudspeakers were arranged in a circle, at 15° intervals, with the listener at the center, and were placed in a room that was 2.5-m wide×4.4-m long ×2.5-m tall. The room was not anechoic, but its walls were acoustically treated with sound-absorbing foam to reduce its reverberation time to an average across frequency (250–8000 Hz) of about 80 ms. The average *A*-weighted background level was about 20 dB [for further details see Akeroyd *et al.* (2007), especially Figs. 1 and 2].

A two-interval, two-alternative forced-choice method was used to determine psychometric functions for judgments of relative distance. Forty-eight trials per point were run for each psychometric function. In the reference interval, the simulated distance was either 2 or 5 m. In the comparison interval the simulated distance was either 1.00, 1.33, 1.67, 2.00, 2.33, 2.67, or 3.00 m (for the 2-m reference), or 2, 3, 4, 5, 6, 7, or 8 m (for the 5-m reference). The order of the intervals was randomized across trials, and they were separated by an inter-stimulus interval of 1200 ms [Akeroyd *et al.*, 2007) used 500 ms; the change was made here in order to give the hearing-aid compressors longer to stabilize between stimuli]. JNDs were measured for four conditions, corresponding to changes in distance closer-than a reference of 2 m ("2-closer"), farther-than 2 m ("2-farther"), closer-than 5 m ("5-closer"), and farther-than 5 m ("5-farther") (unless otherwise stated, all the statistics for the data reported below are listed as quadruplets, in the order 2-closer/2-farther/5-closer/5-farther). The JNDs were found from fitting psychometric functions to the data and then finding the change in distance corresponding to a d' of 1.0. The psychometric function was based on the assumption that the detectability of a change in distance (d') was proportional to the value of the change. The stimuli in the two intervals were short spoken sentences, one spoken by a female and the other by a male; the gender of the sentences used for the reference and comparison distances was randomized across trials (the sentences were chosen randomly from the "BKB" and "ASL" corpora; Bench and Bamford, 1979; Macleod and Summerfield, 1990). Listeners were required to decide which of the sentences was farther away. No feedback was given. The level of the stimuli was set so that a very long sequence of ASL sentences set to a synthesized distance of 2 m gave a long-term average level of 65 dB(A) at a microphone at the center of the loudspeaker ring.

Twenty-six listeners participated, aged 45–78 years (mean=65 years), with better-ear average hearing-losses (average of pure-tone thresholds at 500, 1000, 2000, and 4000 Hz, termed "BEA") of 28–84 dB (mean=57 dB) and worse-ear losses of 33–98 dB (mean=66 dB). They were chosen from the pool available to the Institute of Hearing Research, sourced from postal surveys, attendees at local clinics, and occasional other sources.

All the listeners were experienced hearing-aid users and wore their own aids that had been fitted by the local audiological services. The aids came from a variety of manufacturers: Unitron, Siemens, Oticon, Phonak, Puretone, and Microtech. The compression parameters of the hearing aids were derived from measurements of their output levels for input levels from 55 to 75 dB sound pressure level (SPL) at frequencies of 500, 1000, 2000, and 4000 Hz, which were collected using a hearing-aid analyzer (Siemens Unity 2).[2] It was convenient to have a simple overall classification of whether an aid was linear or compressive: if the measured compression ratio between 60 and 65 dB SPL, at 1000 Hz, was greater than 1.25, then the aid was considered as compressive. Twenty-eight of the aids were so classified: for these, the median compression ratios at 250, 500, 1000, 2000, and 4000 Hz were 1.02, 1.3, 1.8, 1.8, and 1.6. 9 listeners wore two linear aids (termed "linear/linear"), 6 listeners wore one linear aid and one compressive aid ("linear/comp"), and 11 listeners wore two compressive aids ("comp/comp").

III. RESULTS

Figure 1 shows the JNDs, in meters, for the judgments of distance. The open circles plot the data from the linear/linear listeners, the diamonds plot the data from the linear/comp listeners, and the hourglasses plot the data from the comp/comp listeners. The dots re-plot the JNDs measured using 77 *unaided* listeners from Akeroyd *et al.* (2007). Each line shows an exponential curve fitted to the data for the unaided listeners (i.e., JND$\propto e^{cH}$, where H is the hearing-loss and c is a constant): the lines were calculated over BEAs from 20 to 59 dB—the largest hearing-loss included in Akeroyd *et al.* (2007)—and then extrapolated to the larger hearing-losses of the present aided listeners. The scatter in the data points for the aided listeners is comparable to that found for unaided listeners, and there was no obvious effect of hearing-aid compression: none of the aided groups yielded thresholds substantially above the extrapolation from the unaided listeners. This was demonstrated by calculating the difference between the measured JND for each aided listener and the predicted JND given the curve-fit to the data for the unaided listeners, and then comparing, by one-sample *t*-test, the across-listener means of the differences from zero; all

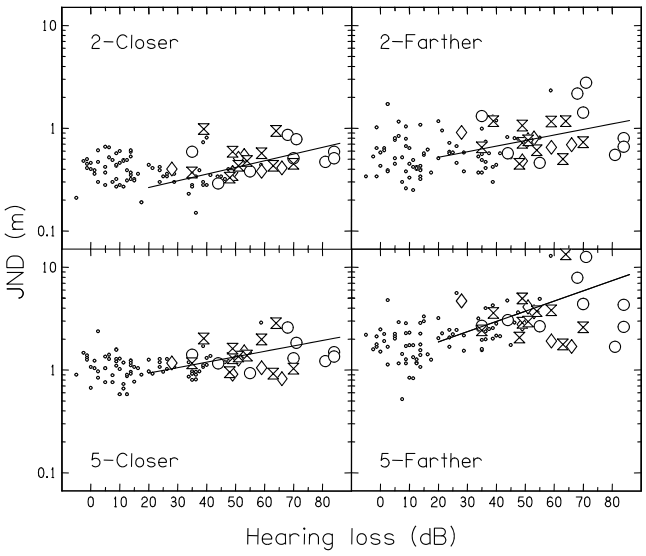


FIG. 1. The JNDs, in meters, plotted as a function of the better-ear hearing-loss for listeners fitted with two linear aids (open circles), one linear and one compressive aid (diamonds), and two compressive aids (hourglasses). The dots plot the JNDs measured using 77 unaided listeners from Akeroyd *et al.* (2007), and the lines show exponential curves fitted to these data. The largest hearing-loss for the unaided listeners was 59 dB, and so the lines have been extrapolated above that value to the higher hearing-losses of the aided listeners. The four panels are for the four tasks (see text).

four were nonsignificant [$t(25)=1.3, 0.79, -0.71, -0.94$; $df=25$; $p>0.05$].

A set of four one-way analyses of variance showed no significant effect of compression, in that the differences between the three groups of data for aided listeners—i.e., linear/linear vs linear/comp vs comp/comp—were not statistically significant for any of the JNDs [$F(2,23)=0.83, 1.91, 1.32, 0.46$, all $p>0.05$]. A correlation analysis confirmed this: a large number of correlations were calculated between each of the four JNDs and some measure of compression ratio, but all were found to be nonsignificant.[3] Figure 2 illustrates one of these analyses: the four panels show the four JNDs as a function of the 60–65 dB compression ratio, averaged across 250, 500, 1000, 2000, and 4000 Hz and then across ears. The lack of any effect of compression ratio is clear.

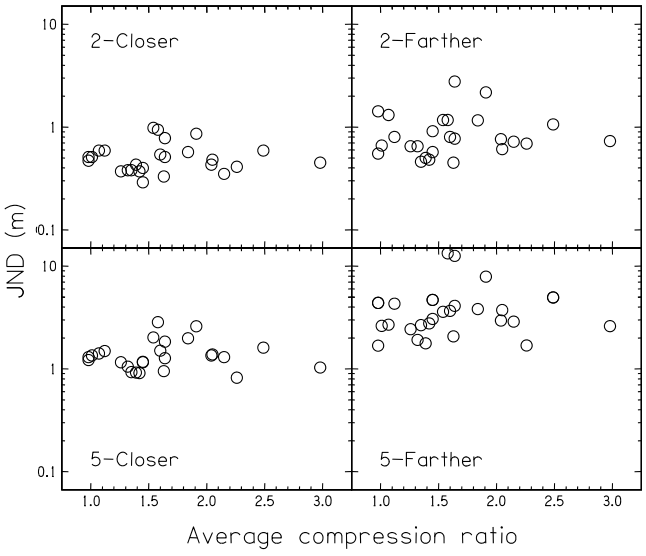


FIG. 2. The JNDs plotted as a function of the average compression ratio. The four panels are for the four tasks.

IV. DISCUSSION

This experiment did not demonstrate any substantial or reliable effect of hearing-aid compression on judgments of relative distance. This was somewhat unexpected, as the cues to auditory distance are based on the levels of sounds and the purpose of hearing-aid compression is to modify level.

The expectation is particularly clear where the inverse-square law dominates, and the overall level reduces at a rate of 6 dB per doubling of distance. The formal relationship for the level difference ΔL between distances d_1 and d_2 is thus $\Delta L = 6 \log_2 d_2 - 6 \log_2 d_1$. But, by the definition of compression ratio, a change in the output of a hearing aid of 1 dB requires a C dB change in its input, where C is the value of the compression ratio. If the JND for distance is determined by the value of ΔL received by the listener—i.e., *after* compression—then it depends on ΔL before compression and the reference distance d_1 thus

$$\mathrm{JND}_C = (2^{C\Delta L/6} - 1) d_1. \quad (1)$$

where the JND is taken as the difference, in meters, between d_1 and d_2. The ratio of the JND with compression (JND_C) to the "normal" JND without compression (JND_N) is therefore

$$\frac{\mathrm{JND}_C}{\mathrm{JND}_N} = \frac{2^{C\Delta L/6} - 1}{2^{\Delta L/6} - 1}. \quad (2)$$

For example, if the limiting value of ΔL is 2 dB and the compression ratio is 3, then the JND will be 3.8× its non-compressive value. For small values of ΔL and compression ratio, Eq. (2) can be approximated by the power function, $\mathrm{JND}_C/\mathrm{JND}_N \approx C^p$, where the exponent p equals 1.25 for a ΔL of 2 dB and a range of C of 1–4.[4] Thus, the JND with compression should depend strongly on the value of the compression ratio.

At least three reasons can be offered as to why this effect was not observed. It may be the case that hearing aids have their effects in the overall scale of the perceived auditory world rather than in relative judgments within that world. Given that in normal hearing the perceptual scale of auditory distance is a compressive function of physical distance, with listeners generally underestimating how far away a sound actually is [summarized by Zahorik (2002a) and Zahorik *et al.* (2005)], it is possible that a compressive hearing aid may exacerbate this by further reducing the scale of the auditory world. This could be tested experimentally by including judgments of absolute distance in the design; it would be expected that the absolute distance of sounds would be judged to be closer with hearing aids than without. A second possible reason is that the listeners—who were all experienced hearing-aid users—were acclimatized to the effects of their own hearing aids on intensity and so had learned to overcome the effect of compression [e.g., Robinson and Gatehouse (1996); for a review, see Munro (2008)]. If so, it would be of value to study first-time, unacclima-

tized, users, especially as any immediate adverse effects on judgments of distance may be influential in the self-reports used by the audiologist to fine-tune the hearing-aid. A third reason is that speech has rapid, large dynamic fluctuations in level that may interfere with any judgments based on overall level. But a fast compressor will reduce these fluctuations—though they would only disappear to zero if the compression was instantaneous and the compression ratio was infinite—and so might lead to an improvement in performance via a reduction in interference (and always assuming that the effect was large enough to dominate the overall level change due to the compression). This was examined informally using measurements of the attack and release time that were taken at the same time as the input/output measurements (see Sec. II), but none of the correlations between either time constant and either the four JNDs or their means was statistically significant. Nevertheless, the question of the effect of speed of compression on intensity resolution is somewhat neglected and deserves future study.[5]

The wider issue underlying these analyses is the assumption that the ability to detect changes in relative distance was due to the ability to detect changes in level: i.e., that performance is founded on intensity resolution. The effects of hearing impairment on intensity resolution are complex, and different experiments have found performance to be worse, better, or about the same as for normal-hearing controls [for a review, see Moore (2007)]. Most studies of intensity discrimination with hearing aids have used linear amplification, however (e.g., Robinson and Gatehouse, 1996), and none has used speech stimuli with hearing-impaired listeners. In principle, if the JND for intensity is x dB unaided, then it should be Cx dB with very fast compression. One might therefore expect adverse effects on intensity discrimination, which may, for an unacclimatized listener, be of consequence during fitting. It would be of interest to measure the effects of compression on the ability to discriminate changes in the level of free-field presented speech.

ACKNOWLEDGMENTS

Stuart Gatehouse contributed substantially to the experimental design before his death in February 2007. I thank David McShefferty, Pat Howell, Helen Lawson, and Fiona Guy for their assistance in collecting the experimental data; Michael Stone and Owen Brimijoin for comments on an earlier draft; and Brian Moore (Associate Editor) and two anonymous reviewers for their recommendations during submission. The Scottish Section of IHR is supported by intramural funding from the Medical Research Council and the Chief Scientist Office of the Scottish Government.

[1]For a full discussion of the system, see Akeroyd *et al.* (2007). It is worth noting here (1) that the system uses loudspeakers, not headphones, and therefore avoids the problems due to individualization of head-related transfer functions that can affect virtual acoustics, but (2) the system is limited to presenting sounds in the horizontal plane, and (3) coloration of reflections is not (yet) implemented.

[2]Three other listeners also participated, but for them either no or unreliable compression data were obtained, and so their results are excluded.

[3]For these analyses any compression ratio found to be larger than 10 was reset to 10. This was done as any larger than 10 were felt to be adversely affected by experimental error—recall that they were derived from input/output functions over a 5-dB range—and therefore unreliably large. Almost 150 separate measures of compression were obtained, using compression ratios calculated from the hearing-aid input/output functions at 55–60, 60–65, or 70–75 dB SPL, taken either at individual octave frequencies from 250 to 4000 Hz or various averages of and either the least compressive across ears, the most compressive, or the average across ears. As a large number of correlations were conducted, a Bonferroni correction was applied to the α level for statistical significance. Even without this correction, however, only three of the correlations had a probability less than the conventional α of 0.05. Given such a small number—each of which involved a compression ratio measured at 3000 Hz but no other frequency—it is not clear that these occasional correlations offer any predictive insight or reliability.

[4]The exponent p depends slightly on both the level difference at threshold (ΔL) and how much the level changes per doubling in distance: e.g., for a ΔL of 1 dB and 6 dB per doubling, then p is 1.12, but for a ΔL of 1 and 4 dB per doubling then p is 1.18.

[5]The attack/release times used here were taken at 1000 Hz. The analyses used the data from 16 of the 17 people with either one or two hearing aids; if they had two, the left and right values were averaged; one listener was excluded as an outlier in both JND and release time.

Akeroyd, M. A., Gatehouse, S., and Blaschke, J. (**2007**). "The detection of differences in the cues to distance by elderly hearing-impaired listeners," J. Acoust. Soc. Am. **121**, 1077–1089.

Allen, J. B., and Berkley, D. A. (**1979**). "Image method for efficiently simulating small-room acoustics," J. Acoust. Soc. Am. **65**, 943–950.

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

Blauert, J. (**1997**). *Spatial Hearing: The Psychophysics of Human Sound Localization* (MIT, Cambridge, MA).

Brungart, D. S., Durlach, N. I., and Rabinowitz, W. M. (**1999**). "Auditory localization of nearby sources. II. Localization of a broadband source," J. Acoust. Soc. Am. **106**, 1956–1968.

Coleman, P. D. (**1963**). "An analysis of the cues to auditory depth perception in free space," Psychol. Bull. **60**, 302–315.

Dillon, H. (**2001**). *Hearing Aids* (Boomerang, Sydney).

Gatehouse, S., and Noble, W. (**2004**). "The Speech, spatial and qualities of hearing scale (SSQ)," Int. J. Audiol. **43**, 85–99.

Hafter, E., and Seeber, B. (**2004**). "The simulated open field environment for auditory localization research," in Proc. ICA 2004, 18th Int. Congress on Acoustics, Kyoto, Japan, pp. 3751–3754 (Int. Commission on Acoustics).

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

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

Moore, B. C. J. (**2008**). "The choice of compression speed in hearing aids: Theoretical and practical considerations and the role of individual differences," Trends Amplif. **12**, 103–112.

Munro, K. J. (**2008**). "Reorganization of the adult auditory system: Perceptual and physiological evidence from monaural fitting of hearing aids," Trends Amplif. **12**, 85–102.

Robinson, K., and Gatehouse, S. (**1996**). "The time course of effects of intensity discrimination following monaural fittings of hearing aids," J. Acoust. Soc. Am. **99**, 1255–1258.

Zahorik, P. (**2002a**). "Assessing auditory distance perception using virtual acoustics," J. Acoust. Soc. Am. **111**, 1832–1846.

Zahorik, P. (**2002b**). "Direct-to-reverberant energy ratio sensitivity," J. Acoust. Soc. Am. **112**, 2110–2117.

Zahorik, P., Brungart, D. S., and Bronkhorst, A. W. (**2005**). "Auditory distance perception in humans: A summary of past and present research," Acta. Acust. Acust. **91**, 409–420.

Seismoelectric reflection and transmission at a fluid/porous-medium interface

Menne Schakel[a] and David Smeulders
Department of Geotechnology, Delft University of Technology, 2600 GA Delft, The Netherlands

(Received 9 March 2009; revised 14 October 2009; accepted 14 October 2009)

The dispersion relation for seismoelectric wave propagation in poroelastic media is formulated in terms of effective densities comprising all viscous and electrokinetic coupling effects. Using Helmholtz decomposition, two seismoelectric conversion coefficients are derived, for an incident *P*-wave upon an interface between a compressible fluid and a poroelastic medium. These coefficients relate the incident *P*-wave to a reflected electromagnetic wave in the fluid, and a transmitted electromagnetic wave in the porous medium. The dependency on angle of incidence and frequency is computed. Using orthodox and interference fluxes, it is shown that energy conservation is satisfied. A sensitivity analysis indicates that electrolyte concentration, viscosity, and permeability highly influence seismoelectric conversion.
 [DOI: 10.1121/1.3263613]

PACS number(s): 43.20.El, 43.20.Gp, 43.40.Ph, 43.30.Ma [RR] Pages: 13–21

I. INTRODUCTION

When the grain surfaces of soils and rocks are in contact with a fluid electrolyte, they typically acquire a chemically bound surface charge that is balanced by mobile counterions in a thin fluid layer surrounding the grains. The bound charge is immobile, whereas the counterions can move. The distribution of mobile ions is determined by a balance between electrostatic forces and thermal diffusivity. At the interface between the immobile and counterions the so-called zeta-potential is defined. The potential varies exponentially when one moves away from the interface. The corresponding characteristic length scale is called the Debye length, which is on the order of some tens of nanometers for typical grain-electrolyte combinations (Pride, 1994). The ensemble of bound and mobile charge layers is referred to as the electrochemical double layer.

A compressional wave in fluid-saturated porous media will create pressure peaks and troughs on the scale of the wavelength. The resulting hydraulic flow will transport the counterions relative to the immobile, bound charge. In this way, counterions accumulate in pressure troughs and bound charge becomes exposed in pressure peaks, creating an electric coseismic field at the scale of the wavelength (Haines *et al.*, 2007). This electric field drives a conduction current that exactly balances the hydraulic current flow. Thus there is no net electric current for a propagating compressional wave within a homogeneous material. For interfaces, however, this is no longer the case. When a compressional or shear wave traverses an interface with a contrast in electrical or mechanical properties, an electric current imbalance is produced resulting in electromagnetic (EM)-waves that can propagate outside the support of the seismic waves (Haartsen and Pride, 1997). The generation of these waves at interfaces can thus be associated with new, electrokinetic, reflection, and transmission coefficients.

Several studies measured the conversion of seismic to electromagnetic energy in the field (Thompson and Gist, 1993; Butler *et al.*, 1996; Mikhailov *et al.*, 1997; Garambois and Dietrich, 2001; Haines *et al.*, 2007). The inverse electroseismic conversion was also measured (Thompson *et al.*, 2007; Zhu *et al.*, 2008). So-called electroseismograms in layered media were presented by several authors (e.g., Haartsen and Pride, 1997; Han and Wang, 2001; White and Zhou, 2006; Haines and Pride, 2006). In order to obtain the electrokinetic reflection and transmission coefficients at a fluid/porous-medium interface Block and Harris (2006) approximated Pride's (1994) electrokinetic theory by solving the Biot reflection problem only, from which the relative pore fluid flow is obtained. This flow subsequently acts as a source term in the EM boundary conditions of Pride's (1994) theory. In this paper, we solve the complete boundary value problem in a procedure similar to Denneman *et al.* (2002), where an acoustic wave from a compressible fluid impinges on a poroelastic medium. Two new parameters are defined, describing the conversion from seismic to EM-waves. The incident fluid wave is transmitted into fast and slow compressional waves, a shear wave, and into an EM-wave. Part of the incident wave is reflected as a fluid wave and a fluid EM-wave. In Sec. II we introduce the governing equations and we present the electrokinetic dispersion relations. The dependency of the seismoelectric reflection and transmission coefficients on angle of incidence and frequency is derived in Sec. III. In Sec. IV we recast the reflection and transmission coefficients in vertical energy flux coefficients and perform a sensitivity analysis.

II. GOVERNING EQUATIONS

We consider a poroelastic matrix saturated by an electrolyte and adopt an $\exp[i\omega t]$ convention for time varying

[a]Author to whom correspondence should be addressed. Electronic mail: m.d.schakel@tudelft.nl

fields. The governing linearized equations in an isotropic, homogeneous poroelastic medium were given by Pride and Haartsen (1996). We rewrite the equations here in a somewhat different form because we aim to derive the dispersion relations in terms of effective densities. We will confirm full consistency with the original Pride equations in the forthcoming. The governing equations read as

$$-\nabla\cdot\hat{\underline{\underline{\boldsymbol{\sigma}}}}-(1-\phi)\nabla\hat{p}=-\omega^2[\rho_{11}(\omega)\hat{\mathbf{u}}+\rho_{12}(\omega)\hat{\mathbf{U}}]+\frac{\eta\phi L(\omega)}{k(\omega)}\hat{\mathbf{E}}, \tag{1}$$

$$-\phi\nabla\hat{p}=-\omega^2[\rho_{12}(\omega)\hat{\mathbf{u}}+\rho_{22}(\omega)\hat{\mathbf{U}}]-\frac{\eta\phi L(\omega)}{k(\omega)}\hat{\mathbf{E}}, \tag{2}$$

$$\hat{\underline{\underline{\boldsymbol{\sigma}}}}=-G[\nabla\hat{\mathbf{u}}+(\nabla\hat{\mathbf{u}})^T]-\left[\left(A-Q\frac{1-\phi}{\phi}\right)\nabla\cdot\hat{\mathbf{u}}+\left(Q-R\frac{1-\phi}{\phi}\right)\nabla\cdot\hat{\mathbf{U}}\right]\underline{\underline{\mathbf{I}}}, \tag{3}$$

$$\hat{p}=-\frac{1}{\phi}(Q\nabla\cdot\hat{\mathbf{u}}+R\nabla\cdot\hat{\mathbf{U}}), \tag{4}$$

$$\hat{\mathbf{J}}=\sigma(\omega)\hat{\mathbf{E}}+L(\omega)(-\nabla\hat{p}+\omega^2\rho_f\hat{\mathbf{u}}), \tag{5}$$

$$\nabla\times\hat{\mathbf{H}}=i\omega\varepsilon\hat{\mathbf{E}}+\hat{\mathbf{J}}, \tag{6}$$

$$\nabla\times\hat{\mathbf{E}}=-i\omega\mu\hat{\mathbf{H}}. \tag{7}$$

Hats over field variables indicate small frequency-domain quantities, ω is the angular frequency, η is the pore fluid viscosity, ϕ is the porosity, ρ_f is the pore fluid density, μ is the magnetic permeability, ε is the bulk electrical permittivity given by $\varepsilon=\varepsilon_0[\phi(\epsilon_f-\epsilon_s)/\alpha_\infty+\epsilon_s]$, with solid and pore fluid relative permittivities ϵ_s and ϵ_f, and vacuum permittivity ε_0. $\underline{\underline{\mathbf{I}}}$ denotes the identity matrix. The field variables are solid and (pore) fluid displacements $\mathbf{u}$ and $\mathbf{U}$, intergranular stress $\underline{\underline{\boldsymbol{\sigma}}}$, pore pressure p, electric current density $\mathbf{J}$, and the electric and magnetic fields $\mathbf{E}$ and $\mathbf{H}$. A, Q, and R are the generalized elastic coefficients that are related to the bulk modulus of the skeleton grains K_s, the shear modulus G, the bulk modulus of the pore fluid K_f, and the bulk modulus of the framework of grains K_b as follows (Biot and Willis, 1957):

$$A=\frac{(1-\phi)^2K_sK_f-(1-\phi)K_bK_f+\phi K_sK_b}{K_f\left(1-\phi-\frac{K_b}{K_s}\right)+\phi K_s}-\frac{2}{3}G, \tag{8}$$

$$Q=\frac{\phi[K_s(1-\phi)-K_b]K_f}{K_f\left(1-\phi-\frac{K_b}{K_s}\right)+\phi K_s}, \tag{9}$$

$$R=\frac{\phi^2K_sK_f}{K_f\left(1-\phi-\frac{K_b}{K_s}\right)+\phi K_s}. \tag{10}$$

The dynamic permeability $k(\omega)$ describes the transition from viscosity toward inertia-dominated flow. The electrokinetic coupling coefficient $L(\omega)$ couples mechanical and EM-wave behavior. Note that for $L(\omega)=0$, mechanical and EM-fields are decoupled in Eqs. (1)–(7). Equations (1)–(4) then transform into the original Biot equations (Biot, 1956a, 1956b) and Eqs. (5)–(7) into Ohm's law and Maxwell's equations. The conductivity of the pore fluid is denoted as $\sigma(\omega)$. Expressions for $k(\omega)$, $L(\omega)$, and $\sigma(\omega)$ are given in Appendix A. As the dynamic permeability and the fluid and solid density terms always appear together, it is convenient to define effective densities $\rho_{11}(\omega)$, $\rho_{12}(\omega)$, and $\rho_{22}(\omega)$ as follows:

$$\rho_{11}(\omega)=(1-\phi)\rho_s-\rho_{12}(\omega), \tag{11}$$

$$\rho_{12}(\omega)=\phi\rho_f\left[1+i\frac{\phi\eta}{\omega\rho_fk(\omega)}\right], \tag{12}$$

$$\rho_{22}(\omega)=\phi\rho_f-\rho_{12}(\omega), \tag{13}$$

where ρ_s is the solid density. Adding Eqs. (1) and (2), we obtain the original dynamic equation (3) by Pride and Haartsen (1996) (no source term). With the definitions for $\rho_{12}(\omega)$ and $\rho_{22}(\omega)$, Eq. (2) is written as

$$i\omega\phi(\hat{\mathbf{U}}-\hat{\mathbf{u}})=L(\omega)\hat{\mathbf{E}}+\frac{k(\omega)}{\eta}(-\nabla\hat{p}+\omega^2\rho_f\hat{\mathbf{u}}), \tag{14}$$

which is Eq. (5) from Pride and Haartsen (1996) (no source term). By adding Eqs. (3) and (4), we obtain an expression for total stress $\hat{\underline{\underline{\boldsymbol{\tau}}}}=-\hat{\underline{\underline{\boldsymbol{\sigma}}}}-p\underline{\underline{\mathbf{I}}}$, which corresponds to Eq. (6) by Pride and Haartsen (1996). Full consistency with the original Pride equations is thus confirmed.

Substituting Eqs. (3) and (4) into Eqs. (1) and (2) yields

$$G\nabla^2\hat{\mathbf{u}}+(A+G)\nabla\nabla\cdot\hat{\mathbf{u}}+Q\nabla\nabla\cdot\hat{\mathbf{U}}=-\omega^2[\rho_{11}(\omega)\hat{\mathbf{u}}+\rho_{12}(\omega)\hat{\mathbf{U}}]+\frac{\eta\phi L(\omega)}{k(\omega)}\hat{\mathbf{E}}, \tag{15}$$

$$Q\nabla\nabla\cdot\hat{\mathbf{u}}+R\nabla\nabla\cdot\hat{\mathbf{U}}=-\omega^2[\rho_{12}(\omega)\hat{\mathbf{u}}+\rho_{22}(\omega)\hat{\mathbf{U}}]-\frac{\eta\phi L(\omega)}{k(\omega)}\hat{\mathbf{E}}. \tag{16}$$

Eliminating $(-\nabla\hat{p}+\omega^2\rho_f\hat{\mathbf{u}})$ from Eqs. (5) and (14) we obtain

$$i\omega\phi(\hat{\mathbf{U}}-\hat{\mathbf{u}})=L(\omega)\hat{\mathbf{E}}+\frac{k(\omega)}{\eta}\left(\frac{\hat{\mathbf{J}}-\sigma(\omega)\hat{\mathbf{E}}}{L(\omega)}\right). \tag{17}$$

Substituting Eq. (6) into Eq. (17) we obtain

$$i\omega\phi(\hat{\mathbf{U}}-\hat{\mathbf{u}})=L(\omega)\hat{\mathbf{E}}+\frac{k(\omega)}{\eta L(\omega)}\left(\nabla\times\hat{\mathbf{H}}-[i\omega\varepsilon+\sigma(\omega)]\hat{\mathbf{E}}\right). \tag{18}$$

Finally, substituting Eq. (7) into Eq. (18) we obtain

TABLE I. Parameters of the poroelastic medium and fluid. Parameters of poroelastic medium from Denneman *et al.* (2002).

Parameters poroelastic medium	
Bulk modulus framework of grains, K_b	5.8×10^9 Pa
Shear modulus framework of grains, G	3.4×10^9 Pa
Bulk modulus skeleton grains, K_s	40×10^9 Pa
Bulk modulus pore fluid, K_f	2.22×10^9 Pa
Pore fluid viscosity, η	0.001 Pa s
Pore fluid density, ρ_f	1000 kg/m^3
Solid density, ρ_s	2760 kg/m^3
Tortuosity of the porous medium, α_∞	2.3
Porosity of the porous medium, ϕ	0.24
Permeability, k_0	0.390×10^{-12} m^2
Temperature, T	295 K
Hydrogen exponent, pH	7
Relative permittivity of the pore fluid, ϵ_f	80
Relative permittivity of the solid, ϵ_s	4
Magnetic permeability ($=\mu_0$), μ	$4\pi\times10^{-7}$ H/m
Valence of species-1 ion, z_1	1
Valence of species-2 ion, z_2	–1
Mobility of species-1 ion, b_1 [a]	3.246×10^{11} m/(N s)
Mobility of species-2 ion, b_2 [a]	4.931×10^{11} m/(N s)
Zeta-potential, ζ [b]	–0.065 V
Similarity parameter, M	1
Concentration of the electrolyte, C	0.001 mol/l
Parameters fluid	
Bulk modulus fluid, K_{fl}	2.22×10^9 Pa
Fluid density, ρ_{fl}	1000 kg/m^3
Fluid relative permittivity, ϵ_{fl}	80
Fluid magnetic permeability ($=\mu_0$), μ_{fl}	$4\pi\times10^{-7}$ H/m
Fluid conductivity, σ_{fl}	5.0×10^{-3} S/m

[a]Nabighian and Corbett, 1988.
[b]From Eq. (A5).

$$\omega^2\mu\bar{\varepsilon}(\omega)\hat{\mathbf{E}}+\omega^2\mu\frac{\eta\phi L(\omega)}{k(\omega)}(\hat{\mathbf{U}}-\hat{\mathbf{u}})=\nabla\nabla\cdot\hat{\mathbf{E}}-\nabla^2\hat{\mathbf{E}}, \quad (19)$$

where $\bar{\varepsilon}(\omega)$ is the newly defined effective electrical permittivity

$$\bar{\varepsilon}(\omega)=\varepsilon-i\frac{\sigma(\omega)}{\omega}+i\frac{\eta L^2(\omega)}{\omega k(\omega)}. \quad (20)$$

Equations (15), (16), and (19) form a closed set of equations for the fields **u**, **U**, and **E**. A summary of electrokinetic symbols is given in Table I.

A. Electrokinetic wave velocities

Electrokinetic theory in isotropic, homogeneous media predicts the existence of uncoupled longitudinal and transversal modes. There are two longitudinal waves, the fast *P*-wave and the slow *P*-wave, and two transversal waves: an electromagnetic wave and a seismic shear wave (Pride and Haartsen, 1996). We derive complex electrokinetic wave velocities in a procedure similar to that of Allard (1993). First, we define the following Helmholtz decompositions:

$$\hat{\mathbf{u}}=\nabla\hat{\varphi}_s+\nabla\times\hat{\boldsymbol{\Psi}}_s, \quad (21)$$

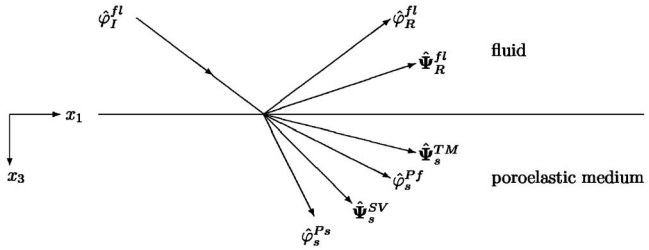


FIG. 1. Seismoelectric reflection and transmission potentials due to an incident fluid wave.

$$\hat{\mathbf{U}}=\nabla\hat{\varphi}_f+\nabla\times\hat{\boldsymbol{\Psi}}_f, \quad (22)$$

$$\hat{\mathbf{E}}=\nabla\hat{\varphi}_E+\nabla\times\hat{\boldsymbol{\Psi}}_E. \quad (23)$$

These are substituted in Eqs. (15), (16), and (19). This yields two separate sets: one for the longitudinal waves and one for the transversal waves. Subsequently, plane wave propagation is introduced in the $(x_1,0,x_3)$ plane (see Fig. 1), e.g., $\hat{\varphi}_s=\tilde{\varphi}_s\exp[-i\mathbf{k}\cdot\mathbf{x}]$ and $\hat{\boldsymbol{\Psi}}_s=(0,\tilde{\psi}_s\exp[-i\mathbf{k}\cdot\mathbf{x}],0)^T$, where **k** is the wavenumber vector and **x** is the position vector in that plane. Through elimination of $\tilde{\varphi}_E$ in the longitudinal wave equations a system of equations for $\tilde{\varphi}_s$ and $\tilde{\varphi}_f$ is obtained. The nontrivial solution yields the squared complex slowness (i.e., inverse complex-valued wave velocities) of the longitudinal waves $s_l^2(\omega)$

$$s_l^2(\omega)=\frac{-d_1(\omega)}{2d_2}\mp\frac{1}{2}\sqrt{\left[\frac{d_1(\omega)}{d_2}\right]^2-4\frac{d_0(\omega)}{d_2}}, \quad (24)$$

where $l=Pf, Ps$ refers to the fast and slow longitudinal waves, respectively, and

$$d_0(\omega)=\bar{\rho}_{11}(\omega)\bar{\rho}_{22}(\omega)-[\bar{\rho}_{12}(\omega)]^2,$$

$$d_1(\omega)=-[P\bar{\rho}_{22}(\omega)+R\bar{\rho}_{11}(\omega)-2Q\bar{\rho}_{12}(\omega)],$$

$$d_2=PR-Q^2, \quad (25)$$

where $P=A+2G$. As the effective densities and the electrokinetic coupling factor always appear together, it is convenient to define new effective densities $\bar{\rho}_{11}(\omega)$, $\bar{\rho}_{12}(\omega)$, and $\bar{\rho}_{22}(\omega)$ as follows:

$$\bar{\rho}_{11}(\omega)=\rho_{11}(\omega)-E_K(\omega), \quad (26)$$

$$\bar{\rho}_{12}(\omega)=\rho_{12}(\omega)+E_K(\omega), \quad (27)$$

$$\bar{\rho}_{22}(\omega)=\rho_{22}(\omega)-E_K(\omega), \quad (28)$$

$$E_K(\omega)=\frac{\eta^2\phi^2L^2(\omega)}{k^2(\omega)\bar{\varepsilon}(\omega)\omega^2}. \quad (29)$$

We note that both viscous and electrokinetic coupling effects are now elegantly comprised in effective frequency-dependent densities.

A similar approach is used for the transversal waves. Through elimination of $\tilde{\boldsymbol{\Psi}}_E$ in the transversal wave equations, a system of equations for $\tilde{\boldsymbol{\Psi}}_s$ and $\tilde{\boldsymbol{\Psi}}_f$ is obtained. The nontrivial solution yields the squared complex slowness of

the transversal waves $s_l^2(\omega)$ according to Eq. (24), where now $l=EM, S$ refers to the EM and seismic shear waves, respectively, and

$$d_0(\omega) = \mu\bar{\varepsilon}(\omega)\frac{\bar{\rho}_{11}(\omega)\bar{\rho}_{22}(\omega) - [\bar{\rho}_{12}(\omega)]^2}{G},$$

$$d_1(\omega) = -\mu\bar{\varepsilon}(\omega)\bar{\rho}_{22}(\omega) - \frac{\rho_{11}(\omega)\rho_{22}(\omega) - [\rho_{12}(\omega)]^2}{G},$$

$$d_2(\omega) = \rho_{22}(\omega). \tag{30}$$

From the complex-valued slowness, phase velocities and attenuation coefficients are readily obtained.

III. REFLECTION AND TRANSMISSION AT A PLANE FLUID/POROUS-MEDIUM INTERFACE

A compressible fluid overlays an isotropic, homogeneous porous medium. The horizontal interface is at $x_3=0$ (see Fig. 1). We adopt a right-hand (x_1,x_2,x_3) coordinate system with the x_3-axis pointing downward. A plane fluid wave (scalar potential $\hat{\varphi}_I^{fl}$) impinges on the interface where it reflects as a P-wave (scalar potential $\hat{\varphi}_R^{fl}$) and a transverse magnetic (*TM*) EM-wave (vector potential $\hat{\mathbf{\Psi}}_R^{fl}$). It transmits as a *TM*-wave (vector potential $\hat{\mathbf{\Psi}}_s^{TM}$), a fast longitudinal wave (scalar potential $\hat{\varphi}_s^{Pf}$), a slow longitudinal wave (scalar potential $\hat{\varphi}_s^{Ps}$), and a vertical shear (*SV*)-wave (vector potential $\hat{\mathbf{\Psi}}_s^{SV}$). The reflection and transmission coefficients are obtained from the (open-pore) boundary conditions, which are continuity of fluid volume displacement, fluid pressure, the horizontal magnetic and electric fields, and vanishing vertical and horizontal intergranular stresses (Deresiewicz and Skalak, 1963; Pride and Haartsen, 1996)

$$(1-\phi)\hat{u}_3 + \phi\hat{U}_3 = \hat{U}_3^{fl}, \tag{31}$$

$$\hat{p} = \hat{p}^{fl}, \tag{32}$$

$$\hat{\sigma}_{13} = \hat{\sigma}_{33} = 0, \tag{33}$$

$$\hat{H}_2 = \hat{H}_2^{fl}, \tag{34}$$

$$\hat{E}_1 = \hat{E}_1^{fl}. \tag{35}$$

The pressure in the upper fluid layer is denoted by p^{fl}. Magnetic and electric fields in the upper fluid layer are indicated by $\mathbf{H}^{fl}$ and $\mathbf{E}^{fl}$, respectively. We relate p^{fl} to upper fluid layer displacement $\mathbf{U}^{fl}$ by $p^{fl}=-K_{fl}\nabla\cdot\mathbf{U}^{fl}$, where K_{fl} is the bulk modulus of the upper fluid layer. The upper fluid layer wave speed c is related to the upper layer fluid density ρ_{fl} by $c=\sqrt{K_{fl}/\rho_{fl}}$. Viscosity does not play a role in the overlying fluid because the ratio of viscous and inertia terms is $\eta\omega/c^2\rho_{fl}$, which is on the order $10^{-11}-10^{-6}$ for the frequencies considered in this paper.

Scalar potentials $\hat{\varphi}_s$, $\hat{\varphi}_f$, and $\hat{\varphi}_E$, see Eqs. (21)–(23), are composed of contributions of the longitudinal waves, e.g., $\hat{\varphi}_s=\hat{\varphi}_s^{Pf}+\hat{\varphi}_s^{Ps}$, and, similarly, vector potentials $\hat{\mathbf{\Psi}}_s$, $\hat{\mathbf{\Psi}}_f$, and $\hat{\mathbf{\Psi}}_E$ are composed of contributions of the transversal waves, e.g., $\hat{\mathbf{\Psi}}_s=\hat{\mathbf{\Psi}}_s^{TM}+\hat{\mathbf{\Psi}}_s^{SV}=(0,\hat{\psi}_s^{TM},0)^T+(0,\hat{\psi}_s^{SV},0)^T$. The dispersion relations given in Eq. (24) not only predict the phase velocities and attenuation coefficients, but also the fluid-solid amplitude ratios β^m and β^n and the amplitude ratios α^m and α^n $(\mathrm{V/m^2})$ that describe the strength of the electric field with respect to the solid displacement amplitude, for each wave

$$\beta^m(\omega) = \frac{\hat{\varphi}_f^m}{\hat{\varphi}_s^m} = \frac{\bar{\rho}_{11}(\omega) - Ps_m^2(\omega)}{Qs_m^2(\omega) - \bar{\rho}_{12}(\omega)}, \tag{36}$$

$$\beta^n(\omega) = \frac{\hat{\psi}_f^n}{\hat{\psi}_s^n} = \frac{Gs_n^2(\omega) - (1-\phi)\rho_s}{\phi\rho_f}, \tag{37}$$

$$\alpha^m(\omega) = \frac{\hat{\varphi}_E^m}{\hat{\varphi}_s^m} = \frac{\eta\phi L(\omega)}{k(\omega)\bar{\varepsilon}(\omega)}[1-\beta^m(\omega)], \tag{38}$$

$$\alpha^n(\omega) = \frac{\hat{\psi}_E^n}{\hat{\psi}_s^n} = \frac{\mu\eta\phi L(\omega)}{k(\omega)[\mu\bar{\varepsilon}(\omega) - s_n^2(\omega)]}[1-\beta^n(\omega)], \tag{39}$$

for $m=Pf$ or Ps, and $n=TM$ or SV. We now specify the following plane wave propagation expressions:

$$\hat{\varphi}_I^{fl} = \tilde{\varphi}_I^{fl}\exp[-i(k_1x_1 + k_3^{fl}x_3)], \tag{40}$$

$$\hat{\varphi}_R^{fl} = \tilde{\varphi}_R^{fl}\exp[-i(k_1x_1 - k_3^{fl}x_3)], \tag{41}$$

$$\hat{\mathbf{\Psi}}_R^{fl} = (0,\tilde{\psi}_R^{fl}\exp[-i(k_1x_1 - k_3^{E}x_3)],0)^T, \tag{42}$$

$$\hat{\varphi}_s^m = \tilde{\varphi}_s^m\exp[-i(k_1x_1 + k_3^{m}x_3)], \tag{43}$$

$$\hat{\mathbf{\Psi}}_s^n = (0,\tilde{\psi}_s^n\exp[-i(k_1x_1 + k_3^{n}x_3)],0)^T, \tag{44}$$

where k_3^{fl} and k_3^E are the vertical wavenumbers of the incident fluid wave and reflected EM-wave, respectively. The reflection and transmission coefficients are

$$R^E = \frac{\tilde{\psi}_R^{fl}}{\tilde{\varphi}_I^{fl}}, \quad R^M = \frac{\tilde{\varphi}_R^{fl}}{\tilde{\varphi}_I^{fl}}, \quad T^{Pf} = \frac{\tilde{\varphi}_s^{Pf}}{\tilde{\varphi}_I^{fl}},$$

$$T^{Ps} = \frac{\tilde{\varphi}_s^{Ps}}{\tilde{\varphi}_I^{fl}}, \quad T^{TM} = \frac{\tilde{\psi}_s^{TM}}{\tilde{\varphi}_I^{fl}}, \quad T^{SV} = \frac{\tilde{\psi}_s^{SV}}{\tilde{\varphi}_I^{fl}}, \tag{45}$$

where R^M indicates the mechanical reflection coefficient and R^E indicates the seismoelectric reflection coefficient. Applying boundary conditions (31)–(35) yields a linear system of six equations and six unknowns

$$\underline{\underline{\mathbf{A}}}\cdot(R^E,R^M,T^{Pf},T^{Ps},T^{TM},T^{SV})^T = (k_3^{fl},\phi\rho_{fl},0,0,0,0)^T, \tag{46}$$

where elements of matrix $\underline{\underline{\mathbf{A}}}$ are given in Appendix B. We solve Eq. (46) by Cramer's rule and show the results for R^E and T^{TM} as a function of incidence angle θ at a fixed frequency of 10 Hz in Fig. 2. The parameter values, representative for a water/shallow-sandstone interface, are given in Table I. We use these values in all computations of the reflection and transmission coefficients. For sin θ between $\sim$0 and $\sim$0.6, $|R^E|$ is approximately constant. Note that for

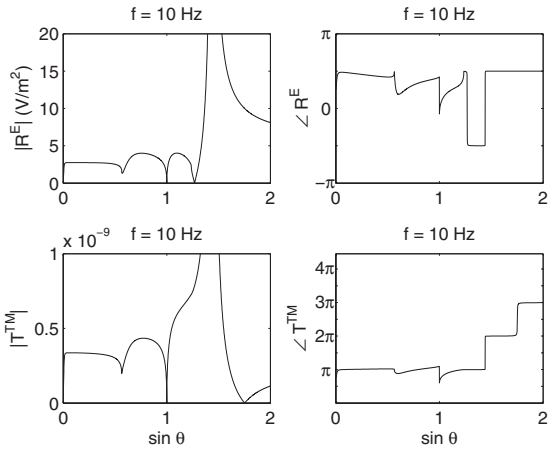


FIG. 2. Amplitudes (left) and phase values (right) of seismoelectric reflection (R^E) and transmission (T^{TM}) coefficients at a fluid/porous-medium boundary.

sin θ=0, no reflected EM-wave is generated due to the prescribed boundary conditions. For the parameters of Table I, $|R^E|$ is approximately 3 V/m^2, which is equivalent to $\sim 8 \times 10^{-7}$ V/Pa at 10 Hz. At sin $\theta \approx 0.57$, $|R^E|$ has a local minimum associated with critical reflection of the fast *P*-wave. Obviously, sin θ=1 corresponds to grazing angle of incidence and there is no seismoelectric conversion ($|R^E|$=0). We associate sin $\theta \approx 1.24$ with critical reflection of the SV-wave and we associate the peak at sin $\theta \approx 1.44$ with the Stoneley wave. The above mentioned behavior is also visible in the phase behavior of R^E, and the behavior of T^{TM}. At sin θ=0 and sin θ=1, R^E and T^{TM} are identically zero. We note that generally $|T^{TM}|$ is on the order 10^{-10}.

Next we show in Fig. 3 the frequency-dependence of $|R^E|$ and $|T^{TM}|$ for 45° angle of incidence. Both seismoelectric coefficients increase strongly with frequency, which corresponds to increasing EM-wave velocities in the diffusive regime for both the fluid and poroelastic medium. We note that the increase in $|T^{TM}|$ sharply diminishes at $\omega=\omega_t$, i.e., for the frequency that separates viscosity-dominated flow from inertia-dominated flow.

IV. SENSITIVITY ANALYSIS

A natural way to perform a sensitivity analysis is to analyze the ratio of reflected/transmitted wave energy fluxes to the incident flux; see, e.g., Dutta and Odé (1983) and Santos *et al.* (1992).

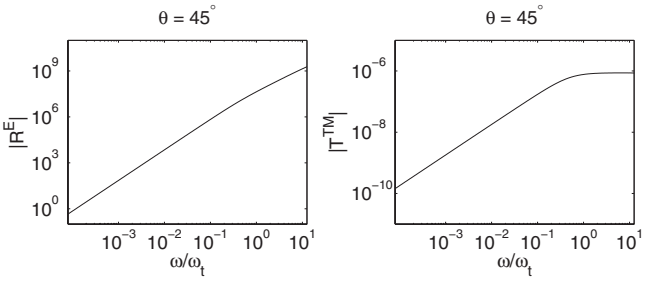


FIG. 3. Magnitude of seismoelectric coefficients R^E and T^{TM} as a function of normalized frequency at θ=45° angle of incidence.

A. Vertical electrokinetic energy fluxes

The electrokinetic vertical energy flux $Y_3(\mathbf{x},t)$ of the porous medium, defined by the dot product of the Poynting vector $\mathbf{Y}$ and the unit normal to the interface $\mathbf{n}=(0,0,1)^T$, is given by (Haartsen, 1995)

$$\begin{aligned} Y_3(\mathbf{x},t) &= [\mathbf{E}(\mathbf{x},t)\times\mathbf{H}(\mathbf{x},t) + \underline{\boldsymbol{\tau}}(\mathbf{x},t)\cdot\dot{\mathbf{u}}(\mathbf{x},t) \\ &\quad - \phi p(\mathbf{x},t)\big(\dot{\mathbf{U}}(\mathbf{x},t) - \dot{\mathbf{u}}(\mathbf{x},t)\big)]\cdot\mathbf{n} \\ &= E_1(\mathbf{x},t)H_2(\mathbf{x},t) + \tau_{31}(\mathbf{x},t)\dot{u}_1(\mathbf{x},t) \\ &\quad + \tau_{33}(\mathbf{x},t)\dot{u}_3(\mathbf{x},t) - \phi p(\mathbf{x},t) \\ &\quad \times \big(\dot{U}_3(\mathbf{x},t) - \dot{u}_3(\mathbf{x},t)\big). \end{aligned} \tag{47}$$

Similar expressions to Eq. (47) were provided by Pride and Haartsen (1996), Dutta and Odé (1983), and Santos *et al.*, (1992). We note that in our geometry (see Fig. 1) $H_1(\mathbf{x},t) = E_2(\mathbf{x},t) = \tau_{32}(\mathbf{x},t) = u_2(\mathbf{x},t) = 0$. We now write real parts of all frequency-domain variables and calculate a time-average over 1 cycle in time so that $\langle Y_3(\mathbf{x},t)\rangle$ becomes

$$\begin{aligned} \langle Y_3(\mathbf{x},t)\rangle &= \frac{1}{2\pi}\int_0^{2\pi} \Big(\mathrm{Re}\big[\hat{E}_1\exp[i\omega t]\big]\mathrm{Re}\big[\hat{H}_2\exp[i\omega t]\big] \\ &\quad + \mathrm{Re}\big[\hat{\tau}_{31}\exp[i\omega t]\big]\mathrm{Re}\big[i\omega\hat{u}_1\exp[i\omega t]\big] \\ &\quad + \mathrm{Re}\big[\hat{\tau}_{33}\exp[i\omega t]\big]\mathrm{Re}\big[i\omega\hat{u}_3\exp[i\omega t]\big] \\ &\quad - \mathrm{Re}\big[\hat{p}\exp[i\omega t]\big]\mathrm{Re}\big[i\omega\phi(\hat{U}_3 - \hat{u}_3) \\ &\quad \times\exp[i\omega t]\big]\Big)d(\omega t) \\ &= \frac{1}{2}[\hat{E}_1'\hat{H}_2' + \hat{E}_1''\hat{H}_2''] + \frac{\omega}{2}[\hat{\tau}_{31}''\hat{u}_1' - \hat{\tau}_{31}'\hat{u}_1''] \\ &\quad + \frac{\omega}{2}[\hat{\tau}_{33}''\hat{u}_3' - \hat{\tau}_{33}'\hat{u}_3''] - \frac{\omega}{2}[\phi\hat{p}''(\hat{U}_3' - \hat{u}_3') \\ &\quad - \phi\hat{p}'(\hat{U}_3'' - \hat{u}_3'')], \end{aligned} \tag{48}$$

where single and double primes denote real and imaginary parts of complex quantities, respectively. Denoting complex conjugation by an asterisk, we find that

$$\begin{aligned} \langle Y_3(\mathbf{x},t)\rangle &= \tfrac{1}{4}\Big\{(\hat{E}_1\hat{H}_2^\star + \hat{E}_1^\star\hat{H}_2) - i\omega\Big[(\hat{\tau}_{31}\hat{u}_1^\star - \hat{\tau}_{31}^\star\hat{u}_1) \\ &\quad + (\hat{\tau}_{33}\hat{u}_3^\star - \hat{\tau}_{33}^\star\hat{u}_3) - \big(\phi\hat{p}(\hat{U}_3^\star - \hat{u}_3^\star) \\ &\quad - \phi\hat{p}^\star(\hat{U}_3 - \hat{u}_3)\big)\Big]\Big\}. \end{aligned} \tag{49}$$

Each of the components in Eq. (49) is composed of contributions from the four electrokinetic waves, e.g., $\hat{E}_1 = \hat{E}_1^{Pf} + \hat{E}_1^{Ps} + \hat{E}_1^{TM} + \hat{E}_1^{SV}$. Writing all terms in Eq. (49) in separate contributions from each of the waves, we will have four orthodox fluxes, resulting from wave amplitudes multiplied by themselves, and six interference fluxes, resulting from cross-terms (Dutta and Odé, 1983; Santos *et al.*, 1992). In the non-dissipative fluid layer there is only an orthodox reflected EM-flux $\langle Y_3^{fl}(\mathbf{x},t)\rangle^{E,E}$, and incident and reflected orthodox acoustic fluxes $\langle Y_3^{fl}(\mathbf{x},t)\rangle^{Pin,Pin}$ and $\langle Y_3^{fl}(\mathbf{x},t)\rangle^{Pr,Pr}$. Vertical energy flux conservation requires

TABLE II. Seismoelectric reflection and transmission coefficients for $\theta=30°$ at $f=10$ Hz and $f=500$ kHz.

	$f=10$ Hz	$f=500$ kHz
$T_E^{Pf,Pf}$	4.6700×10^{-1}	4.7334×10^{-1}
$T_E^{Ps,Ps}$	1.8083×10^{-3}	1.3026×10^{-1}
$T_E^{TM,TM}$	-2.6500×10^{-10}	-4.9895×10^{-7}
$T_E^{SV,SV}$	1.3084×10^{-1}	7.0295×10^{-2}
$T_E^{Pf,Ps}$	-2.0028×10^{-10}	-8.0868×10^{-7}
$T_E^{Pf,TM}$	-2.7470×10^{-12}	-2.5059×10^{-9}
$T_E^{Pf,SV}$	7.4887×10^{-11}	3.0585×10^{-7}
$T_E^{Ps,TM}$	6.7059×10^{-10}	5.2570×10^{-7}
$T_E^{Ps,SV}$	-1.0420×10^{-8}	-3.7939×10^{-5}
$T_E^{TM,SV}$	4.3970×10^{-14}	1.0814×10^{-7}
$R_E^{E,E}$	7.4497×10^{-11}	1.1871×10^{-7}
$R_E^{Pr,Pr}$	-4.0035×10^{-1}	-3.2615×10^{-1}
$T_E-R_E^{E,E}-R_E^{Pr,Pr}$	1.0000×10^{0}	1.0000×10^{0}

$$\langle Y_3(\mathbf{x},t)\rangle-\langle Y_3^{fl}(\mathbf{x},t)\rangle^{E,E}-\langle Y_3^{fl}(\mathbf{x},t)\rangle^{Pr,Pr}=\langle Y_3^{fl}(\mathbf{x},t)\rangle^{Pin,Pin}. \quad (50)$$

Introducing energy transmission coefficients, e.g., $T_E=\langle Y_3(\mathbf{x},t)\rangle/\langle Y_3^{fl}(\mathbf{x},t)\rangle^{Pin,Pin}$, we write

$$T_E-R_E^{E,E}-R_E^{Pr,Pr}=1. \quad (51)$$

We note, again, that T_E consists of four orthodox flux transmission coefficients and six interference flux transmission coefficients. Numerical examples are given in Tables II and III. In Table II, the angle of incidence θ is fixed at 30° and frequencies f of 10 Hz (left column) and 500 kHz (right column) are chosen. Higher frequencies favor seismoelectric conversion. In Table III, the angular frequency is fixed at 10^6 rad/s, and angles of incidence $\theta=30°$ (left column) and $\theta=45°$ (right column) are chosen. Equation (51) is satisfied indeed. $T_E^{Pf,Ps}$ and $T_E^{Ps,SV}$ are negative, because of the out-of-phase behavior of the slow P-wave. $R_E^{Pr,Pr}$ is negative, because the reflected acoustic pressure and fluid displacement have opposite phase. $T_E^{Pf,Pf}$ approaches zero for $\theta=45°$ because this angle is beyond the critical angle of incidence of the fast P-wave.

TABLE III. Seismoelectric reflection and transmission coefficients for $\omega=10^6$ rad/s at $\theta=30°$ and $\theta=45°$.

	$\theta=30°$	$\theta=45°$
$T_E^{Pf,Pf}$	4.7389×10^{-1}	1.6454×10^{-7}
$T_E^{Ps,Ps}$	1.2478×10^{-1}	2.7971×10^{-1}
$T_E^{TM,TM}$	-5.0880×10^{-7}	-1.6581×10^{-6}
$T_E^{SV,SV}$	7.2674×10^{-2}	6.6128×10^{-1}
$T_E^{Pf,Ps}$	-2.6036×10^{-6}	3.2114×10^{-5}
$T_E^{Pf,TM}$	-3.3667×10^{-9}	4.4496×10^{-9}
$T_E^{Pf,SV}$	9.8734×10^{-7}	5.2354×10^{-5}
$T_E^{Ps,TM}$	7.1950×10^{-7}	2.4385×10^{-6}
$T_E^{Ps,SV}$	-1.2373×10^{-4}	1.9413×10^{-3}
$T_E^{TM,SV}$	1.3113×10^{-7}	8.4593×10^{-7}
$R_E^{E,E}$	1.3503×10^{-7}	5.4312×10^{-7}
$R_E^{Pr,Pr}$	-3.2877×10^{-1}	-5.6982×10^{-2}
$T_E-R_E^{E,E}-R_E^{Pr,Pr}$	1.0000×10^{0}	1.0000×10^{0}

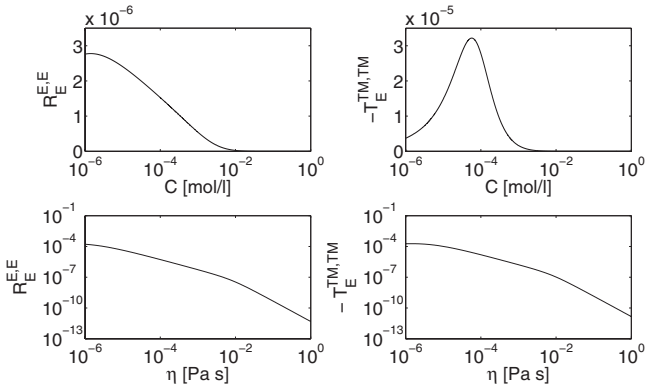


FIG. 4. Variation of $R_E^{E,E}$ and $T_E^{TM,TM}$ as a function of electrolyte concentration (C) and pore fluid viscosity (η) at $\omega=10^6$ rad/s and $\theta=45°$.

B. Parameter variations

Seismoelectric conversions are generated at contrasts in (i) elastic properties (e.g., elastic moduli), (ii) fluid chemistry (pH and electrolyte concentration), and (iii) transport properties such as permeability (Haartsen and Pride, 1997). Electrolyte concentration (C), pore fluid viscosity (η), and permeability (k_0) are dominant seismoelectric conversion parameters while pH, pore fluid bulk modulus (K_f), porosity (ϕ), and tortuosity (α_∞) influence the seismoelectric conversion to a smaller extent (see also Haartsen and Pride, 1997; Mikhailov *et al.*, 2000; Garambois and Dietrich, 2002; Bordes *et al.*, 2006). We investigate the variation of $R_E^{E,E}$ and $T_E^{TM,TM}$ as a function of these parameters in Figs. 4–7 at fixed angular frequency and angle of incidence: $\omega=10^6$ rad/s and $\theta=45°$.

In Fig. 4, $R_E^{E,E}$ decreases with increasing electrolyte concentration because the absolute value of the zeta-potential decreases, which reduces the electrokinetic coupling, as can be seen from Eq. (A4). $T_E^{TM,TM}$ peaks around $C=10^{-4}$ mol/l. This behavior is related to the vertical electrokinetic energy flux, which depends on the TM-wave velocity. We found that the corresponding maximum attenuation of this wave is at $C=10^{-4}$ mol/l.

Both coefficients strongly decrease with increasing fluid viscosity. This is intuitively explained because a low viscosity will favor EM-wave generation as fluid circulation is enhanced, as was noted by Garambois and Dietrich (2002).

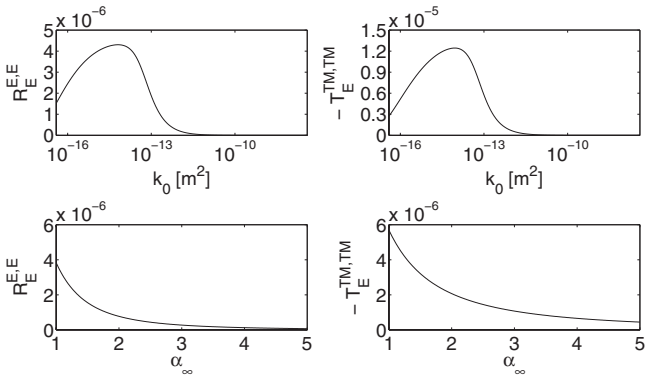


FIG. 5. Variation of $R_E^{E,E}$ and $T_E^{TM,TM}$ as a function of permeability (k_0) and tortuosity (α_∞) at $\omega=10^6$ rad/s and $\theta=45°$.

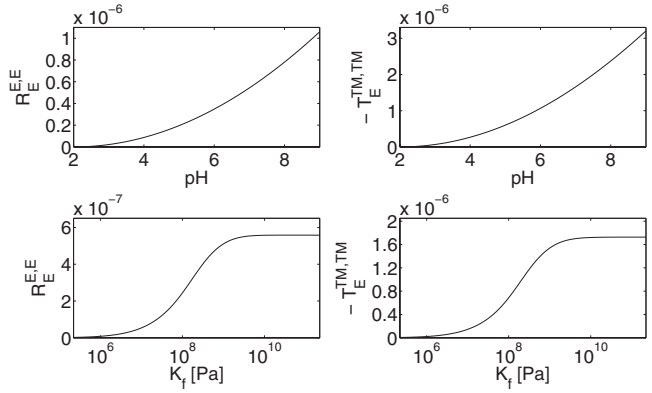


FIG. 6. Variation of $R_E^{E,E}$ and $T_E^{TM,TM}$ as a function of pH and bulk modulus of the pore fluid (K_f) at $\omega=10^6$ rad/s and $\theta=45°$.

In the upper two graphs of Fig. 5 we see that seismoelectric coefficients vary over several orders of magnitude as a function of the permeability k_0. This behavior is attributed to the influence of k_0 on the electrokinetic coupling coefficient $L(\omega)$. From Eq. (A3) we see that by increasing permeability we also increase the pore volume-to-surface ratio Λ, which has two competing effects on $L(\omega)$. Initially the length scale of Λ approaches the characteristic length of the electric double layer d so that $L(\omega)$ increases with increasing Λ. For permeabilities larger than 10^{-14} m^2 the influence of Λ^2 in the denominator of $L(\omega)$ becomes dominant so that the seismoelectric coefficients decrease. The effect of tortuosity on the seismoelectric coefficients is shown in the lower two graphs of Fig. 5. Tortuosity also influences $L(\omega)$, but its effect is smaller than that of k_0 because of its smaller variability. Again, increasing tortuosity effectively increases Λ. For these parameter value combinations, d/Λ is always negligible so that the seismoelectric coefficients decrease with increasing tortuosity.

$R_E^{E,E}$ and $T_E^{TM,TM}$ increase approximately linearly with pH (upper two graphs of Fig. 6), which is a result of the linear relation used for the zeta-potential [see Eq. (A5)]. Also $L(\omega)$ is directly proportional to ζ [see Eq. (A4)].

In the lower two graphs of Fig. 6, we notice that as the pore fluid becomes stiffer, seismoelectric coefficients $R_E^{E,E}$ and $T_E^{TM,TM}$ increase toward a limiting value for large pore fluid bulk moduli. This increase corresponds with a larger contrast in seismoelectric properties across the interface as the compressional wave velocities increase with increasing incompressibility of the pore fluid.

$R_E^{E,E}$ and $T_E^{TM,TM}$ as a function of porosity ϕ are shown in Fig. 7. The increase with porosity of both seismoelectric coefficients is explained by the fact that $L(\omega)$ depends approximately linearly on ϕ [see Eq. (A4)], implying that an increase in surface area increases electrokinetic coupling.

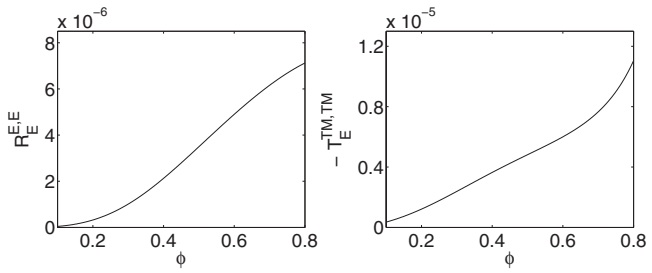


FIG. 7. Variation of $R_E^{E,E}$ and $T_E^{TM,TM}$ as a function of porosity (ϕ) at $\omega=10^6$ rad/s and $\theta=45°$.

This sensitivity analysis indicates that for a water-saturated natural sandstone, electrolyte concentration, pore fluid viscosity, and permeability should be sufficiently low ($C\leq 10^{-3}$ mol/l, $\eta\leq 10^{-3}$ Pa s, and $k_0\leq 10^{-12}$ m^2), and the pH sufficiently high (say pH=7 or more) in order to have appreciable seismoelectric conversion.

V. CONCLUSIONS

The dispersion relations for seismoelectric wave propagation in poroelastic media were solved in terms of effective densities. This allows an elegant incorporation of both viscous and electrokinetic coupling in the effective densities. We solved the seismoelectric reflection and transmission coefficients at a fluid/porous-medium boundary. The dependence of the reflected and transmitted EM-waves due to an incident fluid wave on angle of incidence and frequency was shown. Recasting seismoelectric reflection and transmission coefficients to vertical energy fluxes, while accounting for orthodox and interference fluxes, it was asserted that energy conservation is obeyed. For a water/shallow-sandstone interface, seismoelectric conversion at 10 Hz is on the order 10^{-10} and at 500 kHz it is on the order 10^{-7} for energy flux coefficients. A sensitivity analysis was performed. Electrolyte concentration, pore fluid viscosity, and permeability were varied within realistic ranges and it was shown that this influenced seismoelectric conversions over several orders of magnitude. This sensitivity analysis allows the possibility to devise experiments in which the seismoelectric effect can be detected. Moreover, it allows us to appreciate the use of the electrokinetic effect as an exploration technique that is complementary to conventional seismic surveys for exploring hydrocarbon reserves.

ACKNOWLEDGMENTS

The authors thank Christiaan Schoemaker for fruitful discussions and Evert Slob for in-depth comments. This work was performed under the financial support of Shell-FOM within the research program "The physics of fluids and sound propagation."

APPENDIX A: DYNAMIC COEFFICIENTS

The dynamic permeability $k(\omega)$ is given by (Johnson *et al.*, 1987)

$$k(\omega)=k_0\left[\sqrt{1+i\frac{\omega}{\omega_t}\frac{M}{2}}+i\frac{\omega}{\omega_t}\right]^{-1}, \tag{A1}$$

where k_0 is permeability. The transition frequency ω_t and the similarity parameter M are given by

$$\omega_t=\frac{\phi\eta}{\alpha_\infty k_0\rho_f}, \tag{A2}$$

$$M = \frac{8\alpha_\infty k_0}{\phi\Lambda^2}, \tag{A3}$$

where α_∞ is the tortuosity and Λ is a weighted pore volume-to-surface ratio. The electrokinetic coupling coefficient $L(\omega)$ is obtained from Pride's (1994) theory

$$L(\omega) = -\frac{\phi}{\alpha_\infty}\frac{\varepsilon_0\epsilon_f\zeta}{\eta}\left(1-2\frac{d}{\Lambda}\right) \times\left[1+2i\frac{\omega}{M\omega_t}\left(1-2\frac{d}{\Lambda}\right)^2\left(1+d\sqrt{\frac{i\omega\rho_f}{\eta}}\right)^2\right]^{-1/2}, \tag{A4}$$

where d is the Debye length (defined below) and ζ is the zeta-potential. A semi-empirical relationship between the zeta-potential, electrolyte concentration (C), and pH is used (Pride and Garambois, 2005) as follows:

$$\zeta = [0.010 + 0.025\ \log_{10}(C)]\frac{\mathrm{pH}-2}{5}. \tag{A5}$$

The dynamic conductivity $\sigma(\omega)$ is given by

$$\sigma(\omega) = \frac{\phi\sigma_f}{\alpha_\infty}\left[1+\frac{2[C_{\mathrm{em}}+C_{\mathrm{os}}(\omega)]}{\sigma_f\Lambda}\right], \tag{A6}$$

where σ_f is conductivity of the pore fluid, C_{em} is excess conductance associated with electromigration of excess charge, and $C_{\mathrm{os}}(\omega)$ is excess conductance associated with electroosmosis. Expressions for σ_f, C_{em}, and $C_{\mathrm{os}}(\omega)$ are

$$\sigma_f = \sum_{l=1}^{N_i}(ez_l)^2 b_l N_l, \tag{A7}$$

$$C_{\mathrm{em}} = 2d\sum_{l=1}^{N_i}(ez_l)^2 b_l N_l\left[\exp\left(-\frac{ez_l\zeta}{2k_BT}\right)-1\right], \tag{A8}$$

$$C_{\mathrm{os}}(\omega) = \frac{(\varepsilon_0\epsilon_f)^2\zeta^2}{2d\eta}P\left[1+\frac{2}{P}d\sqrt{\frac{i\omega\rho_f}{\eta}}\right]^{-1}, \tag{A9}$$

$$P = \frac{8k_BTd^2}{\varepsilon_0\epsilon_f\zeta^2}\sum_{l=1}^{N_i}N_l\left[\exp\left(-\frac{ez_l\zeta}{2k_BT}\right)-1\right], \tag{A10}$$

$$d = \sqrt{\left[\sum_{l=1}^{N_i}\frac{(ez_l)^2N_l}{\varepsilon_0\epsilon_f k_BT}\right]^{-1}}, \tag{A11}$$

where k_B is Boltzmann's constant, e is electron charge, T is temperature, and N_l is the amount of species l ions per volume. In this article we consider a binary symmetric electrolyte (N_i=2; z_1=−z_2=1; N_1=N_2) with mobilities b_1 and b_2 (see Table I).

APPENDIX B: ELEMENTS OF $\underline{\underline{\mathbf{A}}}$

Applying the boundary conditions for the seismoelectric reflection and transmission problem yields a linear system

$$\underline{\underline{\mathbf{A}}}\cdot(R^E,R^M,T^{Pf},T^{Ps},T^{TM},T^{SV})^T = (k_3^{fl},\phi\rho_{fl},0,0,0,0)^T, \tag{B1}$$

where elements of matrix $\underline{\underline{\mathbf{A}}}$ are

$$\begin{aligned}
A(1,1) &= 0,\\
A(1,2) &= k_3^{fl},\\
A(1,3) &= k_3^{Pf}[1-\phi+\phi\beta^{Pf}(\omega)],\\
A(1,4) &= k_3^{Ps}[1-\phi+\phi\beta^{Ps}(\omega)],\\
A(1,5) &= k_1[1-\phi+\phi\beta^{TM}(\omega)],\\
A(1,6) &= k_1[1-\phi+\phi\beta^{SV}(\omega)],
\end{aligned} \tag{B2}$$

$$\begin{aligned}
A(2,1) &= 0,\\
A(2,2) &= -\phi\rho_{fl},\\
A(2,3) &= [Q+R\beta^{Pf}(\omega)]s_{Pf}^2(\omega),\\
A(2,4) &= [Q+R\beta^{Ps}(\omega)]s_{Ps}^2(\omega),\\
A(2,5) &= 0,\\
A(2,6) &= 0,
\end{aligned} \tag{B3}$$

$$\begin{aligned}
A(3,1) &= 0,\\
A(3,2) &= 0,\\
A(3,3) &= k_1k_3^{Pf},\\
A(3,4) &= k_1k_3^{Ps},\\
A(3,5) &= k_1^2-\frac{\omega^2}{2}s_{TM}^2(\omega),\\
A(3,6) &= k_1^2-\frac{\omega^2}{2}s_{SV}^2(\omega),
\end{aligned} \tag{B4}$$

$$\begin{aligned}
A(4,1) &= 0,\\
A(4,2) &= 0,\\
A(4,3) &= k_1^2-\omega^2s_{Pf}^2(\omega)\frac{N_1(\omega)}{2G},\\
A(4,4) &= k_1^2-\omega^2s_{Ps}^2(\omega)\frac{N_2(\omega)}{2G},\\
A(4,5) &= -k_1k_3^{TM},\\
A(4,6) &= -k_1k_3^{SV},
\end{aligned} \tag{B5}$$

$$A(5,1) = -\frac{1}{\mu_{fl}} s_E^2(\omega),$$

$$A(5,2) = 0,$$

$$A(5,3) = 0,$$

$$A(5,4) = 0,$$

$$A(5,5) = \frac{\alpha^{TM}(\omega)}{\mu} s_{TM}^2(\omega),$$

$$A(5,6) = \frac{\alpha^{SV}(\omega)}{\mu} s_{SV}^2(\omega), \tag{B6}$$

$$A(6,1) = -k_3^E,$$

$$A(6,2) = 0,$$

$$A(6,3) = k_1 \alpha^{Pf}(\omega),$$

$$A(6,4) = k_1 \alpha^{Ps}(\omega),$$

$$A(6,5) = -k_3^{TM} \alpha^{TM}(\omega),$$

$$A(6,6) = -k_3^{SV} \alpha^{SV}(\omega), \tag{B7}$$

where $N_1(\omega) = P - Q(1-\phi)/\phi + [Q - R(1-\phi)/\phi]\beta^{Pf}(\omega)$, $N_2(\omega) = P - Q(1-\phi)/\phi + [Q - R(1-\phi)/\phi]\beta^{Ps}(\omega)$, and $s_E(\omega)$ is the slowness of the electromagnetic wave in the fluid, which can be calculated using $s_E^2(\omega) = (\omega^2 \mu_{fl} \varepsilon_0 \epsilon_{fl} - i\omega\mu_{fl}\sigma_{fl})/\omega^2$ (Nabighian and Corbett, 1988) with upper fluid layer permeability, relative permittivity, and conductivity denoted by μ_{fl}, ϵ_{fl}, and σ_{fl}, respectively.

Allard, J. F. (**1993**). *Propagation of Sound in Porous Media* (Elsevier Applied Science, London).

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

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

Biot, M. A., and Willis, D. G. (**1957**). "The elastic coefficients of the theory of consolidation," ASME J. Appl. Mech. **24**, 594–601.

Block, G. I., and Harris, J. G. (**2006**). "Conductivity dependence of seismoelectric wave phenomena in fluid-saturated sediments," J. Geophys. Res. **111**, B01304.

Bordes, C., Jouniaux, L., Dietrich, M., Pozzi, J.-P., and Garambois, S. (**2006**). "First laboratory measurements of seismo-magnetic conversions in fluid-filled Fontainebleau sand," Geophys. Res. Lett. **33**, L01302.

Butler, K. E., Russell, R. D., Kepic, A. W., and Maxwell, M. (**1996**). "Measurement of the seismoelectric response from a shallow boundary," Geophysics **61**, 1769–1778.

Denneman, A. I. M., Drijkoningen, G. G., Smeulders, D. M. J., and Wapenaar, K. (**2002**). "Reflection and transmission of waves at a fluid/porous-medium interface," Geophysics **67**, 282–291.

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

Dutta, N. C., and Odé, H. (**1983**). "Seismic reflections from a gas-water contact," Geophysics **48**, 148–162.

(**1988**). *Electromagnetic Methods in Applied Geophysics—Theory*, edited by M. N. Nabighian and J. D. Corbett (Society of Exploration Geophysicists, Tulsa, OK), Vol. **1**.

Garambois, S., and Dietrich, M. (**2001**). "Seismoelectric wave conversions in porous media: Field measurements and transfer function analysis," Geophysics **66**, 1417–1430.

Garambois, S., and Dietrich, M. (**2002**). "Full waveform numerical simulations of seismoelectromagnetic wave conversions in fluid-saturated stratified porous media," J. Geophys. Res. **107**, ESE5.1–ESE5.18.

Haartsen, M. W. (**1995**). "Coupled electromagnetic and acoustic wavefield modeling in poro-elastic media and its applications in geophysical exploration," Ph.D. thesis, Massachusetts Institute of Technology, Cambridge, MA.

Haartsen, M. W., and Pride, S. R. (**1997**). "Electroseismic waves from point sources in layered media," J. Geophys. Res. **102**, 24745–24769.

Haines, S. S., and Pride, S. R. (**2006**). "Seismoelectric numerical modeling on a grid," Geophysics **71**, N57–N65.

Haines, S. S., Pride, S. R., Klemperer, S. L., and Biondi, B. (**2007**). "Seismoelectric imaging of shallow targets," Geophysics **72**, G9–G20.

Han, Q., and Wang, Z. (**2001**). "Time-domain simulation of SH-wave-induced electromagnetic field in heterogeneous porous media: A fast finite-element algorithm," Geophysics **66**, 448–461.

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

Mikhailov, O. V., Haartsen, M. W., and Toksöz, M. N. (**1997**). "Electroseismic investigation of the shallow subsurface: Field measurements and numerical modeling," Geophysics **62**, 97–105.

Mikhailov, O. V., Queen, J., and Toksöz, M. N. (**2000**). "Using borehole electroseismic measurements to detect and characterize fractured (permeable) zones," Geophysics **65**, 1098–1112.

Pride, S. R. (**1994**). "Governing equations for the coupled electromagnetics and acoustics of porous media," Phys. Rev. B **50**, 15678–15696.

Pride, S. R., and Garambois, S. (**2005**). "Electroseismic wave theory of Frenkel and more recent developments," J. Eng. Mech. **131**, 898–907.

Pride, S. R., and Haartsen, M. W. (**1996**). "Electroseismic wave properties," J. Acoust. Soc. Am. **100**, 1301–1315.

Santos, J. E., Corbero, J. M., Ravazzoli, C. L., and Hensley, J. L. (**1992**). "Reflection and transmission coefficients in fluid-saturated porous media," J. Acoust. Soc. Am. **91**, 1911–1923.

Thompson, A. H., and Gist, G. A. (**1993**). "Geophysical applications of electrokinetic conversion," The Leading Edge **12**, 1169–1173.

Thompson, A. H., Hornbostel, S., Burns, J., Murray, T., Raschke, R., Wride, J., McCammon, P., Sumner, J., Haake, G., Bixby, M., Ross, W., White, B. S., Zhou, M., and Peczak, P. (**2007**). "Field tests of electroseismic hydrocarbon detection," Geophysics **72**, N1–N9.

White, B. S., and Zhou, M. (**2006**). "Electroseismic prospecting in layered media," SIAM J. Appl. Math. **67**, 69–98.

Zhu, Z., Toksöz, M. N., and Burns, D. R. (**2008**). "Electroseismic and seismoelectric measurements of rock samples in a water tank," Geophysics **73**, E153–E164.

On Dowell's simplification for acoustic cavity-structure interaction and consistent alternatives

Jerry H. Ginsberg
G. W. Woodruff School of Mechanical Engineering, Georgia Institute of Technology, Atlanta, Georgia 30332-0405

(Received 22 September 2009; accepted 4 November 2009)

A widely employed description of the acoustical response in a cavity whose walls are compliant, which was first proposed by Dowell and Voss [(1962). AIAA J. **1**, 476–477], uses the modes of the corresponding cavity with rigid walls as basis functions for a series representation of the pressure. It yields a velocity field that is not compatible with the movement of the boundary, and the system equations do not satisfy the principle of reciprocity. The simplified formulation is compared to consistent solutions of the coupled field equations in the time and frequency domains. In addition, this paper introduces an extension of the Ritz series method to fluid-structure coupled systems that satisfies all continuity conditions by imposing constraint equations to enforce any such conditions that are not identically satisfied by the series. A slender waveguide terminated by an oscillator is analyzed by each method. The simplified formulation is found to be very accurate for light fluid loading, except for the pressure field at frequencies below the fundamental rigid-cavity resonance, whereas the Ritz series solution is found to be extremely accurate in all cases.
 [DOI: 10.1121/1.3268606]

PACS number(s): 43.20.Tb, 43.55.Br [ADP] Pages: 22–32

I. INTRODUCTION

A fully consistent model of the interaction of a structure in motion and a fluid must account for the surface tractions exerted by the fluid on the structure, while it simultaneously enforces continuity of the motions of each medium at their interface. Linearized acoustics for inviscid fluids reduces surface tractions to a pressure loading, and only velocity continuity in the direction normal to the interface is considered. Furthermore, this interface is taken to be situated at the original location prior to structural displacement. Even with these simplifications, the interaction problem can be quite daunting. In part, the problem stems from the complexity of the structural dynamic laws for all but the simplest systems. In addition, real systems seldom have a configuration that is suitable for solution of the wave equation by classical mathematical methods. The development and implementation of computational methods that are not limited by a domain's shape, notably boundary and finite elements, is an active and widely engaged subject. It is not the purpose here to add to that line of research. Rather, the focus is on analytical techniques for deriving equations governing the coupled response.

A simplified method for determining the acoustic field in an enclosed cavity was proposed more than 45 years ago by Dowell and Voss (1962). It has been implemented in diverse applications, including active noise control (Gopinathan *et al.*, 2001). The most recent paper in this journal to rely on the simplification appears to be the work by Venketasham *et al.* (2008), who analyzed sound radiating from a rectangular cavity. The simplification stems from the usage of the acoustic modes of the same cavity, but with rigid walls, as a set of basis functions for a series representation of the pressure field. Such a representation cannot consistently satisfy the velocity continuity condition at movable boundaries. Instead, the formulation uses Green's theorem to incorporate the overall effect of the surface velocity into the governing pressure equations. Davis *et al.* (2008) surveyed the method and applied it to liquid oxygen tanks. They found that the system equations lack symmetry, which raised the degree of computational effort required for their modal solution. Asymmetry also is a source of concern from a theoretical standpoint because it means that the equations describe a model that does not obey the law of reciprocity (Pierce, 1993).

Dowell *et al.* (1977) noted that the response derived with the simplified analysis will not actually satisfy velocity continuity, with references to Dowell's (1974) aeroelasticity text and the text by Morse and Ingard (1968) to explain why this is acceptable. Ultimately the justification lies on the following reasoning: If the structure is sufficiently stiff, then structural velocities, and therefore the normal component of the fluid's particle velocity field at the interface, will be much less than the overall maxima observed throughout the field. Correspondingly, the normal particle acceleration at the interface should be comparatively small, which according to Euler's momentum equation means that the actual pressure field should have a small gradient of pressure normal to the interface. This leads to the expectation that the actual pressure field should strongly resemble the pressure field obtained from an analysis that *a priori* sets to zero the normal derivative at the interface. One is left to conjecture when a structure is sufficiently stiff to justify the simplification. Furthermore, a structure's compliance is frequency dependent, and modern structures, through the use of engineered materials, are becoming increasingly lightweight.

The present investigation employs a one-dimensional waveguide terminated by a one-degree-of-freedom oscillator as a prototypical system that serves multiple purposes. One

objective is to compare the procedure entailed in using Dowell's simplification to the mathematical operations required to solve the field equations. The analytical solutions are a time-domain modal analysis of the wave equation and a frequency domain solution of the Helmholtz equation, each accompanied by boundary conditions that enforce velocity continuity and transfer the pressure load to the oscillator. A fully coupled modal analysis is likely to be forbidding for multidimensional systems, but frequency domain solution of the Helmholtz equation subject to a Robin-type boundary condition is widely feasible and results of such analyses can be converted to the time domain by inverse fast Fourier transform (FFT) processing. An assessment of the accuracy of Dowell's simplification, which is the second of the present objectives, is obtained by evaluating the complex frequency response derived by each method. Although the prototypical system contains the fundamental features of acoustic-structure interaction, it is described by only two nondimensional system parameters, which makes it easier to identify trends.

The third objective of the present work is to introduce an extension of the general Ritz series formalism (Ritz, 1908) to address the coupled response of a structure and an acoustic cavity. It is a time-domain method that consistently accounts for all interaction effects. Implementation of a Ritz series analysis requires a fundamental variational principle, which is Hamilton's principle in the present context. Lagrange multipliers are used to account for the forces that impose any continuity constraints that are not implicitly satisfied by the Ritz series. The reader is referred to the article by Pierce (1993) for the derivation of Hamilton's principle for acoustics, and a general derivation that includes Lagrange multipliers may be found in the text by Ginsberg (2008). The Ritz series technique has been widely employed to study acoustical and structural systems individually, but much less so for coupled problems. The variational formulation developed here is most closely related to the work by Moussou (2005), with some important generalizations, notably assurance that the simulation represents an irrotational fluid flow, and allowance for continuity conditions to be explicitly stated as auxiliary conditions.

II. THE SIMPLIFIED FORMULATION

The one-dimensional waveguide in Fig. 1 captures the fundamental issues for fluid-structure interaction. A narrow tube having cross-sectional area A is rigidly terminated at $x=L$, while a movable piston of mass M supported by spring K and driven by force F closes $x=0$. The time-domain equations to be satisfied are the wave equation and the oscillator's equation of motion, accompanied by acceleration boundary conditions,

$$\frac{1}{c^2}\frac{\partial^2 p}{\partial t^2}-\frac{\partial^2 p}{\partial x^2}=0,$$

$$\left.\frac{\partial p}{\partial x}\right|_{x=L}=0,$$

$$\left.\frac{\partial p}{\partial x}\right|_{x=0}=-\rho\ddot{w},\quad M\ddot{w}+Kw=F-\left.p\right|_{x=0}A. \tag{1}$$

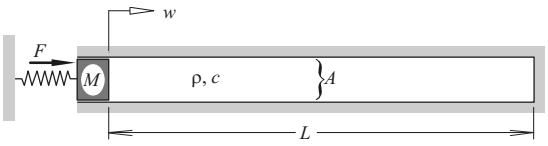


FIG. 1. A one-dimensional waveguide terminated by a one-degree-of-freedom oscillator.

The pressure is coupled to the oscillator's motion by the first of the conditions at $x=0$, which imposes continuity of acceleration between the piston and the fluid particles, and by the pressure loading applied to the oscillator.

The technique proposed by Dowell *et al.* (1962, 1977) expands the pressure in a series whose basis functions are the mode functions $\Phi_{R,n}$ the cavity would have if all walls were rigid,

$$p=\rho c^2\sum_{n=1}^{\infty}P_n(t)\Phi_{R,n}(x). \tag{2}$$

The fundamental properties of the rigid-cavity modes for this system are

$$\frac{\partial^2\Phi_{R,n}}{\partial x^2}=-\frac{\Omega_{R,n}^2}{c^2}\Phi_{R,n},$$

$$\left.\frac{\partial\Phi_{R,n}}{\partial x}\right|_{x=0}=\left.\frac{\partial\Phi_{R,n}}{\partial x}\right|_{x=L}=0,$$

$$\int_0^L(\Phi_{R,n})^2dx=L, \tag{3}$$

where the choice of L as the normalizing value in the last equation is a convenient way of nondimensionalizing $\Phi_{R,n}$. The functions satisfying these equations are

$$\Phi_{R,n}=\sqrt{2}\cos\left(\frac{n\pi x}{L}\right),\quad \Omega_{R,n}=n\pi\frac{c}{L}. \tag{4}$$

At this juncture, it is obvious that there is a dilemma, in that using Eq. (2) to evaluate $\partial p/\partial x$ at $x=0$ will lead to a zero value. According to the first boundary condition at $x=0$, this means that $\ddot{w}=0$, which is contrary to the indication of the piston's equation of motion. This difficulty is circumvented by introducing Green's theorem, which in the present case of a one-dimensional system states that

$$\int_0^L\left(\Phi_{R,n}\frac{\partial^2 p}{\partial x^2}-p\frac{\partial^2\Phi_{R,n}}{\partial x^2}\right)dx=\left.\left(\Phi_{R,n}\frac{\partial p}{\partial x}-p\frac{\partial\Phi_{R,n}}{\partial x}\right)\right|_{x=0}^{x=L}. \tag{5}$$

Substitution of the boundary conditions in Eqs. (1) and (3) reduces the preceding to

$$\int_0^L \left(\Phi_{R,n} \frac{\partial^2 p}{\partial t^2} + \Omega_{R,n}^2 p \Phi_{R,n} \right) dx = \rho c^2 \, \Phi_{R,n}|_{x=0} \ddot{w}. \tag{6}$$

In effect, incorporating the acceleration continuity condition into Green's theorem accounts for it in an overall sense. The basic formulation concludes by observing that the rigid-cavity functions form an orthogonal set over $0<x<L$, so the pressure coefficients corresponding to any pressure distribution are given by

$$P_n = \frac{1}{\rho c^2 L} \int_0^L p \Phi_{R,n} dx. \tag{7}$$

This property converts Eq. (6) to

$$\ddot{P}_n + \Omega_{R,n}^2 P_n = \Phi_{R,n}|_{x=0} \frac{\ddot{w}}{L}, \quad n = 1, \ldots, N. \tag{8}$$

When the pressure series is truncated at N terms, the preceding leads to N differential equations for $N+1$ unknowns consisting of the P_m coefficients and w. The formulation is completed by including the piston's equation of motion, which becomes

$$M\ddot{w} + kw = -A\rho c^2 \sum_{n=1}^{N} \Phi_{R,n}|_{x=0} P_n + F. \tag{9}$$

In matrix notation these equations may be written as

$$\begin{bmatrix} [I] & -\{\Phi_R\} \\ [0] & \sigma \end{bmatrix} \frac{d^2}{dt^2} \begin{Bmatrix} \{P\} \\ w/L \end{Bmatrix} + \begin{bmatrix} [(\Omega_R)^2] & \{0\} \\ \dfrac{c^2}{L^2}\{\Phi_R\}^T & \sigma(\omega_s)^2 \end{bmatrix} \begin{Bmatrix} \{P\} \\ w/L \end{Bmatrix} = \begin{Bmatrix} \{0\} \\ F/\rho AL \end{Bmatrix}, \tag{10}$$

where

$$\sigma = \frac{M}{\rho AL}, \quad \omega_s = \left(\frac{K}{M} \right)^{1/2}. \tag{11}$$

The matrix form of the differential equations makes it evident that the system equations are not symmetric.

III. ANALYTICAL MODAL ANALYSIS

Analytical time-domain solutions become prohibitive, or even impossible, in multi-dimensional systems whose geometry is not commensurate with one of the coordinate systems in which the Helmholtz equation is separable, or if the boundary conditions are not independent of position along a boundary, or if the structural motion itself is difficult to analyze. None of these conditions applies in the case of the system in Fig. 1. The modes of this system are found by seeking a free vibration solution in which the undamped oscillator and the pressure in the waveguide vary harmonically at the same frequency. The analysis is expedited by a formulation in terms of the velocity potential,

$$v = \frac{\partial \phi}{\partial x}, \quad p = -\rho \dot{\phi} \tag{12}$$

so the free response is taken to be

$$\phi = \mathrm{Re}[-i\omega L^2 \Phi(x) \exp(-i\omega t)], \quad w = \mathrm{Re}[LW \exp(-i\omega t)]. \tag{13}$$

Note that factors are introduced in the preceding in order that Φ and W be real and dimensionless. This ansatz must satisfy the velocity potential version of Eq. (1) with $F=0$, which leads to

$$\frac{d^2\Phi}{dx^2} + k^2 \Phi = 0, \quad k = \omega/c,$$

$$\left. \frac{d\Phi}{dx} \right|_{x=L} = 0,$$

$$L \left. \frac{d\Phi}{dx} \right|_{x=0} = W, \quad \sigma(\omega_s^2 - \omega^2) W + A\omega^2 \Phi|_{x=0} = 0. \tag{14}$$

The solution of Helmholtz' equation consistent with the boundary condition at $x=L$ is

$$\Phi = B \cos[k(L-x)]. \tag{15}$$

The first of the above conditions at $x=0$, which enforces velocity continuity, then requires that

$$W = BkL \sin(kL). \tag{16}$$

The equation of motion of the oscillator, which is the second condition at $x=0$, then reduces to

$$\Delta(kl) \equiv \sigma \left[(kL)^2 - \left(\frac{\omega_s L}{c} \right)^2 \right] \sin(kL) - kL \cos(kL) = 0. \tag{17}$$

This is the characteristic equation, whose roots are proportional to the true natural frequencies Ω_n according to

$$\Delta(k_n L) = 0, \quad \Omega_n \equiv c k_n. \tag{18}$$

The corresponding mode function is

$$\Phi_n = B_n \cos\left[k_n L \left(1 - \frac{x}{L} \right) \right], \quad W_n = B_n k_n L \sin(k_n L). \tag{19}$$

At this juncture B_n is arbitrary, but it will be set when the modes are normalized.

The characteristic equation has an infinite number of roots. The transcendental nature of the equation requires that the roots be found numerically. Good estimates to begin such a computation may be obtained from a graphical representation of the characteristic equation, which may be written equivalently as

$$\tan(kL) = \frac{kL}{\sigma[(kL)^2 - (\omega_s L/c)^2]}. \tag{20}$$

Figure 2 depicts each side of Eq. (20) as a function of kL for several values of $\omega_s L/c$; the eigenvalues are the kL values at which the functions intersect. Much insight can be gained from an inspection of these graphs. An important property is that the function on the right side of Eq. (20) for fixed $\omega_s L/c$ rapidly approaches zero as $|kL - \omega_s L/c|$ increases. It follows that away from its singularity, the intersections of this function with $\tan(kL)$ are well approximated by the zeros of

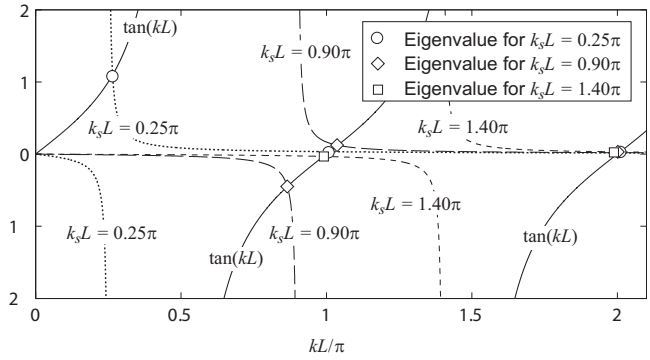


FIG. 2. Graphs of the left and right sides of the characteristic equation, Eq. (20), for three values of $\omega_s L/c$.

$\tan(kL)$, which are multiples of π. These limiting values are eigenvalues of the waveguide with rigid terminations. Figure 2 shows that there are alternate possibilities for eigenvalues in the vicinity of $\omega_s L/c$. If $\omega_s L/c$ is not close to a multiple of π, then there will be a single intersection of the two functions, which might be slightly smaller or larger than $\omega_s L/c$. In contrast, if $\omega_s L/c$ is sufficiently close to a multiple of π, the branches of the right-side function intersect $\tan(kL)$ on opposite sides of that multiple of π. Thus, if $\omega_s L/c \approx n\pi$, then there will be two close eigenvalues straddling $kL=n\pi$. The quantity on the right side of Eq. (20) decreases monotonically at fixed kL as $\sigma \equiv M/\rho AL$ increases, so these observed trends become increasingly accurate as the fluid becomes lighter.

Forced response of the system may be analyzed by using Φ_n and W_n paired set as basis functions for series representations of the pressure and displacement. Modal coordinates q_n describe the contribution of each mode function to the displacement, so the coefficients of the series for ϕ are the modal velocities $\dot{q}_n$. Thus, the modal series are

$$\phi(x,t) = L\sum_{n=1}^{\infty} \Phi_n \dot{q}_n(t), \quad w(t) = \sum_{n=1}^{\infty} W_n q_n(t),$$

$$v = \sum_{n=1}^{\infty} \frac{d\Phi_n}{d(x/L)} \dot{q}_n(t), \quad u = \sum_{n=1}^{\infty} \frac{d\Phi_n}{d(x/L)} q_n(t). \tag{21}$$

The pressure series is found from Eq. (12) to be

$$p = -\rho L \sum_{n=1}^{\infty} \Phi_n \ddot{q}_n(t). \tag{22}$$

The mode functions satisfy orthogonality conditions that are derived from the potential and kinetic energies of the system, which are

$$T = \frac{1}{2}\rho A \int_0^L v^2 dx + \frac{1}{2} M \dot{w}^2,$$

$$V = \frac{1}{2}\rho c^2 A \int_0^L \left(\frac{\partial u}{\partial x}\right)^2 dx + \frac{1}{2} K w^2. \tag{23}$$

The modal series for v and u are substituted into these expressions. The orthogonality conditions for the mode functions are obtained by setting to zero the coefficients that couple different generalized coordinates or velocities, thereby transforming T and V to sums of squares of the generalized velocities and coordinates, respectively (Ginsberg, 2001). [The alternate form $p^2/2\rho c^2$ is often used for the strain energy density. Doing so would be wrong in the present context because substitution of Eq. (22) would lead to a representation of V as a quadratic sum in the $\ddot{q}_n$ values. Such a description violates the definition of potential energy as a quantity that depends solely on a system's current position.] The energies that result may be written as

$$T = \frac{1}{2}\rho AL \sum_{n=1}^{\infty} (\dot{q}_n)^2, \quad V = \frac{1}{2}\rho AL \sum_{n=1}^{\infty} (\Omega_n)^2 (q_n)^2. \tag{24}$$

These forms are obtained when the following orthogonality conditions are satisfied:

$$\int_0^L \frac{d\Phi_j}{d(x/L)} \frac{d\Phi_n}{d(x/L)} d\left(\frac{x}{L}\right) + \sigma W_j W_n = \delta_{jn},$$

$$\left(\frac{c}{L}\right)^2 \int_0^L \frac{d^2\Phi_j}{d(x/L)^2} \frac{d^2\Phi_n}{d(x/L)^2} d\left(\frac{x}{L}\right) + \sigma(\omega_s)^2 W_j W_n = (\Omega_n)^2 \delta_{jn}. \tag{25}$$

The first of these relations in the case where $j=n$ defines the modal mass to be the mass of the fluid domain. This definition serves to normalize the B_n coefficients because substitution of the expressions in Eq. (19) into this normalization gives

$$\left[\frac{2k_n L + \sin(2k_n L)}{4k_n L} + \sigma \sin(k_n L)^2\right](B_n)^2 = 1. \tag{26}$$

The effect of the excitation F appears in the virtual work. The displacement of the oscillator mass in terms of the modal coordinates appears in Eq. (21), and a virtual displacement of the system is produced by differential increments of these coordinates. It follows that

$$\delta w = \sum_{n=1}^{\infty} W_n \delta q_n. \tag{27}$$

The virtual work is $F\delta w$, so the modal generalized forces are

$$Q_n = W_n F. \tag{28}$$

The equations of motion are derived by using these forces and the mechanical energies in Eq. (24) to form Lagrange's equations. The result is

$$\ddot{q}_n + (\Omega_n)^2 q_n = \frac{W_n F}{\rho AL}. \tag{29}$$

The uncoupled nature of these equations greatly simplifies the evaluation of transient response. If one is interested in response to nonzero initial conditions, the first condition in Eq. (25) can be used in conjunction with the modal series to map an initial field into the modal space. A detailed description of the procedure is provided by Ginsberg (2001).

IV. RITZ SERIES VARIATIONAL SOLUTION

A different approach for time-domain modeling of cavity-structure interaction is an extension of the Ritz series method (Ritz, 1908) to the coupled structural acoustic system. Ritz concept was to represent the response as a series whose terms are products of selected spatial functions that form a basis, and unknown time coefficients that describe the contribution of each basis function, and therefore serve as generalized coordinates. Equations for the coefficients are derived by enforcing a variational principle, which is Hamilton's principle in the present application. There are alternative techniques for enforcing the continuity conditions. The one followed here will enforce them explicitly through constraint equations and associated Lagrange multipliers. A general coordinate independent description of the method will be reported elsewhere. The work here serves to introduce the method by applying it to the one-dimensional waveguide.

The fundamental variable for Hamilton's principle is displacement. However, the method begins by describing the velocity potential by a Ritz series. This is done in order to assure that the fluid motion is irrotational, independently of the values of generalized coordinates. (This consideration would be identically satisfied in the one-dimensional waveguide by any series representation of the axial displacement. It is introduced here for the sake of consistency with the more general formulation.) A fundamental difference from the preceding modal analysis is that Eq. (22) used $p=-\rho\dot{\phi}$, which is derived from the momentum equation for the fluid domain. Incorporating this relation into a variational formulation based on Hamilton's principle would be wrong from a conceptual standpoint, because Hamilton's principle can be regarded as the fundamental axiom from which any relation between the force system and momentum can be extracted. The appropriate relation to employ instead is obtained from the continuity equation for a compressible inviscid fluid (Pierce, 1993),

$$p = -\rho c^2 \nabla \cdot \bar{u}_f = -\rho c^2 \frac{\partial u}{\partial x} \tag{30}$$

The Ritz series ansatz for the potential function requires selection of a set of basis functions $\psi_n(x)$. As was the done for the normal mode series, Eq. (21), the coefficients in this series coefficients are represented as generalized velocities $\dot{z}_n$ and the scaling factor L is used in order that ψ_n may be defined nondimensionally,

$$\phi = L\sum_{n=1}^{\infty} \psi_n \dot{z}_n(t). \tag{31}$$

The separation of space and time in each term of the series makes it possible to differentiate ϕ to obtain series for velocity and pressure, and then integrate the former to obtain a description of the displacement,

$$v = \sum_{n=1}^{\infty} \left(\frac{d}{d(x/L)} \psi_n \right) \dot{z}_n, \quad u = \sum_{n=1}^{\infty} \left(\frac{d}{d(x/L)} \psi_n \right) z_n,$$

$$p = -\frac{\rho c^2}{L} \sum_{n=1}^{\infty} \left(\frac{d^2}{d(x/L)^2} \psi_n \right) z_n. \tag{32}$$

In the terminology of classical mechanics, the z_n variables constitute generalized coordinates for the fluid.

There are few theoretical limitations on the choice of basis functions. A standard one is that they must constitute a linearly independent set. This can be achieved by selecting a set of functions having an indicial parameter n that corresponds to increasingly rapid variation as it increases. As is true for structural dynamics, rigid boundaries correspond to geometric boundary conditions. The rigid condition at $x=L$ is satisfied by requiring that $d\psi_n/dx=0$ at $x=L$. Conversely, $d\psi_n/dx$ should be nonzero at $x=0$ for at least some functions. Failure to do so would make it impossible to match the piston and acoustic motions at their interface. It is also necessary that at least some ψ_n be nonzero at $x=0$. This is imposed to avoid having a pressure series that vanishes at $x=0$, and therefore could not describe the acoustic force applied to the collar. It follows that the rigid-cavity modes used in the simplified formulation should not be used to form the Ritz series. A useful set for the present problem is

$$\psi_n = \cos\left[\frac{n\pi(L-x)}{2L} \right], \quad n = 0, 1, \ldots, N. \tag{33}$$

Because the present formulation has a description of displacement, the continuity condition at $x=0$ is restated as $w=u_f$ at $x=0$, which leads to a linear constraint condition,

$$\sum_{n=1}^{N} a_n z_n - w = 0, \quad a_n = \left. \frac{\partial \psi_n}{\partial x} \right|_{x=0}. \tag{34}$$

This constraint equation will be enforced explicitly, so w is henceforth treated as a variable that is independent of the z_n values. Substitution of the Ritz series into the energy functionals in Eq. (23) converts them to quadratic sums,

$$\mathcal{T} = \frac{1}{2}\rho AL \left[\sum_{j=1}^{N} \sum_{n=1}^{N} (M_f)_{jn} \dot{z}_j \dot{z}_n + \sigma \dot{w}^2 \right],$$

$$\mathcal{V} = \frac{1}{2}\rho AL \left[\sum_{j=1}^{N} \sum_{n=1}^{N} \frac{c^2}{L^2} (K_f)_{jn} z_j z_n + \sigma (\omega_s)^2 w^2 \right], \tag{35}$$

where the nondimensional matrix elements are

$$(M_f)_{jn} = \int_0^L \left(\frac{d}{d(x/L)} \psi_j \right) \left(\frac{d}{d(x/L)} \psi_n \right) d\left(\frac{x}{L} \right),$$

$$(K_f)_{jn} = \int_0^L \left(\frac{d^2}{d(x/L)^2} \psi_j \right) \left(\frac{d^2}{d(x/L)^2} \psi_n \right) d\left(\frac{x}{L} \right). \tag{36}$$

The displacement continuity condition, Eq. (34), is enforced by the force exerted between the piston and the fluid column. Consequently, it is manifested in Hamilton's principle as a Lagrange multiplier τ (Weinstock, 1974, Ginsberg 2008). Although τ actually is p $(x=0)$ A, that property is embedded in the principle. Because the Lagrange multiplier accounts for the interface forces, the virtual work accounts solely for F,

$$\delta \mathcal{W} = F \delta w. \tag{37}$$

The constrained Hamilton's principle states that

$$\int_{t_1}^{t_2} \left[\delta \mathcal{T} - \delta \mathcal{V} + \delta \mathcal{W} + \tau \left(\sum_{n=1}^{\infty} a_n \delta z_n - \delta w \right) \right] dt = 0. \tag{38}$$

A variational derivative is applied to the integrand, with the energy functionals substituted from Eq. (35). Integration by parts in time ultimately converts the integrand to a summation of terms in which each virtual increment δz_n and δw is multiplied by a factor that depends on the generalized coordinates and their time derivatives. The fundamental lemma of the calculus of variations requires that each factor of a virtual increment vanish because the increments are arbitrarily selected. For the fluid generalized coordinates, the result is

$$\rho A L \sum_{n=1}^{\infty} \left(M_{mn} \ddot{z}_n + \frac{c^2}{L^2} K_{mn} z_n \right) = \tau a_j, \quad j = 1,2, \ldots, \tag{39}$$

while setting the coefficient of δw to zero yields the oscillator's equation of motion,

$$M[\ddot{w} + (\omega_s)^2 w] = F - \tau. \tag{40}$$

Evaluation of the response requires solution of the differential equations of motion, Eqs. (39) and (40), subject to the constraint equation, Eq. (34). Such equations are said to be differential-algebraic (DAE) because τ does not occur in them as a derivative. A number of sophisticated methods have been developed to solve DAEs. A simple one suitable for the present case is an elimination method. Matrix notation is convenient for the elimination process, so the constraint coefficients are placed in $[a]$, which is a $1 \times N$ array. The constraint equation is

$$w = [a]\{z\}, \tag{41}$$

which transforms the equation of motion for the piston to

$$\tau = F - M[a]\{\ddot{z}\} - K[a]\{z\}. \tag{42}$$

Substitution of this expression into the matrix version of Eq. (39) leads to

$$[[M_f] + \sigma[a]^T[a]]\{\ddot{z}\} + \left[\frac{c^2}{L^2}[K_f] + \sigma(\omega_s)^2[a]^T[a] \right]\{z\} = [a]^T \frac{F}{\rho A L}. \tag{43}$$

The fluid's inertia and stiffness coefficients are $[M_f]$ and $(c^2/L^2)[K_f]$, respectively. This leads to the interpretation of $\sigma[a]^T[a]$ and $\sigma(\omega_s)^2[a]^T[a]$ as the respective effects added by structure. It should be noted that this set of system equations is symmetric.

V. FREQUENCY DOMAIN ANALYSIS

If the shape of a cavity conforms to the constant coordinate surfaces of a coordinate system in which the wave equation is separable, and if the boundary conditions along each such surface do not depend on location, then it is likely that frequency domain analysis is feasible. In the case of the present prototypical system, the result will be formulas for the response. The frequency domain ansatz represents the acoustic pressure, piston displacement, and excitation as $\mathrm{Re}[P(x)\exp(-i\omega t)]$, $\mathrm{Re}[W \exp(-i\omega t)]$, and $\mathrm{Re}[F' \exp(-i\omega t)]$, respectively. This converts the governing equations, Eqs. (1), to

$$\frac{d^2P}{dx^2} + k^2 P = 0, \quad k \equiv \frac{\omega}{c},$$

$$\left. \frac{dP}{dx} \right|_{x=L} = 0,$$

$$\left. \frac{dP}{dx} \right|_{x=0} = \rho \omega^2 w, \quad M(\omega^2 - \omega_s^2) w = P|_{x=0} A - F'. \tag{44}$$

The continuity requirement is satisfied implicitly by using it to formulate the piston's equation of motion, which converts the last equation to a Robin-type boundary condition,

$$\frac{M}{\rho} \left(\frac{\omega^2 - \omega_s^2}{\omega^2} \right) \left. \frac{dP}{dx} \right|_{x=0} = P|_{x=0} A - F'. \tag{45}$$

The solution of Helmholtz equation consistent with the boundary condition at $x=L$ is

$$P = D \cos(k(L-x)) \tag{46}$$

This general solution satisfies Eq. (45) if

$$D = -\frac{kL}{\Delta} \frac{F'}{A}, \tag{47}$$

where Δ is the function of kL forming the modal characteristic equation, Eq. (17). The corresponding displacement is

$$W = \frac{\sin(kL)}{\Delta} \frac{F' L}{\rho c^2 A}. \tag{48}$$

It is evident that resonances occur when $\Delta = 0$, that is, when kL equals a nondimensional natural frequency.

VI. ASSESSMENT

The frequency domain version of the simplified analysis may be obtained by setting

$$P_n = \rho c^2 \, \mathrm{Re}(Y_n \exp(-i\omega t)), \quad w = L \, \mathrm{Re}(W \exp(-i\omega t)) \tag{49}$$

in Eqs. (8) and (9), which leads to

$$[(n\pi)^2 - (kL)^2] Y_n = -\sqrt{2}(kL)^2 \frac{W}{L},$$

$$\sigma \left[\left(\frac{\omega_s L}{c} \right)^2 - (kL)^2 \right] \frac{W}{L} + \sqrt{2} \sum_{n=1}^{N} Y_n = \frac{F'}{\rho c^2 A}. \tag{50}$$

An expression for W is obtained by solving the first equation for Y_n, then substituting that result into the second equation, which yields

$$\frac{W}{L} = \frac{F'}{\rho c^2 A} \left(\frac{1}{S} \right), \quad Y_n = -\sqrt{2} \frac{F'}{\rho c^2 A} \frac{1}{[(kL)^2 - (n\pi)^2] S},$$

$$S = \sum_{n=1}^{N} \frac{2(kL)^2}{(kL)^2 - (n\pi)^2} + \frac{M}{\rho AL}\left[\left(\frac{\omega_s L}{c}\right)^2 - (kL)^2\right]. \quad (51)$$

One term in the denominator of Y_n vanishes when kL is a multiple of π, but that does not represent a resonance because the sum contained in S is singular in that case. Rather the simplified analysis displays a resonance at any kL for which $S=0$.

The uncoupled nature of the modal equations leads to simple expressions for the complex frequency response. The harmonic response is represented by setting

$$q_n = \mathrm{Re}(Y_n \exp(-i\omega t)). \quad (52)$$

An expression for Y_n is obtained as the steady-state solution of Eq. (29), which may be written as

$$Y_n = \frac{W_n}{[(k_n L)^2 - (kL)^2]} \frac{F'L}{\rho c^2 A}. \quad (53)$$

Series representations of the corresponding pressure field and oscillator displacement may be found by substituting Eqs. (52) and (53) into Eqs. (22) and (21). When these series converge, they yield identical results, aside from numerical errors, as the frequency domain analysis because both analyses solve the same set of field equations without approximation.

Frequency domain solution of the differential equations governing the Ritz generalized coordinates is implemented by setting

$$z_n = \mathrm{Re}(Z_n \exp(-i\omega t)), \quad w = \mathrm{Re}(W \exp(-i\omega t)). \quad (54)$$

This representation is substituted into the reduced set of equations, Eq. (43), which yields

$$\{Z\} = [[\hat{K}] - (kL)^2[\hat{M}]]^{-1}[a]^T \frac{F'L}{\rho c^2 A}, \quad W = [a]\{Z\}, \quad (55)$$

where

$$[\hat{M}] = [M_f] + \sigma[a]^T[a],$$

$$[\hat{K}] = [K_f] + \sigma\left(\frac{\omega_s L}{c}\right)^2 [a]^T[a]. \quad (56)$$

Note here that although the solution for $\{Z\}$ is described in terms of a matrix inverse, an efficient method like Gauss elimination would be used in practice.

An important quantity to examine is $|W|/L$ as a function of kL with $F'/\rho c^2 A$ set to unity. Rather than evaluating the pressure at a number of points, a concise metric of the pressure field is the mean-squared pressure averaged over $0<x<L$, that is,

$$(p^2)_{\mathrm{av}} = \frac{1}{L}\int_0^L \frac{1}{2} P(x)P(x)^* dx, \quad (57)$$

where $P(x)$ is the complex amplitude of the pressure distribution. The expression in Eq. (46) gives

$$(p^2_{\mathrm{exact}})_{\mathrm{av}} = \frac{kL[2kL + \sin(2kL)]}{8\Delta^2}\left(\frac{|F|}{A}\right)^2. \quad (58)$$

The consequence of the orthogonality of the rigid-cavity modes is that the average mean-squared pressure obtained from Eq. (2) is

$$(p^2_{\mathrm{approx}})_{\mathrm{av}} = \frac{1}{2}(\rho c^2)^2 \sum_{j=1}^{N} |Y_j|^2. \quad (59)$$

Evaluation of mean-squared pressure associated with the Ritz series begins by using Eqs. (32) and (54) to form the complex amplitude of the pressure distribution,

$$P(x) = -\rho c^2 \sum_{n=1}^{\infty} \left(\frac{d^2}{d(x/L)^2}\psi_n\right)\frac{Z_n}{L}. \quad (60)$$

Substitution of this expression into Eq. (57) yields

$$(p^2_{\mathrm{Ritz}})_{\mathrm{av}} = \frac{1}{2}(\rho c^2)^2 \sum_{j=1}^{N}\sum_{n=1}^{N} (K_f)_{jn}\frac{Z_j Z_n^*}{L^2}$$
$$= \frac{1}{2}(\rho c^2)\left\{\frac{Z}{L}\right\}^T [K_f]\left\{\frac{Z}{L}\right\}^*. \quad (61)$$

A comparison of Eq. (61) with the second of Eqs. (35) shows that this form is proportional to the mean-squared potential energy, which is not surprising because an alternative form of the potential energy density is $p^2/2\rho c^2$. This brings to the fore another reason why $p=-\rho\dot{\phi}$ should not be used in any context of the variational analysis. It could not be used to form the potential energy functional because doing so would have produced a potential energy that depends on generalized accelerations. If it were used here to evaluate the mean-squared pressure, the outcome would exhibit substantial differences with Eq. (58), 6 dB or more, in the regions between resonances. This discrepancy between pressure evaluations apparently arises because $p=-\rho\dot{\phi}$ is derived from the momentum principle, which Hamilton's principle only enforces in a weak sense. [It is interesting to observe that internal stresses associated with a structural dynamics solution for displacement are always computed from stress-strain relations, which are analogous to Eq. (30).]

The assumptions underlying Dowell's method should become increasingly valid with increasing structural mass, so the first set of data describe $M/\rho AL=100$. All evaluations were carried out using eight terms for Dowell's simplification and the Ritz series, both of which are more than necessary to obtain convergent results in the frequency interval below $kL=2.5\pi$ in which three resonances occur. The Ritz basis functions for these evaluations are those in Eq. (33), but others were verified to give equally good results. In all evaluations the Ritz series results were indistinguishable from the frequency domain solution. Past Ritz series analyses of structural dynamics have occasionally encountered difficulties associated with ill-conditioning as the length of a Ritz series is increased. It is possible that the same phenomenon would have been encountered here if a longer series was used in order to cover a broader frequency range. This situation, which would be marked by $[M_f]$ and $[K_f]$ having

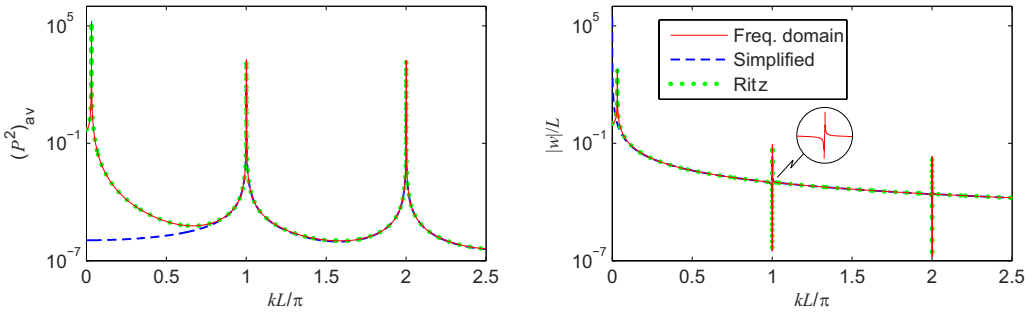


FIG. 3. (Color online) Frequency dependence of the mean-squared pressure averaged over the length of the waveguide and of the displacement amplitude of the piston, $M/\rho AL=100$ and $\omega_s L/c=0$.

extraordinarily large condition numbers, arises because the basis functions cease to be independent in a computational sense, even though they are mathematically linearly independent (Ginsberg, 2001). Such situations may be addressed with singular value decomposition or by changing the set of basis function.

Figures 3–6 compare the average mean-squared pressure and the displacement of the piston for a sequence of $\omega_s L/c$ values commencing from zero, which corresponds to an unsupported piston. All peaks in the graph would rise without limit were it not for roundoff error; energy dissipation in either the fluid or structure would limit their height. These peaks are the resonances at which kL equals a natural frequency. The two peaks that occur close to $kL=\pi$ in Fig. 5 ($\omega_s L/c=\pi$) are manifestations of the general discussion of Eq. (20), where it was noted that if $\omega_s L/c$ is close to $n\pi$, then two natural frequencies will occur slightly above and below $kL=n\pi$. An interesting feature of the graphs of displacement is the occurrence of an anti-resonance when kL is a multiple of π. To understand this phenomenon recall that when σ is reasonably large, the coupled system resonances occur very close to multiples of π. However, Eq. (48) indicates that W is proportional to $\sin(kL)$, so a displacement node occurs at the piston's face when $kL=n\pi$.

An overview of Figs. 3–6 indicates that Dowell's simplification yields results that are usually in close agreement with the consistent solutions, but there are some significant discrepancies. In Fig. 3 there is a very low-frequency resonance that the approximate solution does not exhibit for pressure and places at zero frequency for displacement. The smallness of this resonant frequency means that the pressure amplitude within the waveguide is nearly independent of x, a fact that becomes evident when one writes the cosine term in Eq. (46) as $\cos(kL(1-x/L))$. Because $\omega_s L/c=0$ corresponds to removal of the spring K, it follows that this low-frequency resonance is analogous to a Helmholtz resonance, in which the compressed fluid encounters inertial resistance at the cavity opening, except that here the inertia is that of the piston, rather than the fluid mass in the neck of the opening.

Figures 4–6, for which $\omega_s L/c$ is not small, indicate that the simplified formulation produces excellent results for the displacement across the entire frequency spectrum. Although it greatly underpredicts the low-frequency pressure, it does very well at predicting the pressure in the frequency range of the lowest resonance and upward, regardless of the value of $\omega_s L/c$. Reduction in the mass ratio to $\sigma=1$ still leads to generally good agreement between the simplified analysis and the consistent solutions. Figure 7, for which $\omega_s L/c=0.5\pi$, is representative. It can be seen that the simplified analysis still underpredicts the low frequency pressure distribution. The resonant frequencies are lower than the corresponding true values, but the discrepancies lessen with increasing resonant frequency.

VII. DISCUSSION

In 1962 when Dowell and Voss (1962) first used the rigid-cavity modes as basis functions for the pressure field, doing so was justifiable because of the absence of fully developed numerical techniques, as well as the relatively limited computational resources that were then available. From its inception it was recognized that although the formulation does incorporate the effect of boundary motion, solutions obtained by this method do not actually satisfy continuity of velocity at the fluid-structure interface. The physical argument supporting the approximation, specifically that the in-

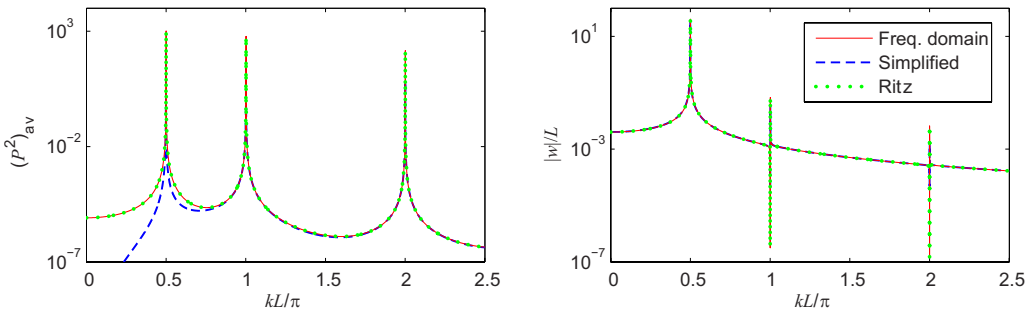


FIG. 4. (Color online) Frequency dependence of the mean-squared pressure averaged over the length of the waveguide and of the displacement amplitude of the piston, $M/\rho AL=100$ and $\omega_s L/c=0.5\pi$.

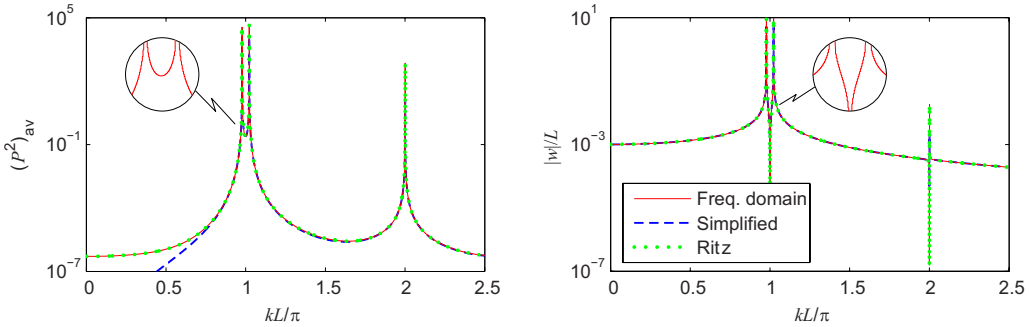


FIG. 5. (Color online) Frequency dependence of the mean-squared pressure averaged over the length of the waveguide and of the displacement amplitude of the piston, $M/\rho AL=100$ and $\omega_s L/c=\pi$.

ternal pressure field should be close to that of the cavity with rigid walls if the walls are sufficiently stiff, is reasonable, but any solution that disagrees with the basic principle of continuity could be quite inaccurate in some situations.

To explore whether such concerns are warranted, a prototypical system consisting of a long waveguide with one end terminated by a one-degree-of-freedom oscillator was used to compare the simplified analysis to three consistent approaches. A time-domain solution of the wave equation for the velocity potential subject to boundary conditions that couple the acoustic and structural responses led to a modal series representation of the coupled pressure field and displacement. This type of solution was feasible because of the one-dimensional nature of the prototypical system, but its extension to multi-dimensional problems in which the fluid-structure interface is a surface is likely to present a formidable task. This is not the case for the second time-domain formulation, which was a Ritz series analysis of the coupled waveguide and oscillator. The Ritz series formulations requires little *a priori* analysis, in that the only requirement is identification of suitable basis functions. The presentation of this formulation was intended to introduce it and demonstrate its validity, but it is readily extended to address a diverse range of multidimensional systems; a general exposition of it will be provided later. Frequency domain solution of the Helmholtz equation was the third consistent formulation described here. Because of the one-dimensional nature of system, this analysis yielded formulas for the complex pressure distribution and oscillator response. In general, replacement of time derivatives by algebraic powers of frequency makes it possible to employ the approach for many multidimensional systems. Furthermore, frequency domain solutions can be converted to the time domain through the application of the inverse FFT algorithm. It is not clear whether analysis of a specific system via Dowell's simplification is easier than a consistent frequency domain solution, but it is this author's opinion that they are equally feasible.

The level of difficulty for any analytical method is significantly increased when the shape of the acoustic cavity is not well described by one of the standard coordinate systems in which the wave equation is separable. However, the various analytical methods are not equally applicable to cavities whose shape conforms to this specification. Dowell's simplification only addresses acoustic cavities bounded by solid surfaces, but some systems feature a boundary on which the pressure is specified, for example, an opening that is exposed to an exterior fluid domain. This limitation becomes evident when one examines the basic derivation underlying the simplification, which is based on Green's theorem for the fluid domain. Wherever the pressure occurs in a term evaluated on the surface, it is multiplied by the normal derivative of the rigid-cavity mode, which is zero. Thus, there is no way to incorporate a known surface pressure into the formulation. Of course, one could attempt to modify the method by using other types of modes. Formal mathematical solutions of field equations, either in the time or frequency domain, are not feasible in situations where the boundaries of the cavity consists of surfaces along which the conditions depend on position. An example of such a situation is an elastic plate that is situated in a portion of a bounding plane, with another portion of that boundary being a rigid baffle or an open port. One alternative in such situations would be to patch a solution using component mode synthesis, which Magalhaes and Ferguson (2005) applied to cavity acoustics. Another alternative is to implement the Ritz series formalism outlined here, but as noted, that is a topic for future exploration. Of course,

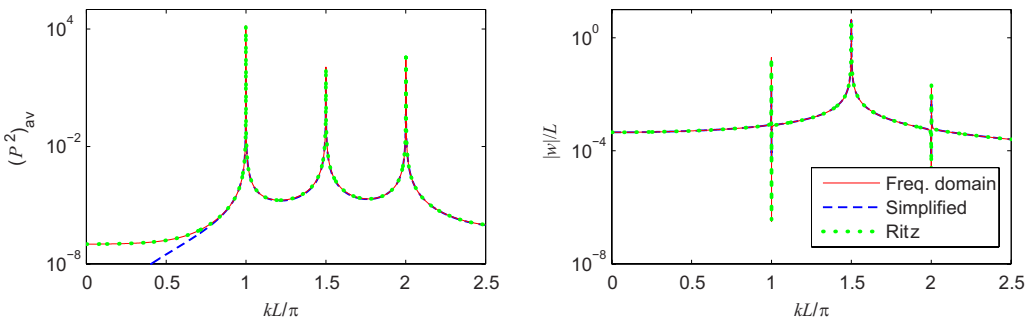


FIG. 6. (Color online) Frequency dependence of the mean-squared pressure averaged over the length of the waveguide and of the displacement amplitude of the piston, $M/\rho AL=100$ and $\omega_s L/c=1.5\pi$.

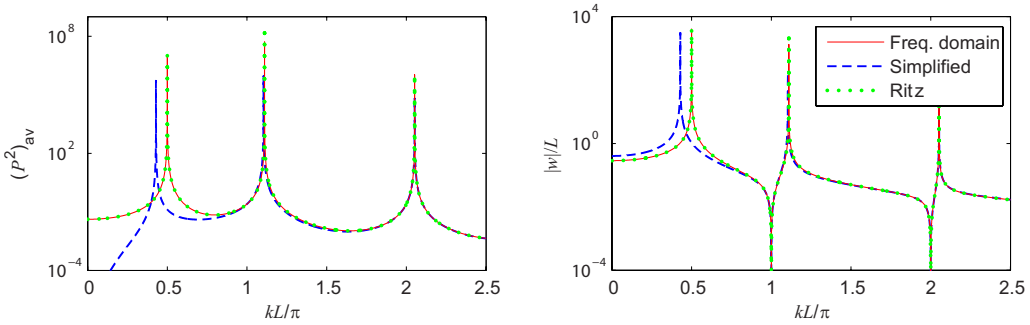


FIG. 7. (Color online) Frequency dependence of the mean-squared pressure averaged over the length of the waveguide and of the displacement amplitude of the piston, $M/\rho AL=1$ and $\omega_s L/c=0.5\pi$.

boundary and finite elements are readily implemented in either the frequency or time domain to describe interior cavities that are intractable to analysis.

A general aspect of the simplified formulation noted here and by Davis *et al.* (2008) is that the associated system equations are not symmetric. This is most evident in Eq. (10) and may be traced to the fact that the boundary term in the Green's function leading to the pressure coefficient equations depends on surface acceleration (a second derivative), whereas the equation of motion for the structure depends on the current pressure, not its second derivative, so that neither the effective inertia nor stiffness matrix is symmetric. For Davis *et al.* (2008), this was a concern because it complicated the solution of the time-domain equations, but there is a more troubling aspect. A general derivation of the inertia and stiffness matrices corresponding to the kinetic and potential energies of any system discloses that both must be symmetric if the system is linear and time invariant (Ginsberg, 2008). Asymmetry can only result from nonconservative forces, such as follower and feedback forces, which are known to destabilize system if their effect is sufficiently large. (Nonconservative here refers to forces that are not derivable from a potential energy function.) Any linear model whose system equations are not symmetric does not obey the principle of reciprocity (Pierce, 1981), but acoustic-structure interaction has been proven to be a reciprocal process (Lyamshev, 1959; Chen and Ginsberg, 1995). The consequences of not satisfying reciprocity were not evident in the current assessment, which focused on the amplitude of complex response variables, but it might be seen in an examination of phase differences. Furthermore, such phenomena might become an issue for transient time-domain solutions of the coefficient equations, and it would not be wise to employ the simplified formulation in conjunction with experimental methods that rely on reciprocity.

An overview of the computed results showed that Dowell's simplification yields results that are essentially correct when the oscillator's mass is large. Discrepancies were found to be failure of the approximate solution to describe a low-frequency Helmholtz-type resonance that occurs when the mass is unsupported by a spring, and under-prediction of the pressure amplitude in the frequency range below the first resonance. It is not surprising that the quality of the simplified analysis improves with increasing frequency, because the impedance of the oscillator increases rapidly. However, the fact that raising the structural stiffness did not improve the low-frequency results suggests that the effective stiffness of the fluid affects the validity of the simplification. When the fluid mass is comparable to the oscillator's, the primary additional errors in the approximate results were found to stem from deviations of the resonant frequencies, which lessened with increasing frequency. In contrast, the results obtained from the Ritz series formulation were indistinguishable from those derived by solving Helmholtz' equation, regardless of the oscillator's properties.

At this juncture, one is left to ponder the basic question: Why select an approximate analytical technique for structural acoustic modeling when consistent alternatives are available? It is true that the prototypical system's response obtained from the approximate formulation were generally quite good, which suggests that prior work based on Dowell's method probably obtained reasonably accurate results. The discrepancies that were found might not be important for airborne structures (light fluid loading) because the pressure levels in the low-frequency range, where the simplification breaks down, are very low. However, the important words in the preceding sentences are "probably" and "might," in that the trends observed here do not necessarily apply to other systems. From a philosophical viewpoint, it is hard to justify usage of any formulation of interaction problems that violates the principles of velocity continuity and reciprocity, especially in mission-critical applications.

Chen, P. T., and Ginsberg, J. H. (**1995**). "Complex power, reciprocity, and radiation modes for submerged bodies," J. Acoust. Soc. Am. **98**, 3343–3351.

Davis, R. B., Virgin, L. N., and Brown, A. M. (**2008**). "Cylindrical shell submerged in bounded acoustic media: A modal approach," AIAA J. **46**, 752–763.

Dowell, E. H. (**1974**). *Aeroelasticity of Plates and Shells* (Nordhoff, Leyden).

Dowell, E. H., and Voss, H. M. (**1962**). "The effect of a cavity on panel vibration," AIAA J. **1**, 476–477.

Dowell, E. H., Gorman, G. F., and Smith, D. A. (**1977**). "Acoustoelasticity: General theory, acoustic natural modes and forced response to sinusoidal excitation, including comparisons with experiment," J. Sound Vib. **52**, 519–542.

Ginsberg, J. H. (**2001**). *Mechanical and Structural Vibration* (Wiley, New York) Chap. 6.

Ginsberg, J. H. (**2008**). *Engineering Dynamics* (Cambridge University Press, New York) Sec. 7.1.

Gopinathan, S. V., Varadan, V. V., and Varadan, V. K. (**2001**). "Rayleigh-Ritz formulation for active control of the acoustics of cabin enclosures," Proc. SPIE **4326**, 58–67.

Lyamshev, M. (**1959**). "A question in connection with the principle of reciprocity in acoustics," Sov. Phys. Dokl. **84**, 405–409.

Magalhaes, M. D. C., and Ferguson, N. S. (**2005**). "The development of Component Mode Synthesis (CMS) for three-dimensional fluid-structure interaction," J. Acoust. Soc. Am. **118**, 3679–3690.
Morse, P. M., and Ingard, K. U. (**1968**). *Theoretical Acoustics* (McGraw-Hill, New York).
Moussou, P. (**2005**). "A kinematic method for the computation of fluid-structure interaction systems," J. Fluids Struct. **20**, 643–658.
Pierce, A. D. (**1981**). *Acoustics* (McGraw-Hill, New York), pp. 195–203.
Pierce, A. D. (**1993**). "Variational formulations in acoustic radiation and scattering," in *Underwater Scattering and Radiation*, edited by A. D. Pierce and R. N. Thurston (Academic, San Diego).
Ritz, W. (**1908**). "Über eine neue methode zur lösung gewisser variationsprobleme der mathematischen physik" ("On a new method for solving some variational problems of mathematical physics"), J. Reine Angew. Math. **1**, 1–61.
Venkatesham, B., Tiwari, M., and Munjal, M. L. (**2008**). "Analytical prediction of the breakout noise from a rectangular cavity with one compliant wall," J. Acoust. Soc. Am. **124**, 2952–2962.
Weinstock, R. (**1974**). *Calculus of Variations* (Dover, New York).

An iterative method for the computation of nonlinear, wide-angle, pulsed acoustic fields of medical diagnostic transducers

J. Huijssen[a] and M. D. Verweij[b]
Laboratory of Electromagnetic Research, Faculty of Electrical Engineering, Mathematics and Computer Science, Delft University of Technology, Mekelweg 4, 2628 CD Delft, The Netherlands

(Received 25 June 2009; revised 15 October 2009; accepted 27 October 2009)

The development and optimization of medical ultrasound transducers and imaging modalities require a computational method that accurately predicts the nonlinear acoustic pressure field. A prospective method should provide the wide-angle, pulsed field emitted by an arbitrary planar source distribution and propagating in a three-dimensional, large scale domain holding a nonlinear acoustic medium. In this paper, a method is presented that is free of any assumed wavefield directionality. The nonlinear acoustic wave equation is solved by treating the nonlinear term as a contrast source. This formulation leads to an iterative scheme that involves the repetitive solution of a linear wave problem through Green's function method. It is shown that accurate field predictions may be obtained within a few iterations. Moreover, by employing a dedicated numerical convolution technique, the method allows for a discretization down to two points per wavelength or period of the highest frequency of interest. The performance of the method is evaluated through a number of nonlinear field predictions for pulsed transducers with various geometries. The results demonstrate the directional independence of the method. Moreover, comparison with results from several existing methods shows that the method accurately predicts the nonlinear field for weak to moderate nonlinearity. © *2010 Acoustical Society of America.* [DOI: 10.1121/1.3268599]

PACS number(s): 43.25.Cb, 43.80.Qf [TDM] Pages: 33–44

I. INTRODUCTION

Current-day medical diagnostic ultrasound scanners benefit from the nonlinear distortion of the acoustic wavefield that arises when the field propagates through human tissue. The imaging modality that is called Tissue Harmonic Imaging (THI) is known to significantly improve the quality of the echographic image by specifically imaging the second harmonic frequency component that arises from the nonlinear propagation in tissue.[1,2] Recent studies show that the nonlinear distortion effect may be exploited further by employing the third to the fifth harmonic frequency components[3,4] and by optimizing the ultrasound transducer and the imaging method for the reception of these components. The imaging modality that benefits from the third to the fifth harmonics is called SuperHarmonic Imaging (SHI). The design of a new medical phased array transducer that is optimized for harmonic imaging requires the accurate prediction of the three-dimensional, nonlinear acoustic pressure field generated by such a transducer, even when it is steered over a wide angle. This paper aims at the development of a numerical method that is capable of this. The presented method employs no *a priori* assumption on the directivity of the wavefield, and it shows an excellent accuracy for the nonlinear fields occurring in medical diagnostic applications.

Most currently used three-dimensional, nonlinear, numerical acoustic methods assume forward-wave propagation and are based on an evolution equation that, starting from a source plane, marches the solution from plane to plane in the main direction of propagation. Within this group of methods, a first line of research is based on the Kuznetsov–Zabolotskaya–Khokhlov (KZK) equation.[5] Because it employs the parabolic or paraxial approximation, the KZK equation is only valid for beams with quasi-planar wavefields. For this reason, the region of validity is generally located not too close to the source and not too far off the main transducer axis.[6] Currently, a score of numerical implementations is based on the KZK equation and predicts the nonlinear acoustic field of plane source geometries with an arbitrary aperture.[7–10] A second line of forward-wave methods uses a phenomenological approach and splits each marching step into separate operations that account for the effects of diffraction, absorption, and nonlinearity.[11–14] The diffraction step is incorporated without using a paraxial approximation. Zemp *et al.*[13] employed an efficient angular spectrum approach for this step. Varslot and Taraldsen[15] developed a forward-wave propagating method for heterogeneous media, and within this context they compared both a parabolic and an angular spectrum approach. For the nonlinear propagation step, all known forward-wave methods employ a plane-wave solution in the time domain or in the frequency domain, thus assuming that the main nonlinear distortion is in the direction of the normal to the computational planes. Although at each diffraction step the nonlinear distortion is spread again, this may not be accurate for wave-

[a]Present address: Noise and Vibration Research Group, Division PMA, Department of Mechanical Engineering, Catholic University Leuven, Leuven, Belgium.
[b]Electronic mail: m.d.verweij@tudelft.nl

fields that are strongly focused or that propagate at an angle significantly different from the direction normal to the computational planes. A correction method has been proposed by Christopher and Parker.[11] Further, Fox *et al.*[16] implemented a KZK model with a steered propagation axis, and Christopher[17] did the same for the phenomenological approach. However, when studying wide-angle phenomena such as grating lobes or propagation through heterogeneous media, nonlinear distortion in *all* directions needs to be accounted for. A method that handles nonlinear propagation in all forward directions was proposed by Varslot *et al.*,[18] but this method is limited by its quasilinear approximation.

Apart from the forward-wave methods, a number of full-wave methods have been developed to compute the nonlinear propagation of acoustic fields.[19–24] Because of the computational effort involved, these methods have been limited to two-dimensional Cartesian or cylindrical implementations. None of these methods has an assumed directionality in the wave propagation or in the nonlinear distortion. An issue with full-wave methods is the fine discretization in space and time that is needed for an accurate computation. In this paper, the discretization is expressed by the number of grid points D_ϕ per wavelength and per period of a particular field component with temporal (angular) frequency ϕ. The Nyquist–Shannon sampling theorem for bandlimited signals prescribes a minimum of $D_\Phi=2$ for the maximum relevant frequency of interest Φ. Finite difference and finite element methods need at least $D_\Phi=10$. Since the number of grid points needed for the computation of a three-dimensional, pulsed field is proportional to D_Φ^4, a high value of D_Φ makes these methods particularly unfavorable for application to large scale, three-dimensional problems. The best effort in this respect has been made by Wojcik *et al.*,[24] who presented a pseudospectral method that handles the spatial differentiation in the *k*-space domain. This method requires $D_\Phi=4$ in the spatial dimensions, and with a Courant number of 0.2 it employs at least $D_\Phi=20$ in the temporal dimension.

The numerical method presented in this paper addresses nonlinear, wide-angle, large scale wave propagation of transient acoustic waves in three dimensions. The main advantages of the method are a nonlinear distortion that is independent of the direction of propagation, and a discretization that is close to $D_\Phi=2$ in all dimensions. The former is resolved by an iterative solution of the nonlinear wave problem, and the latter is obtained by an adequate spatiotemporal filtering of the problem. The method will be referred to as the *Iterative Nonlinear Contrast Source* (INCS) method. The size of the computational domain will be in the order of $100\times100\times100$ wavelengths and 100 periods at the maximum frequency of interest. In view of the medical diagnostic application, the focus will be on weak to moderate nonlinearity, with special interest in the generation of the second and higher harmonic frequency components. Moreover, local nonlinear effects will be discarded.

This paper starts with the mathematical formulation of the nonlinear wave propagation problem that results in the lossless Westervelt equation.[25] The nonlinear term in this equation is interpreted as a contrast source,[26] and the nonlinear wave problem is formally solved with an iterative integral equation scheme. The most significant operation in the repeated steps of the iterative scheme is a convolution of Green's function with the primary and contrast sources. The efficient numerical evaluation of this convolution integral is performed with the Filtered Convolution (FC) method. This method has been described in more detail in a previous paper[26] that will here be referred to as the FC paper and that will be briefly summarized in this paper. Subsequently, results are presented and discussed for the nonlinear acoustic fields from several types of sources, including medical phased array transducers, and the obtained predictions are compared with results of a number of methods from literature. The paper ends with a discussion of the performance of the method as compared to other methods, and some conclusions.

II. PROBLEM FORMULATION

A. Nonlinear wave equation

Consider a nonlinear fluid that is homogeneous, isotropic, and lossless. The propagation of acoustic waves in this medium may be described by the set of first-order acoustic equations[27,28]

$$\nabla p + \rho D_t \boldsymbol{v} = \boldsymbol{f}, \tag{1}$$

$$\nabla \cdot \boldsymbol{v} + \kappa D_t p = q, \tag{2}$$

where p [Pa] is the acoustic pressure, $\boldsymbol{v}$ [m s^{-1}] is the particle velocity, and $\boldsymbol{f}$ [N m^{-3}] and q [s^{-1}] are the volume densities of the primary sources of force and injection rate, respectively. The medium behavior is accounted for by the density of mass ρ [kg m^{-3}] and the compressibility κ [Pa^{-1}]. Further, $D_t=\partial_t+\boldsymbol{v}\cdot\nabla$ denotes the total or material time derivative, which is typical for the Eulerian description of wave motion. The second term in D_t is an inherent cause of nonlinearity. Another cause for nonlinear behavior is the nonlinearity of the equation of state for the medium. Appendix A shows that the total nonlinear effect is accounted for until the second-order terms in Eqs. (1) and (2) by taking

$$\rho = \rho_0[1 + \kappa_0 p], \tag{3}$$

$$\kappa = \kappa_0[1 + \kappa_0(1 - 2\beta)p], \tag{4}$$

where ρ_0 is the ambient density of mass, κ_0 is the ambient compressibility, and $\beta=1+B/2A$ is the coefficient of nonlinearity. Appendix B shows that, upon neglection of the local nonlinear effects, Eqs. (1)–(4) lead to the lossless Westervelt equation

$$c_0^{-2}\partial_t^2 p - \nabla^2 p = S + \frac{\beta}{\rho_0 c_0^4}\partial_t^2 p^2, \tag{5}$$

where $c_0=(\rho_0\kappa_0)^{-1/2}$ is the small-signal sound speed and $S=\rho_0\partial_t q-\nabla\cdot\boldsymbol{f}$ represents the primary source.

This paper is concerned with plane source geometries located in the plane $z=0$, e.g., the phased array geometry depicted in Fig. 1. Corresponding to the two terms in S, the plane source may be represented as a jump condition in either $\boldsymbol{v}$ or p. A *velocity jump* of magnitude ΔV is realized by

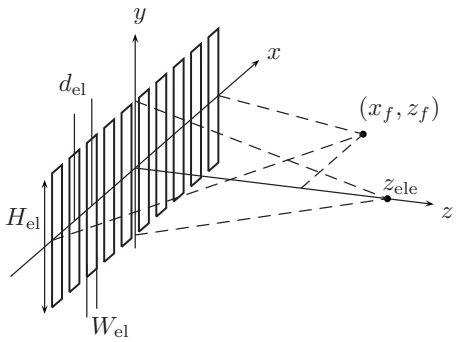


FIG. 1. Plane source geometry for a phased array transducer in the plane $z=0$, employing N_{el} elements of width W_{el}, height H_{el}, and pitch d_{el}, and lateral focus (x_f, z_f) and elevation focus z_{ele}, where z_{ele} is not necessarily equal to z_f.

imposing an injection rate acting in an infinitely thin source volume, yielding

$$S = \rho_0 \partial_t [\Delta V(x,y,t)\,\delta(z)], \tag{6}$$

where $\delta(z)$ is the Dirac delta distribution. A *pressure jump* of magnitude ΔP is realized by imposing a force acting in an infinitely thin source volume and working perpendicular to its surface, yielding

$$S = -\partial_z [\Delta P(x,y,t)\,\delta(z)]. \tag{7}$$

When only one side of the source plane is considered and the medium is homogeneous, these two source descriptions are equivalent to the respective boundary value problems of a transducer with a normal velocity $v_z = \Delta V/2$ in a perfectly rigid baffle, and a transducer with a pressure $p = \Delta P/2$ in a perfectly compliant baffle.

B. Iterative solution

The nonlinear acoustic pressure field $p(\boldsymbol{x},t)$ is interpreted as a summation of a linear field solution $p^{(0)}(\boldsymbol{x},t)$ and a nonlinear field correction $\delta p(\boldsymbol{x},t)$. The linear field solution $p^{(0)}$ may be obtained by solving the linearized wave equation

$$c_0^{-2}\partial_t^2 p^{(0)} - \boldsymbol{\nabla}^2 p^{(0)} = S. \tag{8}$$

The solution method for the linear wave equation is denoted by the symbolic operator equation $p^{(0)} = \mathcal{L}[S]$. To find $\delta p(\boldsymbol{x},t)$, the nonlinear term on the right hand side of Eq. (5) is considered as a *nonlinear contrast source*

$$S_{\mathrm{nl}}(p) = \frac{\beta}{\rho_0 c_0^4}\partial_t^2 p^2, \tag{9}$$

which acts in the homogeneous, lossless, and linear background medium associated with Eq. (8). Now, δp is related to $S_{\mathrm{nl}}(p)$ as $\delta p = \mathcal{L}[S_{\mathrm{nl}}(p)]$. If the total nonlinear pressure field p is dominated by $p^{(0)}$, then a good first estimate of δp is $\delta p^{(1)} = \mathcal{L}[S_{\mathrm{nl}}(p^{(0)})]$. This yields a first estimate $p^{(1)} = p^{(0)} + \delta p^{(1)}$ of p, which may be employed to improve our estimate of $S_{\mathrm{nl}}(p)$. Upon repeated application of these steps, the contrast source formulation enables the computation of a successive estimate $p^{(j)} = p^{(0)} + \delta p^{(j)}$ to p, resulting in the Neumann iterative solution[29]

$$p^{(0)} = \mathcal{L}[S^{(0)}],$$

$$S^{(0)} = S, \tag{10}$$

$$p^{(j)} = p^{(0)} + \mathcal{L}[S^{(j)}],$$

$$S^{(j)} = S_{\mathrm{nl}}(p^{(j-1)}), \quad j \geq 1. \tag{11}$$

This scheme is the formal equivalent of the approach used by Hoffelner *et al.*[23] Moreover, it can be interpreted as a perturbation scheme,[5] where $p^{(1)}$ is identified as the quasilinear solution. With increasing j, this quasilinear solution is iteratively improved toward the full nonlinear wavefield.

One method to compute $\mathcal{L}[S^{(j)}]$ is the evaluation of the convolution integral[30]

$$\mathcal{L}[S^{(j)}](\boldsymbol{x},t) = G(\boldsymbol{x},t) *_{\boldsymbol{x},t} S^{(j)}(\boldsymbol{x},t)$$

$$= \int_{\mathcal{D}_T^{(j)}} \int_{\mathcal{D}_X^{(j)}} G(\boldsymbol{x}-\boldsymbol{x}', t-t') S^{(j)}(\boldsymbol{x}',t')\,\mathrm{d}\boldsymbol{x}'\,\mathrm{d}t', \tag{12}$$

where $G(\boldsymbol{x},t)$ is the three-dimensional Green's function

$$G(\boldsymbol{x},t) = \frac{\delta(t - \|\boldsymbol{x}\|/c_0)}{4\pi\|\boldsymbol{x}\|} \tag{13}$$

of the homogeneous, lossless, and linear background medium.[27,31] In Eq. (12), $\mathcal{D}_T^{(j)}$ and $\mathcal{D}_X^{(j)}$ denote the relevant temporal and spatial domains of integration. For $\mathcal{L}[S^{(0)}]$, the spatial integration is over the support $\mathcal{D}_X^{(0)}$ of the plane source S in $z=0$ and involves integrations in x and y only. For the successive estimates $\mathcal{L}[S^{(j)}]$, the spatial integration is over the three-dimensional support $\mathcal{D}_X^{(j)}$ of the contrast source $S^{(j)}$. In principle, the latter support spans the entire space where an acoustic pressure is observed. In practice, however, the domain in which the contrast sources contribute significantly to $p^{(j)}$ is limited, as will be shown in Sec. V A.

The convergence of the scheme may be illustrated for a nonlinear plane-wave problem by comparing the iterations $p^{(j)}$ to an implicit solution p_B of the lossless Burgers' equation.[32] The iterations $p^{(j)}$ are obtained using a one-dimensional numerical implementation of Eqs. (10)–(12).[28] The medium is water ($\rho_0 = 998$ kg m^{-3}, $c_0 = 1492$ m s^{-1}, and $\beta = 3.52$) and the excitation pulse is a harmonic signal with a Gaussian envelope, yielding a pressure pulse with a center frequency of $f_0 = 1$ MHz, a 2σ pulse width of 2.12 μs, and a maximum surface pressure $P_0 = 500$ kPa at the source. Figure 2 shows the frequency spectrum of $p^{(j)}$ at $x = 100$ mm for $j=0$ to $j=6$. The convergence of the Neumann iterative solution is very fast. The relative error $|\hat{p}^{(j)}(f) - \hat{p}_B(f)|/|\hat{p}_B(f)|$ has decreased to 0.04% at the second harmonic ($2f_0$) for $j=3$, to 0.12% at the third harmonic ($3f_0$) for $j=4$, and to 0.16% at the fourth harmonic ($4f_0$) for $j=5$. As a rule of thumb, for weak to moderate nonlinear distortion, we may suffice with iteration $j = h+1$ for an accurate prediction of the h-th harmonic.

III. DISCRETIZATION AND NUMERICAL EVALUATION METHOD

Even in the simplest cases, the convolution integral in Eq. (12) must be discretized and evaluated numerically. As

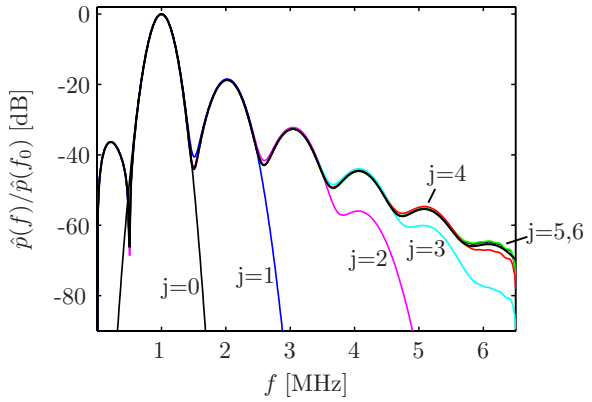


FIG. 2. (Color online) Frequency spectrum of the pressure p_B as obtained by the Burgers' equation (thick line) and of the pressure $p^{(j)}$ as obtained by the Neumann iterative solution for $j=0$ to $j=6$ (thin lines). The point of observation is $x=100$ mm. Each successive iteration $p^{(j)}$ gives a better estimate for increasingly higher harmonics.

has been discussed in Sec. I, for a full-wave, three-dimensional space-time domain method, it is vital to use a discretization that is as coarse as possible, preferably up to the limit of $D_\Phi=2$. An additional motivation for this is that with the described iterative solution the pressure field at the full spatiotemporal grid needs to be evaluated and stored in each iteration. Such a coarse discretization may be realized with the FC method.[26] In the this section the four steps of the FC method will be briefly summarized, and a specific issue in applying this method to the nonlinear wave problem will be discussed.

A. Efficient discretization employing the filtered convolution method

A straightforward approach to numerically evaluate the convolution integral in Eq. (12) is by approximating it with a convolution sum

$$\mathcal{L}[S^{(j)}]_{s,n} = (\Delta x)^3 \Delta t \sum_{\boldsymbol{r}} \sum_{m} G_{\boldsymbol{s}-\boldsymbol{r},n-m} S^{(j)}_{\boldsymbol{r},m}, \tag{14}$$

where $G_{\boldsymbol{r},m}$ and $S^{(j)}_{\boldsymbol{r},m}$ are the sampled versions of $G(\boldsymbol{x},t)$ and $S^{(j)}(\boldsymbol{x},t)$ on a grid with equidistant points $(\boldsymbol{r}\Delta x, m\Delta t)$, Δx and Δt being the spatial and temporal step sizes. The resulting values $\mathcal{L}[S^{(j)}]_{s,n}$ are approximated values of $\mathcal{L}[S^{(j)}]$ on the grid points $(\boldsymbol{s}\Delta x, n\Delta t)$, and the error strongly depends on the chosen step sizes. In order to attain a coarse discretization, it is assumed that a maximum angular frequency of interest Φ may be chosen in such a way that all components of interest in the pressure field have a temporal angular frequency $|\omega| \le \Phi$ and a spatial angular frequency $|\boldsymbol{k}| = (k_x^2+k_y^2+k_z^2)^{1/2} \le \Phi/c_0$. The first steps of the FC method are to filter and window $G(\boldsymbol{x},t)$ and $S^{(j)}(\boldsymbol{x},t)$ in all spatiotemporal dimensions before these functions are sampled. This means that $G(\boldsymbol{x},t)$ and $S^{(j)}(\boldsymbol{x},t)$ are subjected to an ideal low-pass filter with an angular cutoff frequency $\Omega=\pi/\Delta t$ in the temporal domain and ideal, spherical low-pass filter with an angular cutoff frequency $K=\pi/\Delta x$ in the spatial domain. This avoids the aliasing error that would otherwise be caused during the sampling operation by the frequency components $|\omega| > \Omega$ and $\{|k_x|, |k_y|, |k_z|\} > K$. Because of the assumption on the frequency content of interest, Ω may be taken as small as Φ and K as small as Φ/c_0. The resulting step sizes may therefore be as large as $\Delta t = \pi/\Phi$ and $\Delta x = c_0\pi/\Phi$, resulting in a discretization of $D_\Phi=2$. To limit the number of points in each dimension and thus enable the evaluation of a limited convolution sum, a rectangular window is applied to $G(\boldsymbol{x},t)$ and $S^{(j)}(\boldsymbol{x},t)$ in all spatiotemporal dimensions. After the filtering and windowing operations, the second step of the FC method is to sample $G(\boldsymbol{x},t)$ and $S^{(j)}(\boldsymbol{x},t)$ at a coarse discretization of $D_\Phi=2$. The last two steps of the FC method involve the evaluation of the resulting convolution sum in Eq. (14) with a Fast Fourier Transform (FFT) method. To achieve this, in step 3, $G_{\boldsymbol{r},m}$ and $S^{(j)}_{\boldsymbol{r},m}$ are transformed with four-dimensional FFTs, and in step 4 the obtained results are multiplied and transformed back to the original domain with an inverse FFT to yield $\mathcal{L}[S^{(j)}]_{s,n}$.

B. Spatiotemporal filtering of the nonlinear contrast source

In the FC paper,[26] the spatiotemporally filtered and windowed version of the three-dimensional Green's function in Eq. (13) is obtained analytically. In the same paper, the spatiotemporal filtering and windowing of the primary source, as well as the filtering and windowing of a general contrast source, are thoroughly discussed. In this subsection, several aspects are exemplified that are specific to the spatiotemporal filtering of the nonlinear contrast source.

For the evaluation of the contrast source $S^{(j)}$ the pressure field $p^{(j-1)}$ is used. It is observed that $p^{(j-1)}$ is bandlimited in all spatiotemporal dimensions if it has been obtained with the FC method. At first sight it seems that the squaring of $p^{(j-1)}$ generates a result with a spatiotemporal bandwidth that may exceed both the angular cutoff frequency Ω and cutoff wavenumber $K=\Omega/c_0$, and filtering must be applied in all dimensions to prevent aliasing. However, since we may assume that $p^{(j-1)}$ predominantly consists of propagating waves for which the spatial and temporal behaviors are related by $|k|=\omega/c_0$, the numerical evaluation may be considerably simplified by omitting the filtering in the spatial dimensions and applying the filter only in the temporal dimension. This is shown by first studying the behavior of the nonlinear contrast source $S_{\mathrm{nl}}(p)$ for a pressure field $p(\boldsymbol{x},t)$ that consists of two steady-state, two-dimensional, plane waves traveling in the plane $y=0$, having frequencies ω_1 and ω_2 with $0 \le \omega_2 \le \omega_1 \le \Omega$, and crossing each other at an angle $0 \le \alpha \le \pi$,

$$\begin{aligned} p(\boldsymbol{x},t) &= I_{\omega_1} + I_{\omega_2} \\ &= \sin\left[\omega_1 t - \frac{\omega_1}{c_0} s_\alpha x - \frac{\omega_1}{c_0} c_\alpha z\right] \\ &\quad + \sin\left[\omega_2 t + \frac{\omega_2}{c_0} s_\alpha x - \frac{\omega_2}{c_0} c_\alpha z\right], \end{aligned} \tag{15}$$

where $s_\alpha = \sin(\alpha/2)$ and $c_\alpha = \cos(\alpha/2)$. The square of $p(\boldsymbol{x},t)$ that occurs in $S_{\mathrm{nl}}(p)$ consists of five terms

$$p^2(\boldsymbol{x},t) = I_0 + I_{2\omega_1} + I_{2\omega_2} + I_{\omega_1+\omega_2} + I_{\omega_1-\omega_2}, \tag{16}$$

where

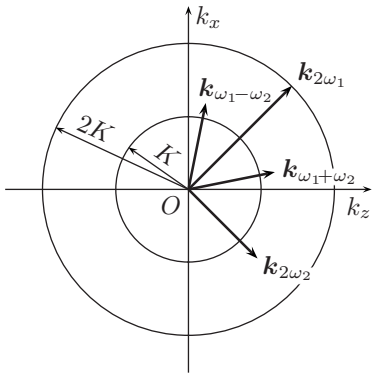


FIG. 3. Location of the wavevectors $\boldsymbol{k}$ associated with the terms $I_{2\omega_1}$, $I_{2\omega_2}$, $I_{\omega_1+\omega_2}$, and $I_{\omega_1-\omega_2}$ for the cases $\omega_1=\Omega$, $\omega_2=(2/3)\Omega$, and $\alpha=90°$.

$$I_0 = 1, \tag{17}$$

$$I_{2\omega_1} = -\frac{1}{2}\cos\left[2\omega_1 t - \frac{2\omega_1}{c_0}s_\alpha x - \frac{2\omega_1}{c_0}c_\alpha z\right], \tag{18}$$

$$I_{2\omega_2} = -\frac{1}{2}\cos\left[2\omega_2 t + \frac{2\omega_2}{c_0}s_\alpha x - \frac{2\omega_2}{c_0}c_\alpha z\right], \tag{19}$$

$$I_{\omega_1+\omega_2} = -\cos\left[(\omega_1+\omega_2)t - \frac{\omega_1-\omega_2}{c_0}s_\alpha x - \frac{\omega_1+\omega_2}{c_0}c_\alpha z\right], \tag{20}$$

$$I_{\omega_1-\omega_2} = \cos\left[(\omega_1-\omega_2)t - \frac{\omega_1+\omega_2}{c_0}s_\alpha x - \frac{\omega_1-\omega_2}{c_0}c_\alpha z\right]. \tag{21}$$

Due to the second-order time derivative in Eq. (9), the term I_0 does not yield a contribution to $S_{\mathrm{nl}}(p)$. The remaining terms are of the form $I_\omega = a\cos(\omega t - k_x x - k_z z)$, in which k_x and k_z may be considered as the components of a wavevector that depends on the particular values of ω_1, ω_2, and α. Figure 3 shows the wavevectors for a typical situation. The temporal filtering removes each term I_ω with $\omega > \Omega$. Inspection of Eqs. (18)–(20) shows that the terms $I_{2\omega_1}$, $I_{2\omega_2}$, and $I_{\omega_1+\omega_2}$ that survive the temporal filtering also have $k_x \leq K$ and $k_z \leq K$. This implies that the terms associated with the wavevectors $\boldsymbol{k}_{2\omega_1}$, $\boldsymbol{k}_{2\omega_2}$, and $\boldsymbol{k}_{\omega_1+\omega_2}$ in Fig. 3 are eliminated by the temporal filtering. On the other hand, the temporal filtering will never remove a term $I_{\omega_1-\omega_2}$ because it always occurs that $(\omega_1-\omega_2) \leq \Omega$. Inspection of Eq. (21) reveals that although for this term $k_z \leq K$, there are circumstances under which $k_x = (\omega_1+\omega_2)s_\alpha/c_0 > K$, and these are not dealt with by the temporal filtering. This implies that the term with the wavevector $\boldsymbol{k}_{\omega_1-\omega_2}$ in Fig. 3 remains present if no spatial filtering is applied. The issue is now to determine for which values of $\omega_1 = \gamma\Omega$ $(0 \leq \gamma \leq 1)$, $\omega_2 = \eta\omega_1$ $(0 \leq \eta \leq 1)$, and α the term $I_{\omega_1-\omega_2}$ will still possess a $k_x \leq K$ and, consequently, the spatial filtering of the contrast source may still be skipped. It turns out that this is the case if

$$(1+\eta)\gamma \sin\left(\frac{\alpha}{2}\right) \leq 1. \tag{22}$$

This inequality always holds if $\gamma \leq 1/2$, i.e., for frequencies $\omega_2 \leq \omega_1 \leq \Omega/2$, and independently if $\alpha \leq \pi/3$, i.e., for a crossing angle smaller than 60°. For other values of γ and α, only specific ranges of η yield a spatial frequency larger than K. In conclusion, for the case of two crossing plane waves, the temporal filtering with a cutoff frequency Ω implies spatial filtering with a cutoff frequency less than or equal to $K=\Omega/c_0$ for all frequencies smaller than $\Omega/2$ and for all crossing angles smaller than 60°.

The considered case of two crossing plane waves can be generalized to a realistic three-dimensional situation where the nonlinear pressure field p consists of a continuum of plane waves, all with temporal frequencies below Ω. In this situation, the pressure field in points that are close to the source shows a different characteristic than the pressure field in points that are some distance away from the source. For points close to the source, the cumulative nature of nonlinear propagation results in a pressure field in which the higher harmonic components are virtually absent, and the field only consists of the fundamental frequency components in a small frequency band around f_0. In any computation involving nonlinear fields, the maximum frequency of interest $F = \Omega/2\pi$ is chosen such that at least the second harmonic is incorporated, i.e., $F \geq 2f_0$. As a consequence, $\gamma \leq 1/2$ for the temporal frequencies involved with the field near the source, and the spatial filtering will be superfluous here. For points at some distance from the source, the constituents of p will travel more or less parallel to each other and cross each other under an angle far less than 60°. This causes that at some distance from the source $\alpha < \pi/3$ and the spatial filtering may safely be omitted as well.

IV. IMPLEMENTATION

The method has been implemented in FORTRAN on a single processor system as well as on a multiprocessor parallel system with distributed memory. The FFTs are performed with routines from the FFTW library.[33] The evaluation of the partial derivatives in the primary and contrast sources utilizes either a high order finite difference scheme or a spectral derivative.[34] The structure of the parallel program may be found in the FC paper.[26] Although with the FC method the discretization occurs at the theoretical limit of two points per wavelength, a number of additional computational strategies may be invoked to realize a further reduction in memory and computation time. First, symmetrical source geometries allow the reduction in the spatial computational domain by only computing the fields for $x \geq 0$ and/or $y \geq 0$, and deriving from them the contrast sources at $x < 0$ and/or $y < 0$. Second, reduction in the temporal dimension may be obtained by using a periodical time axis, which is achieved by omitting the temporal windowing operation on Green's function. Third, the convolution sum may be split up and evaluated in blocks, thus reducing the demand in core memory. Finally, instead of only using fixed rectangular domains, a co-moving time frame along the z-dimension may be employed, or a domain in the shape of a parallellogram in either the

(x,z)-plane or the (y,z)-plane may be used in case of steered beams or grating lobes under a large angle with the z-axis. This results in a smaller time window and/or spatial domain, especially for large observation angles. The skew domain may be wrapped in such a way that a rectangular domain results, which may then be efficiently dealt with by an FFT method.[35]

V. NUMERICAL RESULTS

In this section results are presented that are obtained by the INCS method for the nonlinear field of a point source, an unfocused cylindrical piston transducer with radius R, a rectangular transducer with width W and height H, and a focused and steered phased array transducer as shown in Fig. 1. All plane source representations will employ a pressure jump condition. The source geometry is included by a geometry factor $0 \leq A(x,y) \leq 1$. The pulse that will be used is a harmonic signal with a Gaussian envelope, resulting in a pressure jump

$$\Delta P(x,y,t) = 2P_0 A(x,y)\exp[-(2t/t_w)^2]\sin(2\pi f_0 t), \quad (23)$$

where P_0 is the maximum surface pressure of the source, f_0 is the center frequency, and t_w is the pulse width. The position-dependent delay of the time signature needed for the focusing of the phased array is accounted for in the temporal Fourier domain by using a geometry factor

$$\hat{A}(x,y,\omega) = A(x,y)\exp[-j\omega t_d(x,y)], \quad (24)$$

where $t_d(x,y)$ is the delay time. The medium is water ($\rho_0 = 998$ kg m^{-3}, $c_0 = 1492$ m s^{-1}, and $\beta = 3.52$), and attenuation is considered negligible.

A. Domain of the contrast source

The first results concern the spatial extent of the contrast sources that contribute significantly to the nonlinear field solution. As the nonlinear distortion is formed cumulatively along the propagation path up to a certain point of observation, it may be assumed beforehand that those nonlinear contrast sources that lie in the region in between the primary source and the observation point will mainly contribute to the nonlinear distortion, and that the influence of other contrast sources can be safely ignored. This assumption may be investigated by using the INCS method to compute the nonlinear field correction due to the entire contrast source, and to compare this with the nonlinear field correction obtained with a contrast source that is partially masked. The primary source is a rectangular piston with dimensions $W \times H = 10$ mm $\times 10$ mm, causing a pressure jump. The source emits a harmonic signal with a frequency of $f_0 = 1$ MHz and a Gaussian envelope with a pulse width of 6 cycles, i.e., $t_w = 3/f_0$, and a maximum surface pressure of $P_0 = 500$ kPa. The pressure field is obtained in a domain of interest with a spatial size of 15 mm $\times 15$ mm $\times 52$ mm and a co-moving temporal window of 22 μs. The maximum frequency of interest is $F = 3$ MHz, the discretization is performed at $D_F = 2$, and the linear field solution $p^{(0)}$ and the first nonlinear field correction $\delta p^{(1)}$ are obtained. The profile of $p^{(0)}$ is shown in Fig. 4(a), and the profile of $\delta p^{(1)}$, employing the entire region of interest as the contrast source domain, is depicted in Figs. 4(b) and 4(c). If the contrast source domain is extended to a size of 30 mm $\times 30$ mm $\times 104$ mm, then $\delta p^{(1)}$ does not change noticeably, which indicates that the original domain was indeed sufficiently large to compute the profiles in Figs. 4(b) and 4(c).

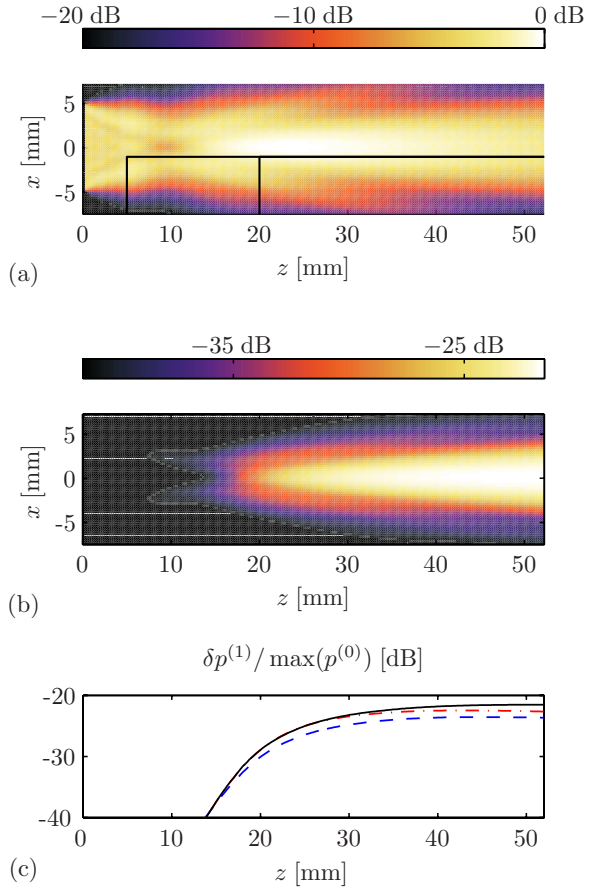


FIG. 4. (Color online) Profiles for the rectangular piston source. (a) The linear field solution $p^{(0)}$ in the plane $y=0$ mm, together with the masking domains. (b) The nonlinear field correction $\delta p^{(1)}$ in the plane $y=0$ mm, for the unmasked case. (c) $\delta p^{(1)}$ on the z-axis, for the unmasked case (solid), for a masking domain starting at $z=5$ mm (dashed) and for a masking domain starting at $z=20$ mm (dashed-dotted). The 0 dB level corresponds to 870 kPa.

Next, the contrast source is masked, i.e., put to zero, in a region $X \times Z = [-7.5 \text{ mm}, -1 \text{ mm}] \times [z, 52 \text{ mm}]$ with $z = \{5, 20\}$ mm, as depicted in Fig. 4(a). In this case an error $p_z^{(1)} - p^{(1)} = \delta p_z^{(1)} - \delta p^{(1)}$ in the first nonlinear field estimate shows up, where the subscript z denotes the specific masking domain. From Fig. 4(c) it becomes clear that the effect of this masking in the axial profile mainly manifests itself as an underestimation of the nonlinear field correction. This effect is larger for a masking domain starting closer to the primary source. The profiles of the errors $\delta p_z^{(1)} - \delta p^{(1)}$ are shown in Fig. 5. From these figures it may be observed that the less the masking blocks the view from an observation point to the primary source, the lower the error in the nonlinear field correction becomes. This confirms that, at least for a plane primary source, the contrast sources that contribute significantly to the nonlinear field in a certain observation point indeed lie in the region in between the observation point and

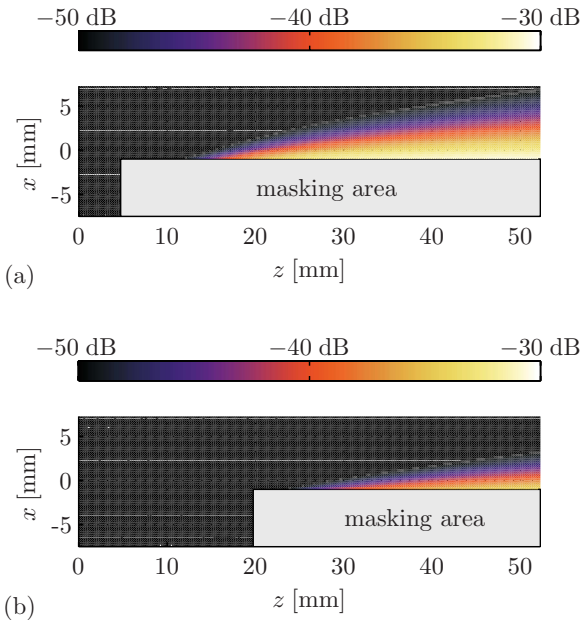


FIG. 5. (Color online) Profiles of the error $\delta p_z^{(1)} - \delta p^{(1)}$ in the nonlinear field corrections caused by the masking of the contrast source domains. The plots show the profiles in the plane $y=0$ mm for the masking domains starting at (a) $z=5$ mm and (b) $z=20$ mm. The 0 dB level corresponds to 870 kPa.

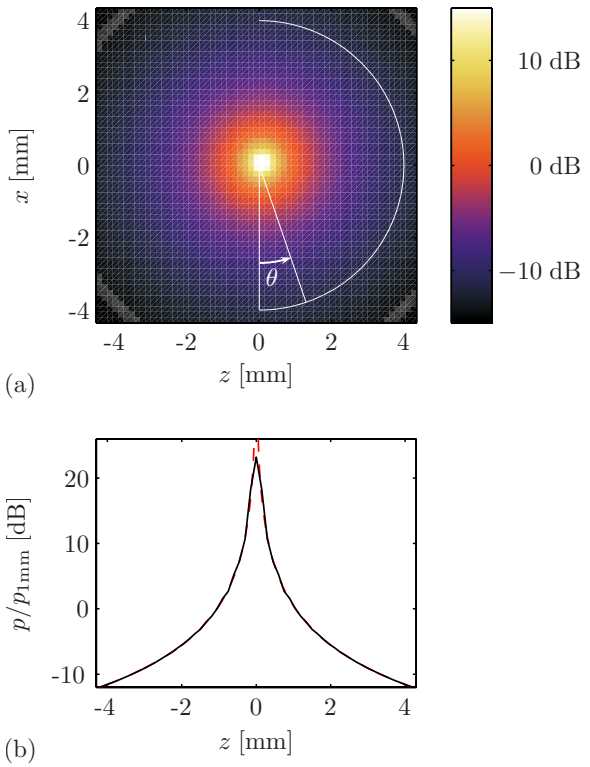


FIG. 6. (Color online) Profile of $p^{(0)}$ for the point source. (a) In the plane $y=0$, together with the half-circle used in Fig. 7. (b) On the z-axis, as obtained from the INCS method (solid) and from the analytical solution in Eq. (26) (dashed). The 0 dB level corresponds to 500 kPa.

the primary source. This knowledge can be used to reduce the support of the contrast source and thus to reduce the numerical effort needed for the iterations of the INCS method.

B. Non-directionality of the nonlinear distortion

An important benefit of the INCS method is the directional independence of the linear and nonlinear field solutions. This may be demonstrated by computing the nonlinear pressure field from a omnidirectional point source located at $\boldsymbol{x}=\boldsymbol{0}$, defined as

$$S = S_0 S(t)\,\delta(\boldsymbol{x}). \tag{25}$$

In the linear case, the field solution for the point source is found by substituting Eq. (25) and the three-dimensional Green's function of Eq. (13) in Eq. (12). This yields

$$p_L(\boldsymbol{x},t) = \frac{S_0 S(t-\|\boldsymbol{x}\|/c_0)}{4\pi\|\boldsymbol{x}\|}, \quad \boldsymbol{x}\neq 0. \tag{26}$$

The signature $S(t)$ is a harmonic signal with a unit amplitude and a Gaussian envelope with $f_0=1$ MHz and $t_w=3/f_0$. The amplitude factor S_0 equals $2\pi\times 10^3$ N m^{-1}, resulting in a maximum linear pressure of $p_L=500$ kPa$=p_{1\text{mm}}$ at $\|\boldsymbol{x}\|=1$ mm. In the nonlinear case, the domain of interest is centered at the origin and has a size of 8.6 mm$\times$8.6 mm $\times$8.6 mm, and a co-moving temporal window of 12 μs is used. The maximum frequency of interest is $F=5$ MHz, the discretization is performed at $D_F=2$, and the iterative scheme is run up to $j=3$. Figure 6(a) shows the profile for the linear field $p^{(0)}$ in the plane $y=0$, and Fig. 6(b) shows the profile on the z-axis, together with the profile of the linear field solution p_L from Eq. (26). As observed from the latter figure, the INCS method reproduces the linear field profile very accurately, except for a very small region around the point source.

The directional independence of the INCS method is demonstrated by the profile of the linear field solution $p^{(0)}$ and the profile of the nonlinear field correction $\delta p^{(3)}=p^{(3)}-p^{(0)}$ on the half-circle with radius $\|\boldsymbol{x}\|=4$ mm that is indicated in Fig. 6(a). Figure 7 shows the relative variation of the mentioned radial profiles around their mean levels, versus

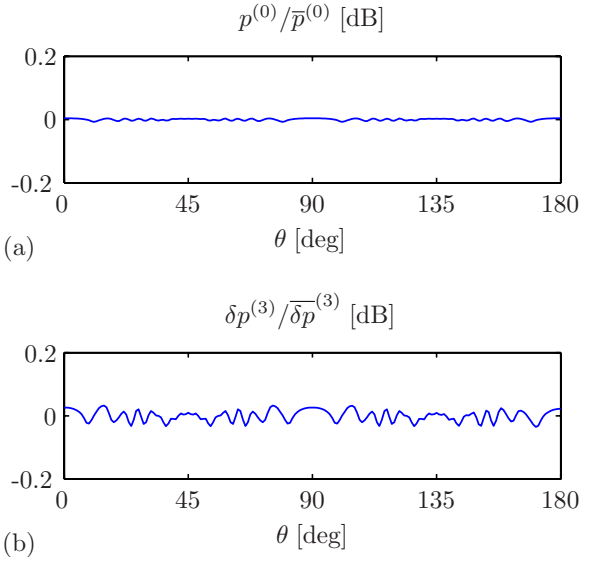


FIG. 7. (Color online) Radial profile levels at a semi-circle with radius $\|\boldsymbol{x}\|=4$ mm as indicated in Fig. 6(a), versus angle θ. (a) For the linear field solution $p^{(0)}$. (b) For the nonlinear field correction $\delta p^{(3)}$. The mean levels of the linear field profile and the nonlinear field correction profile at that radius are $\bar{p}^{(0)}=123$ kPa and $\overline{\delta p}^{(3)}=0.97$ kPa.

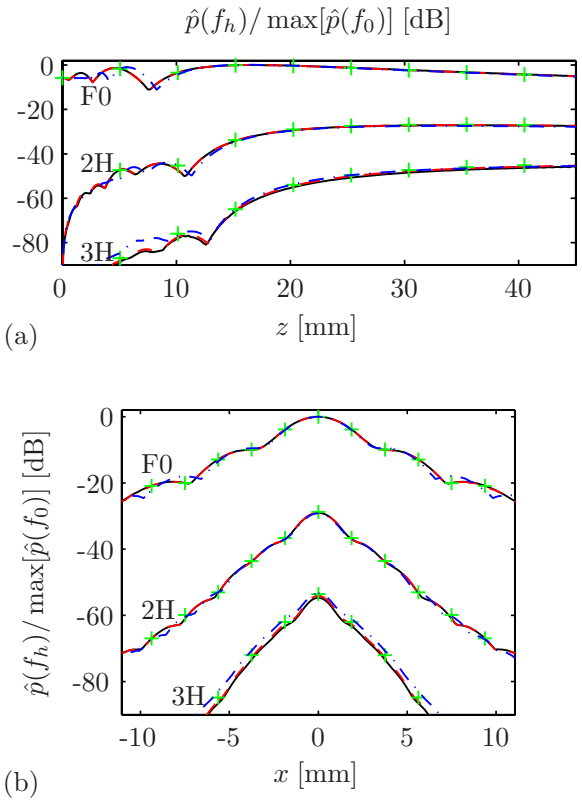


FIG. 8. (Color online) Spectral profiles for the cylindrical source as obtained by the INCS method (solid), the FD-WV method (dashed), the AS-NLP method (+), and the KZK method (dashed-dotted). (a) Axial spectral profiles. (b) Lateral spectral profiles at $z=20$ mm. The labels $F0$, $2H$, and $3H$, respectively, indicate the fundamental, second harmonic, and third harmonic. The 0 dB level corresponds to 983 kPa.

the angle θ. It may be observed that the relative directional variation is less than 0.05 dB. This fact confirms that the INCS method is virtually directionally independent.

C. Comparison with other methods for the field of a cylindrical piston transducer

To verify the accuracy of the INCS method, the results for an unfocused cylindrical piston transducer are compared with those of the following three existing methods: The Finite Difference Westervelt (FD-WV) method,[22,36] which is an FD approximation of the Westervelt equation for axisymmetric configurations, the Angular Spectrum NLP (AS-NLP) method,[13] which is a marching step scheme based on the angular spectrum approach, and a method based on the KZK equation.[37] The employed transducer has a radius $R=5$ mm, and is excited with a source pulse as used before with $f_0=1$ MHz, $t_w=3/f_0$, and $P_0=500$ kPa. The three existing methods are run with parameter settings that should result in an accurate field prediction. The INCS method employs a domain of interest with a spatial size of 17 mm $\times$ 11 mm $\times$ 45 mm and a co-moving temporal window of 39 μs, a maximum frequency of interest of $F=4$ MHz with a discretization at $D_F=2$, and iteration $j=4$ is investigated. Figure 8 shows the spectral profiles of the fundamental and the second and third harmonics on the axis of the transducer, and in a radial direction at $z=20$ mm. The profiles of the INCS method show a good agreement with those of the other methods. The agreement is best for the FD-WV method, with a difference of at most 0.3 dB for the fundamental and second harmonic and 0.9 dB for the third harmonic. With the AS-NLP method, the difference is at most 1.0 dB for the fundamental and second harmonic, and 2.0 dB for the third harmonic. The difference with the KZK code within the paraxial region, which is the region not too close to the source and up to $\pm 16°$ off the transducer axis, is at most 1.0 dB for the fundamental and second harmonic, and 3.0 dB for the third harmonic. For locations near the transducer axis, these comparisons confirm that the INCS method provides an accurate prediction of the fundamental and the higher harmonics.

D. Comparison with the AS-NLP method for the field of a phased array transducer

In a second accuracy test, the nonlinear field profiles are compared as predicted by the INCS method and by the AS-NLP method for a 48 element phased array transducer with element parameters $W_{\mathrm{el}} \times H_{\mathrm{el}}=0.21$ mm $\times 12$ mm, and $d_{\mathrm{el}}=0.5$ mm. The array is excited with a source pulse as used before with $f_0=1$ MHz, $t_w=3/f_0$ and $P_0=250$ kPa. The nonlinear field solution as obtained with the INCS method is evaluated in a domain of interest with a spatial size of 30 mm $\times 18$ mm $\times 72$ mm and a co-moving temporal window of 36 μs. The maximum frequency of interest is $F=4$ MHz, the discretization is performed at $D_F=2$, and the iteration $j=4$ is investigated.

First we consider an unsteered beam. In this case the array is focused at (0 mm, 57 mm), with $z_{\mathrm{ele}}=z_f$. Figure 9 shows the spectral profiles of the fundamental and the second and third harmonics, as obtained with the INCS method. In Fig. 10 the axial spectral profiles and the lateral spectral profiles for $y=0$ mm and $z=58$ mm are shown, as obtained with the INCS method and the AS-NLP method. These figures show an even better resemblance between the spectral profiles of the INCS method and the AS-NLP method than for the unfocused cylindrical source.

From the directional independence observed in Sec. V B, it is assumed that the INCS method will equally well predict the spectral profiles of the unsteered beam as considered above, and of a beam steered 45° off the transducer axis as considered next. The focus in this case is $(x_f, z_f)=$ (40 mm, 40 mm), with $z_{\mathrm{ele}}=z_f$, giving the same focal distance as in the unsteered case. The domain of interest is parallellogram-shaped in the xz-plane, conforming the beam angle, and this skew domain has a size of 30 mm $\times 18$ mm $\times 53$ mm. In Fig. 11, the fundamental, the second harmonic, and the third harmonic spectral profiles in the plane $y=0$ mm are shown, as obtained with the INCS method with the same discretization and number of iterations as in the unsteered situation. Figure 12 shows the lateral spectral profiles for $y=0$ mm and $z=40$ mm, as obtained by the INCS method and the AS-NLP method. For the latter, similar parameter settings as for the unsteered case are utilized. The direction perpendicular to the computational planes is kept in the direction of the z-axis, and in the current case it is therefore not directed along the steered beam axis. For the fundamental component, a perfect agreement be-

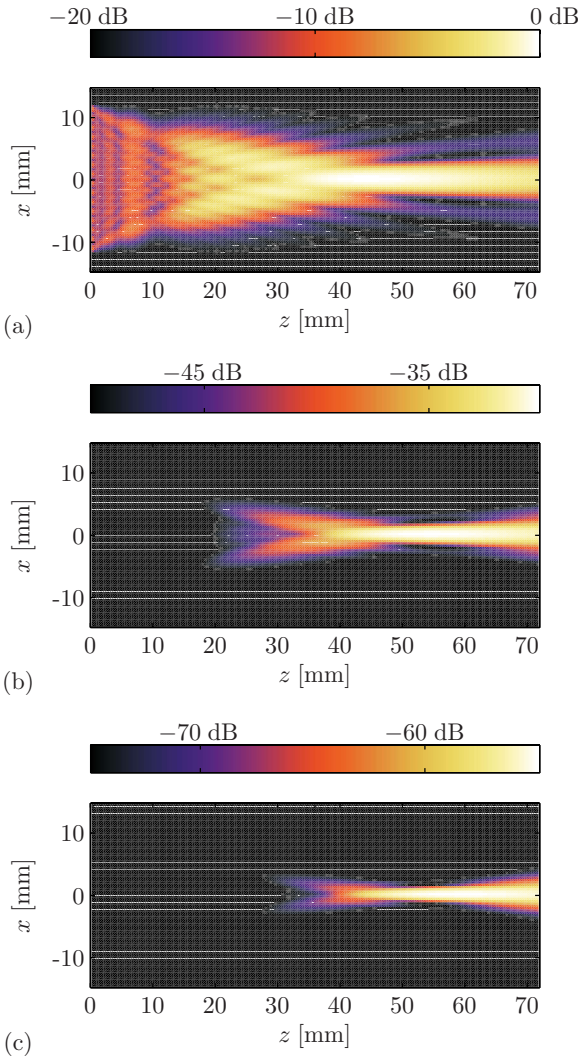


FIG. 9. (Color online) Spectral profiles in the plane $y=0$ mm for the phased array transducer exciting an unsteered beam. (a) Fundamental spectral profile. (b) Second harmonic spectral profile. (c) Third harmonic spectral profile. The 0 dB level corresponds to 414 kPa.

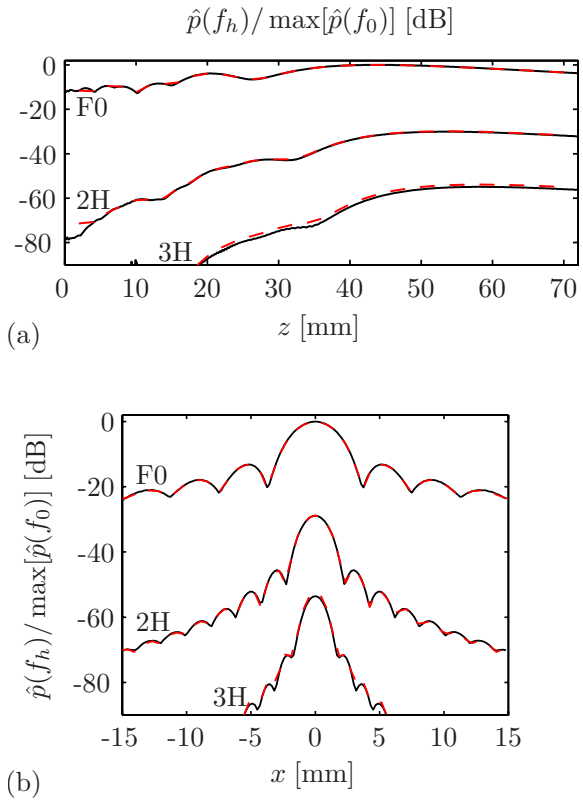


FIG. 10. (Color online) Spectral profiles for the phased array transducer exciting an unsteered beam, as obtained by the INCS method (solid) and by the AS-NLP method (dashed). (a) Axial spectral profiles. (b) Lateral spectral profiles at $y=0$ mm and $z=58$ mm. The 0 dB level corresponds to 414 kPa.

tween both methods is observed. At the main beam, i.e., at $x=40$ mm, the level of the second harmonic component predicted by the AS-NLP method is 2.6 dB lower than the level predicted by the INCS method, and this difference increases for larger x. The third harmonic components obtained by both methods differ by 7.5 dB at the main beam, and the tails have a widely different shape and level. The case of steering off axis reveals two points. First, that the INCS method is capable of dealing with this type of fields, and second, that the quasi-planar wave assumption in the nonlinear step of the AS-NLP method yields a result of limited accuracy for fields that are not propagating in the assumed direction, i.e., perpendicular to the computational planes.

VI. DISCUSSION

Since the INCS method is designed to be a directionally independent method, its discretization in all dimensions is coupled to the highest frequency of interest F. The presented comparisons confirm that the INCS method reproduces the nonlinear acoustic wavefield very accurately up to the limit $D_F=2$. To obtain accurate results for the hth harmonic frequency component at a discretization $D_F=2$, it is found sufficient to take $F=(h+1)f_0$, and detailed investigation by the authors showed that even $F=(h+0.5)f_0$ showed good results. With regard to the number of iterations, it is found that for weak to moderate nonlinear distortion, the iteration $j=h+1$ already provides sufficient accuracy.

The diffraction and the nonlinear distortion are included without any dependence on the direction, which makes the INCS method fit for an arbitrarily directed, focused, or wide-angle nonlinear acoustic wavefield, including sidelobes and grating lobes. The directional independence of the nonlinear distortion is particularly important for the accurate prediction of the higher harmonics of a steered beam. This is clearly illustrated by the comparison between the INCS method and the AS-NLP method in Fig. 12. An extra advantage of the method is that the utilized convolution method gives no reflections on the computational grid boundaries, as occurs for finite difference or finite element methods, and it has no periodical boundary conditions, which are an issue with pseudospectral methods.

Because the FFT method is utilized in the evaluation of the convolution sum, the amount of arithmetical operations required by the scheme is of order $N \log N$, where N is the total number of grid points. The method can therefore be considered as near-linear in its computation time. Because of the coarse discretization at the limit $D_F=2$ in all dimensions, the INCS method is very efficient in terms of computation time and memory usage. Nevertheless, compared to forward-

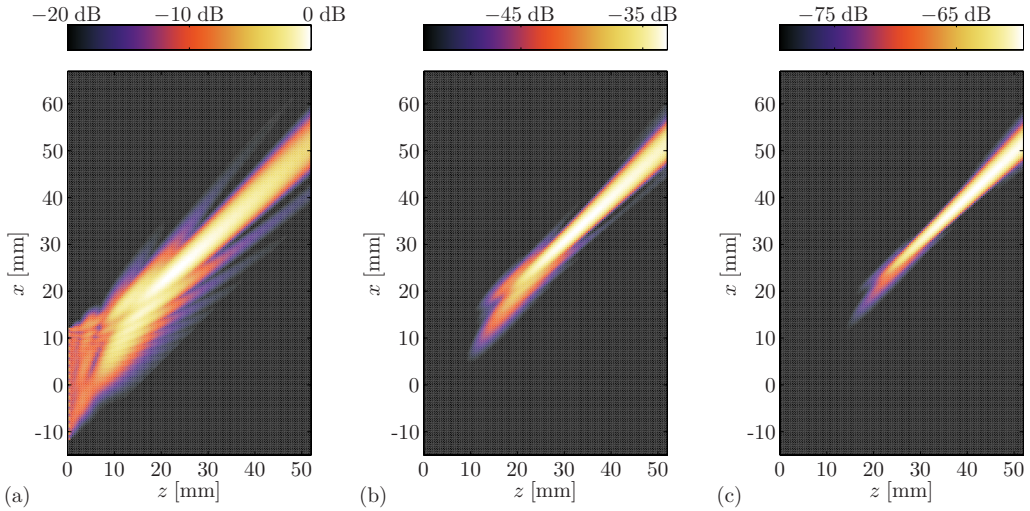


FIG. 11. (Color online) Spectral profiles in the plane $y=0$ mm for the phased array transducer exciting a beam steered at 45°. (a) Fundamental spectral profile. (b) Second harmonic spectral profile. (c) Third harmonic spectral profile. The 0 dB level corresponds to 358 kPa.

wave methods such as the AS-NLP method or KZK-based methods, and compared to full-wave schemes with a marching-on-in-time scheme, an apparent disadvantage of the employed iterative scheme is the necessity to store the field data at the entire four-dimensional grid. This requires a computer with considerable memory, or a significant usage of the harddisk as backup storage, or alternatively the use of a distributed computer system. The simulations of the fields of the point source, the rectangular source, and the cylindrical source were performed on a desktop computer running at 1 GHz and with 2 GB memory, and the execution times varied from 9 min to as much as 130 h. However, the large execution time was caused mainly by disk access time. The computations of the nonlinear field from the phased array were performed on 16 processors of an SGI Altix 3700 clustered multiprocessor system consisting of 416 Intel Itanium processors running at 1.3 GHz and with 2 GB memory per core. These simulations took between 1 and 3 h.

VII. CONCLUSIONS

In this paper, a novel, directionally independent method has been presented that predicts the three-dimensional, nonlinear acoustic pressure field from extended sources that excite a pulsed wavefield propagating under an arbitrarily wide angle. The method is based on the Neumann iterative solution of the lossless Westervelt equation, where the nonlinear term is interpreted as a nonlinear contrast source. In the iterative scheme, the most demanding step in each iteration is the spatiotemporal convolution of Green's function with the primary source or with the nonlinear contrast source. With the employment of the filtered convolution method,[26] this convolution may be performed accurately at a discretization of $D_F=2$ grid points per wavelength and per period at the maximum frequency of interest F. This enables the computation of acoustic wavefields over very large domains. The presented numerical results support the conclusion that the INCS method is directionally independent and produces very accurate results for the nonlinear acoustic wavefield in the case of weak to moderate nonlinearity, as occurs in medical diagnostic applications.

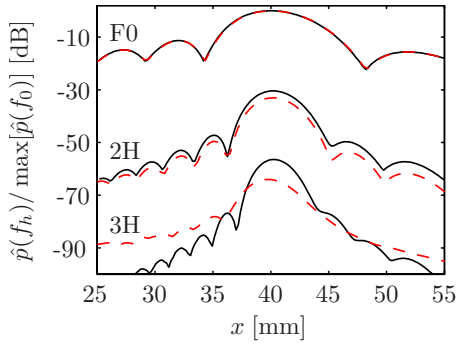


FIG. 12. (Color online) Lateral spectral profiles at $y=0$ mm and $z=40$ mm for the phased array transducer exciting a beam steered at 45°, as obtained by the INCS method (solid) and by the AS-NLP method (dashed). The 0 dB level corresponds to 358 kPa.

ACKNOWLEDGMENTS

This work was supported by the Dutch Technology Foundation (STW) and by the Dutch National Computing Facilities Foundation (NCF). The authors would like to thank Ayache Bouakaz, Richard Cobbold, and Roger Zemp for providing the source code of the comparison methods.

APPENDIX A: DERIVATION OF THE NONLINEAR CONSTITUTIVE PARAMETERS FROM THE EQUATION OF STATE

Often, the nonlinear behavior of the medium is derived from a second-order approximation of the equation of state relating the hydrostatic pressure P to the mass density ρ and the entropy s. Under the condition of isentropy, P may then be written as[5]

$$P = p_0 + c_0^2(\rho - \rho_0) + \frac{c_0^2}{\rho_0}\frac{B}{2A}(\rho - \rho_0)^2, \tag{A1}$$

where p_0 is the ambient pressure, ρ_0 is the ambient mass density, c_0 is the small-signal sound speed, and B/A is the

parameter of nonlinearity. Since in the nonviscous case P is related to the acoustic pressure by $p=P-p_0$, we can use the equation of state to write ρ as a function of p, giving

$$\rho = \rho_0 + \rho_0 \frac{A}{B}\left[\left(1 + 2\frac{B}{A}\kappa_0 p\right)^{1/2} - 1\right], \tag{A2}$$

and to first-order approximation this becomes[38]

$$\rho = \rho_0[1 + \kappa_0 p], \tag{A3}$$

where $\kappa_0 = 1/\rho_0 c_0^2$ is the ambient compressibility.

In order to arrive at an expression for κ, we combine the sourceless version of Eq. (2), i.e.,

$$\nabla \cdot \boldsymbol{v} + \kappa D_t p = 0, \tag{A4}$$

and the conservation law of mass

$$D_t \rho + \rho \nabla \cdot \boldsymbol{v} = 0, \tag{A5}$$

to obtain

$$\kappa D_t p = \frac{D_t \rho}{\rho}. \tag{A6}$$

With $D_t\rho = (\mathrm{d}\rho/\mathrm{d}p) D_t p$ this leads to

$$\kappa = \frac{1}{\rho}\frac{\mathrm{d}\rho}{\mathrm{d}p}. \tag{A7}$$

Upon substitution of Eq. (A2) for ρ we end up with an expression for $\kappa(p)$, from which we can easily derive the first-order approximation[38]

$$\kappa = \kappa_0[1 + \kappa_0(1 - 2\beta)p], \tag{A8}$$

where $\beta = 1 + B/2A$ is the coefficient of nonlinearity.

APPENDIX B: DERIVATION OF THE LOSSLESS WESTERVELT EQUATION FROM THE FIRST-ORDER ACOUSTIC EQUATIONS

Substitution of the constitutive parameters of Eqs. (3) and (4) into field equations (1) and (2), and discarding the terms that involve third and higher order products of p and/or $\boldsymbol{v}$ result in the set of equations

$$\nabla p + \rho_0 \partial_t \boldsymbol{v} = \boldsymbol{f} - \rho_0 \kappa_0 p \partial_t \boldsymbol{v} - \rho_0 (\boldsymbol{v} \cdot \nabla)\boldsymbol{v}, \tag{B1}$$

$$\nabla \cdot \boldsymbol{v} + \kappa_0 \partial_t p = q - \kappa_0^2 (1 - 2\beta) p \partial_t p - \kappa_0 \boldsymbol{v} \cdot \nabla p. \tag{B2}$$

For a rotation free field the substitution $\nabla \boldsymbol{v} = (\nabla \boldsymbol{v})^T$ is valid, and upon employment of the first-order approximation $\nabla p = -\rho_0 \partial_t \boldsymbol{v}$ in the second-order terms, and upon subsequent reordering of the terms, a new set of equations may be obtained as

$$\nabla p + \rho_0 \partial_t \boldsymbol{v} = \boldsymbol{f} - \nabla L, \tag{B3}$$

$$\nabla \cdot \boldsymbol{v} + \kappa_0 \partial_t p = q + \kappa_0^2 \beta \partial_t^2 p^2 + \kappa_0 \partial_t L, \tag{B4}$$

where

$$L = \tfrac{1}{2}\rho_0 \boldsymbol{v} \cdot \boldsymbol{v} - \tfrac{1}{2}\kappa_0 p^2 \tag{B5}$$

is the Lagrangian density. The second-order wave equation is obtained by subtracting the divergence of Eq. (B3) from ρ_0 times the temporal derivative of Eq. (B4). This directly leads to the source term $S = \rho_0 \partial_t q - \nabla \cdot \boldsymbol{f}$ in Eq. (5). Except for the source terms, Eqs. (B3)–(B5) are equivalent to the lossless versions of Eqs. (5)–(8) of Aanonsen *et al.*,[25] who showed that these equations lead to the Westervelt equation upon neglecting the local nonlinear effects that are represented by the terms involving L.

[1] B. Ward, A. C. Baker, and V. F. Humphrey, "Nonlinear propagation applied to the improvement of resolution in diagnostic medical ultrasound equipment," J. Acoust. Soc. Am. **101**, 143–154 (1997).
[2] F. Tranquart, N. Grenier, V. Eder, and L. Pourcelot, "Clinical use of ultrasound tissue harmonic imaging," Ultrasound Med. Biol. **25**, 889–894 (1999).
[3] A. Bouakaz and N. de Jong, "Native tissue imaging at superharmonic frequencies," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **50**, 496–506 (2003).
[4] A. Bouakaz, E. Merks, C. T. Lancée, and N. Bom, "Noninvasive bladder volume measurements based on nonlinear wave distortion," Ultrasound Med. Biol. **30**, 469–476 (2004).
[5] *Nonlinear Acoustics*, edited by M. F. Hamilton and D. T. Blackstock (Academic, San Diego, CA, 1998).
[6] K.-E. Frøysa, J. Naze Tjøtta, and S. Tjøtta, "Linear propagation of a pulsed sound beam from a plane or focusing source," J. Acoust. Soc. Am. **93**, 80–92 (1993).
[7] M. D. Cahill and A. C. Baker, "Numerical simulation of the acoustic field of a phased-array medical ultrasound scanner," J. Acoust. Soc. Am. **104**, 1274–1283 (1998).
[8] A. Bouakaz, C. T. Lancée, P. Frinking, and N. de Jong, "Simulations and measurements of nonlinear pressure field generated by linear array transducers," in Proceedings of the 1999 IEEE Ultrasonics (1999), pp. 1511–1514.
[9] V. A. Khokhlova, A. E. Ponomarev, M. A. Averkiou, and L. A. Crum, "Nonlinear pulsed ultrasound beams radiated by rectangular focused diagnostic transducers," Acoust. Phys. **52**, 481–489 (2006).
[10] X. Yang and R. Cleveland, "Time domain simulation of nonlinear acoustic beams generated by rectangular pistons with application to harmonic imaging," J. Acoust. Soc. Am. **117**, 113–123 (2005).
[11] P. T. Christopher and K. J. Parker, "New approaches to nonlinear diffractive field propagation," J. Acoust. Soc. Am. **90**, 488–499 (1991).
[12] J. Tavakkoli, D. Cathignol, and R. Souchon, "Modeling of pulsed finite-amplitude focused sound beams in time domain," J. Acoust. Soc. Am. **104**, 2061–2072 (1998).
[13] R. J. Zemp, J. Tavakkoli, and R. S. C. Cobbold, "Modeling of nonlinear ultrasound propagation in tissue from array transducers," J. Acoust. Soc. Am. **113**, 139–152 (2003).
[14] J. Wojcik, A. Nowicki, P. A. Lewin, P. E. Bloomfield, T. Kujawska, and L. Filipczynski, "Wave envelopes method for description of nonlinear acoustic wave propagation," Ultrasonics **44**, 310–329 (2006).
[15] T. Varslot and G. Taraldsen, "Computer simulation of forward wave propagation in tissue," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **52**, 1473–1482 (2005).
[16] P. D. Fox, A. Bouakaz, and F. Tranquart, "Computation of steered nonlinear fields using offset KZK axes," in Proceedings of the IEEE Ultrasonics 2005 (2005), pp. 1984–1987.
[17] T. Christopher, "Algorithm for the nonlinear propagation of acoustic beams from phased arrays and nonplanar sources," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **53**, 2188–2192 (2006).
[18] T. Varslot, S.-E. Masoy, T. Johansen, and B. Angelsen, "Aberration in nonlinear acoustic wave propagation," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **54**, 470 (2007).
[19] V. W. Sparrow and R. Raspet, "A numerical method for general finite amplitude wave propagation in two dimensions and its application to spark pulses," J. Acoust. Soc. Am. **90**, 2683–2691 (1991).
[20] T. Yano and Y. Inoue, "Strongly nonlinear waves and streaming in the near field of a circular piston," J. Acoust. Soc. Am. **99**, 3353–3372 (1996).
[21] M. Liebler, S. Ginter, T. Dreyer, and R. E. Riedlinger, "Full wave modeling of therapeutic ultrasound: Efficient time-domain implementation of the frequency power-law attenuation," J. Acoust. Soc. Am. **116**, 2742–2750 (2004).
[22] I. M. Hallaj and R. O. Cleveland, "FDTD simulation of finite-amplitude pressure and temperature fields for biomedical ultrasound," J. Acoust. Soc. Am. **105**, L7–L12 (1999).
[23] J. Hoffelner, H. Landes, M. Kaltenbacher, and R. Lerch, "Finite element

simulation of nonlinear wave propagation in thermoviscous fluids including dissipation," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **48**, 779–786 (2001).
[24]G. Wojcik, J. Mould, L. Carcione, and S. Ayter, "A study of second harmonic generation by focused medical transducer pulses," in Proceedings of the 1998 IEEE Ultrasonics (1998), pp. 1583–1588.
[25]S. I. Aanonsen, T. Barkve, J. Naze Tjotta, and S. Tjotta, "Distortion and harmonic generation in the nearfield of a finite amplitude sound beam," J. Acoust. Soc. Am. **75**, 749–768 (1984).
[26]M. D. Verweij and J. Huijssen, "A filtered convolution method for the computation of acoustic wave fields in very large spatiotemporal domains," J. Acoust. Soc. Am. **125**, 1868–1878 (2009).
[27]A. T. de Hoop, *Handbook of Radiation and Scattering of Waves* (Academic, San Diego, CA, 1995).
[28]J. Huijssen, "Modeling of nonlinear medical diagnostic ultrasound," Ph.D. thesis, Delft University of Technology, Delft, The Netherlands (2008).
[29]J. T. Fokkema and P. M. van den Berg, *Seismic Applications of Acoustic Reciprocity* (Elsevier, Amsterdam, 1993).
[30]G. F. Roach, *Green's Functions* (Cambridge University Press, Cambridge, 1982).
[31]J. A. DeSanto, *Scalar Wave Theory—Green's Functions and Applications* (Springer-Verlag, Berlin, 1992).
[32]R. O. Cleveland, M. F. Hamilton, and D. T. Blackstock, "Time-domain modeling of finite-amplitude sound in relaxing fluids," J. Acoust. Soc. Am. **99**, 3312–3318 (1996).
[33]M. Frigo and S. G. Johnson, "The design and implementation of FFTW3," Proc. IEEE **93**, 216–231 (2005).
[34]B. Fornberg, *A Practical Guide to Pseudospectral Methods* (Cambridge University Press, New York, 1996).
[35]J. de Koning, "Parallelizing Parnac," Technical Report No. TR06-03, VORtech Computing, Delft, The Netherlands, 2007.
[36]J. Huijssen, A. Bouakaz, M. D. Verweij, and N. de Jong, "Simulations of the nonlinear acoustic pressure field without using the parabolic approximation," in Proceedings of the IEEE Ultrasonics 2003 (2003), pp. 1851–1854.
[37]Y.-S. Lee and M. F. Hamilton, "Time-domain modeling of pulsed finite-amplitude sound beams," J. Acoust. Soc. Am. **97**, 906–917 (1995).
[38]J. Huijssen and M. D. Verweij, "Nonlinear constitutive equations derived for fluids obeying an ideal gas, a Tait-Kirkwood or a B/A type equation of state," in Proceedings of the ISNA17 (2005), pp. 552–555.

Dislodgement and removal of dust-particles from a surface by a technique combining acoustic standing wave and airflow

Di Chen and Junru Wu
Department of Physics and Materials Science Program, The University of Vermont, Burlington, Vermont 05405

(Received 18 April 2009; revised 16 October 2009; accepted 19 October 2009)

It is known that there are many fine particles on the moon and Mars. Their existence may cause risk for the success of a long-term project for NASA, i.e., exploration and habitation of the moon and Mars. These dust-particles might cover the solar panels, making them fail to generate electricity, and they might also penetrate through seals on space suits, hatches, and vehicle wheels causing many incidents. The fine particles would be hazardous to human health if they were inhaled. Development of robust dust mitigation technology is urgently needed for the viable long-term exploration and habilitation of either the moon or Mars. A feasibility study to develop a dust removal technique, which may be used in space-stations or other enclosures for habitation, is reported. It is shown experimentally that the acoustic radiation force produced by a 13.8 kHz 128 dB sound-level standing wave between a 3 cm-aperture tweeter and a reflector separated by 9 cm is strong enough to overcome the van der Waals adhesive force between the dust-particles and the reflector-surface. Thus the majority of fine particles (>2 μm diameter) on a reflector-surface can be dislodged and removed by a technique combining acoustic levitation and airflow methods. The removal efficiency deteriorates for particles of less than 2 μm in size.
 [DOI: 10.1121/1.3268507]

PACS number(s): 43.25.Uv, 43.25.Gf [RMW] Pages: 45–50

I. INTRODUCTION

Exploration and habitation of the moon and Mars have become a long-term project for NASA (Calle *et al.*, 2008). The presence of fine particles in Martian and lunar soil poses a significant threat to the success of the project. Since there is no water or other liquids in either environment, the mundane operation of washing away the fine particulates from any surface becomes impractical. The majority of particulates existing on the moon and Mars are silicon dioxides (SiO_2). The particle-size range is broad; it can be as large as submillimeter and as small as 10 nm. It has been reported that transient dust clouds exist above the lunar surface even the atmosphere (mixture of H, He, Ar, Na, and K) pressure is very low on the moon surface (Stern, 1999). It was suggested that the possible physical mechanism may be as follows: The dust-particles on the moon may be electrically charged by charged particles from the solar wind as well as UV radiation (Horanyi, 1998). Consequently, levitation and migration of dust particles may take place near lunar surface. The solar panels brought to the moon during exploration by the astronauts, from that the electricity can be generated, may be quickly covered by dust-particles, making the electrical equipment nonoperational. Furthermore, the fact that dust-can penetrate through seals on space suits, hatches, vehicle wheels, etc., may also become troublesome. Furthermore, the lunar and Martian fine particles are hazardous to health by causing respiratory disease if they are inhaled (Schrunk *et al.*, 1999); removal of dust-particles from these surfaces should be important.

Development of robust dust mitigation technology is urgently needed for the viable long-term exploration and habitation of either the moon or Mars. Such technology should be effective for dislodging particles of diameter ranging from 0.01 to 100 μm from a surface of a solar panel or space suits, etc. Existing dust mitigation technology, such as electrostatic precipitators and mechanical filters, either uses large concentration of ions or requires frequent filter changes or manual cleaning.

It has been reported that lunar or Martian environment is very different from that on Earth. For example, more than 95% of Martian atmosphere is carbon dioxide. Mean surface level pressure is 600 Pa, only 0.6%–0.7% of the Earth atmosphere. The lunar atmosphere is even less significant in comparison with that of Mars and the Earth. Therefore, application of acoustic technique in an open-space would be ineffective particularly for lunar applications. However, in a space station or an equivalent enclosure for research and habitation, the atmosphere condition should be very similar to that of the Earth. Thus the acoustic method may provide an alternative surface cleaning technique.

This study is designed to investigate the feasibility of using acoustic levitation method in dislodging dust from a dust-covered-surface possibly related to exploration and habitation of the moon and Mars.

II. THEORY

A one-dimensional incident traveling plane-wave with frequency f and wavelength λ propagating along x direction (vertical downward, as shown in Fig. 1) can be represented by $u(x,t)=u_0\cos(kx-\omega t)$, where $k=2\pi/f$, $\omega=2\pi f$, $u(x,t)$, and u_0 are instantaneous particle velocity and particle velocity amplitude, respectively. Now when we place a perfect

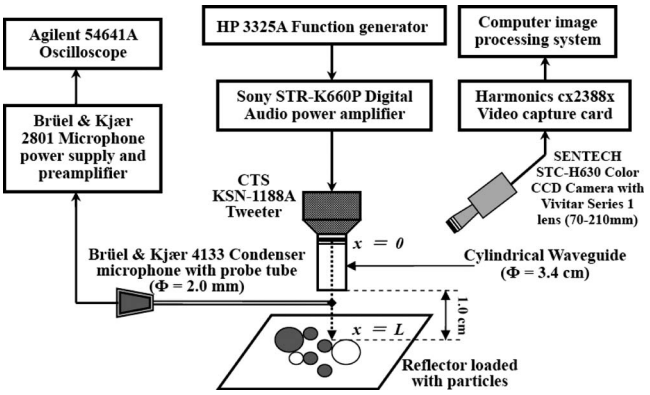


FIG. 1. Illustration of experimental setup (not scaled).

plane reflector at $x=L$, perpendicular to the wave propagation direction and ignore the attenuation, letting the incident wave be superposed on another traveling in the $-x$ direction (reflected plane-wave), $u_0\cos(kx+\omega t)$, we obtain a standing wave given by $u=2u_0\cos kx\cos \omega t$. The boundary condition at the reflector (assuming that the reflector is rigid) requires $kL=(n+1/2)\pi$, i.e., $L=(n+1/2)\lambda/2$. Particle velocity antinodes (maxima) or pressure nodes are found when $\cos kx=\pm 1$, $x=0,\lambda/2,\lambda,3\lambda/2,\ldots$. Likewise, pressure antinodes (maxima) or velocity nodes are found where $\cos kx=0$ at positions $x=\lambda/4,3\lambda/4,5\lambda/4,\ldots$. The acoustic pressure antinodes should locate at $L=\lambda(2n+1)/4$, where n are integers. The spacing between two nearest pressure nodes or pressure antinodes is $\lambda/2$ (wavelength/2).

The second order nonlinear effect of a sound field can produce a force on a particle in a fluid exposed to the sound field. The basis for such forces was expressed by Gorkov (1962) in an equation for the acoustic radiation force on a spherical particle of volume v in a medium where the time-averaged densities of the kinetic energy density $\langle E_k\rangle$ and potential energy $\langle E_p\rangle$ are known. Letting β be the ratio of the compressibility of the particle material and the surrounding medium, and ρ_s and ρ_o are the densities for the particle and the surrounding medium, respectively, Gorkov's equation, applied to the x component F_x of the force, is

$$F_x = vDd\Phi/dx, \tag{1}$$

where $D=3(\rho_s-\rho_o)/(2\rho_s+\rho_o)$ and Φ is called the acoustic force potential and is given by

$$\Phi = \langle E_k\rangle - (1-\beta)\langle E_p\rangle. \tag{2}$$

Here the time-averaged kinetic energy $\langle E_k\rangle$ is proportional to the square of the particle velocity amplitude. Since the velocity amplitude is just $2u_0\cos kx$ in a planar standing wave field, maxima of $\langle E_k\rangle$ occur one-half wavelength apart, i.e., when $\cos kx$ is equal to ± 1. Since the net energy transported in a perfect standing wave is zero, it follows that $d\langle E_p\rangle/dx$ is equal and opposite to $d\langle E_k\rangle/dx$. Since the dust-particles are denser and less compressible than the surrounding fluid, $D>1$ and $\beta<1$. Hence F_x is in the direction of increasing $\langle E_k\rangle$ or decreasing $\langle E_p\rangle$. In other words, theory predicts that when a plane standing sound wave is established in space, the dust-particles will migrate toward planes where acoustic pressure amplitude is minimum. These planes are separated by a distance of $\lambda/2$ (one-half wavelength). If the particles lodge on a reflector located at the pressure antinodes (velocity nodes), they may be levitated and move to the pressure nodes (velocity antinodes).

We have used a circular rigid walled waveguide to couple with a sound source (a tweeter). For this type of waveguide of radius a to satisfy the normal particle velocity being zero at $r=a$, the acoustic pressures of possible modes may be written as (Kinsler *et al.*, 1982)

$$p_{ml} = A_{ml}J_m(k_{ml}r)\cos(m\theta)\exp[i(\omega t - k_z z)]. \tag{3}$$

Here J_m is the mth-order Bessel function

$$k_z = \sqrt{\left(\frac{2\pi f}{c}\right)^2 - k_{ml}^2}, \tag{4}$$

and

$$k_{ml} = j'_{ml}/a, \tag{5}$$

where j'_{ml} is the lth root of $dJ_m(x)/dx=0$.

The possible modes, which can propagate in the waveguide excited by a sound source with circular symmetry and its maximum vibration velocity amplitude locating at the center, may be limited to $m=0$ modes. For $m=0$ and $n=1$, it is a plane-wave which can always propagate in the waveguide with a phase velocity $c=343$ m/s in air at 20 °C. In addition to this plane-wave mode, it is possible to have the excitation frequency f of the acoustic source at $z=0$, satisfying the following condition: $ck_{02}/2\pi<f<ck_{03}/2\pi$ (Kinsler *et al.*, 1982), the nonplanar (0, 2) mode may also propagate in the waveguide. Its mathematic expression is given by

$$p_{02} = A_{02}J_0(k_{02}r)\exp[i(\omega t - k_z z)], \tag{6}$$

where $k_{02}=j'_{02}/a=3.83/a$. For our application, $a=1.7$ cm, $c=343$ m/s (Kinsler *et al.*, 1982), $k_{02}=3.83/1.7$ cm, $k_{03}=7.02/1.7$ cm, $f_{02}=ck_{02}/(2\pi)=12.3$ kHz, and $f_{03}=ck_{03}/(2\pi)=22.5$ kHz. Since the sound source (tweeter) has the highest transmitting efficiency at $f=13.8$ kHz, thus it was chosen for the excitation frequency to maximize the acoustic radiation force, which is between f_{02} and f_{03}. The (0, 2) mode phase velocity $c_{p,02}=c/\sqrt{1-(f_{02}/f)^2}=756$ m/s and its wavelength along the propagation direction $\lambda_{02}=c_{p,02}/f=5.48$ cm (Kinsler *et al.*, 1982). For the planar wave, $\lambda=c/f=2.48$ cm.

III. EXPERIMENTAL METHODS AND RESULTS

The main experimental setup (Fig. 1) consists a tweeter as a sound source (KSN-1188A, CTS, Elkhart, IN) coupled with a cylindrical waveguide, a plane reflector, loaded with Mars (JSC Mars-1A, Planet LLC, Madison, WI), or lunar (JSC-1A, Planet LLC, Madison, WI) dust simulants installed perpendicular to the acoustic axis of the tweeter by keeping $L=9\pm 0.1$ cm, and a color charge coupled device (CCD) camera (STC-H630, SENTECH, Carrollton, TX) used to monitor the movements of the dust simulants. The sinusoidal electric signal of 13.8 kHz generated by a function generator (HP 3325, Hewlett-Packard, Palo Alto, CA) and amplified by an audio power amplifier (Sony STR-K660P amplifier, New

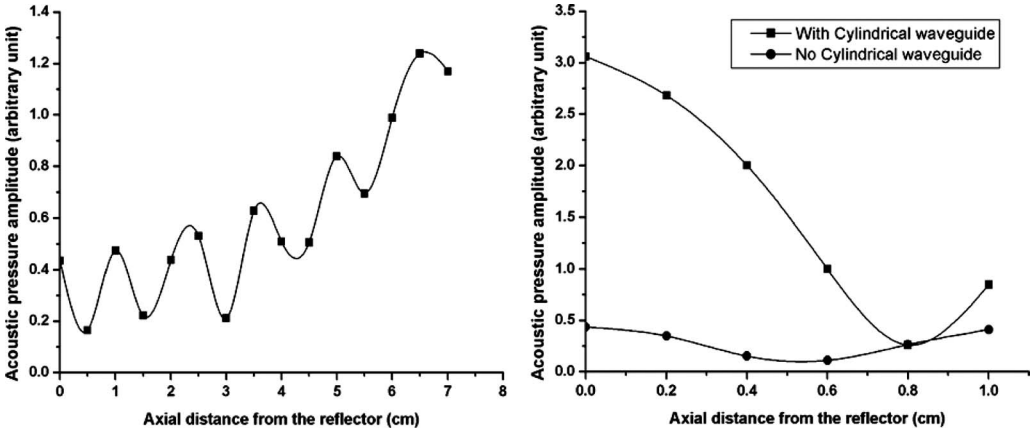


FIG. 2. Microphone measurements. Left: Acoustic pressure amplitude between the tweeter without cylindrical waveguide and silicon wafer. Right: Acoustic pressure amplitude is significantly improved by using a cylindrical waveguide coupled to the tweeter.

York, NY) was applied to the tweeter to establish a standing wave field between the tweeter and reflector. A $\frac{1}{4}$ in. microphone (model 4133, Brüel & Kjær, Denmark) coupled with a hollow 2 mm diameter metallic tube (Fig. 1) and associated preamplifier was used to measure sound pressure amplitude as a function of axial distance from the reflector. The output of the preamplifier was connected to the input of a digital oscilloscope (54641A, Agilent Technology, Palo Alto, CA). Activities of the dust-particles on the reflector were monitored through the CCD camera with a lens system (Vivitar Series 1, 70–210 mm, Vivitar Co., Edison, NJ). A video capture card (Cx2388x, Harmonics, Newport Beach, CA) was used to capture the video signals which was then converted and stored in a digital format into a computer for further processing.

The acoustic pressure amplitude was measured spatially every 0.5 cm step upward from the reflector. The best-fit solid curve for the acoustic pressure amplitudes vs the distance between the reflector and tweeter was obtained using the "spline" mode of a software (ORIGIN 6.0, OriginLab Corporation, Northampton, MA), as shown in Fig. 2 (left panel); it suggests that a standing wave acoustic field was established in space. To further enhance the acoustic field, a cylindrical lucite waveguide (8 cm long, 3.4 cm diameter) was then closely coupled to the tweeter. The axial acoustic pressure amplitude near the reflector-surface was measured and was found to be significantly enhanced by about sixfold or by 15.6 dB magnification, as shown in Fig. 2 (right panel). Due to the presence of the waveguide, the space limitation only allowed us to measure acoustic pressure amplitude in a small region near the reflector where the waveguide was terminated. We expected that the similar enhancement would take place within the waveguide. The spatial-averaged acoustic intensity over $\frac{1}{4}$ in. diameter cross-section was estimated by using a $\frac{1}{4}$ in. sound-level meter with the linear mode setting (model 215, Quest Electronics, Oconomowoc, WI). The spatial-averaged acoustic intensity generated by the tweeter with the cylindrical waveguide was 128 dB, 8 dB higher than that without the waveguide. The waveguide played a role of concentration of sound energy in space and also altered the wavelength as it left the waveguide and propagated into open-space, as described in Eq. (4) and demonstrated on the right panel of Fig. 2. It was observed that the majority of dust-particles were levitated in the acoustic standing wave field. It is suggested that the acoustic radiation force was strong enough to overcome gravity of the particle and the adhesive van der Waals force between the particles and reflector-surface. We then turned on the airflow generated by a small fan (diameter of 13.5 cm). The dust-particles were quickly blown away after they were levitated from the surface. Without the acoustic field, the airflow alone had no effect on the dust-particles lodged on the reflector-surface due to the strong van der Waals force.

We studied the particle-removal efficiency on four different reflectors of 9×9 cm^2 size made of different materials: silicon wafer, commercial solar panel, synthetic leather, and Teflon, as shown in Fig. 3. All these four different reflectors were placed on top of an aluminum square block (10×10 cm^2). All parameters including L, frequency, and acoustic intensity were kept the same for all four materials. Each reflector was exposed to the acoustic excitation and airflow for 90 s during the removal operation. Size-

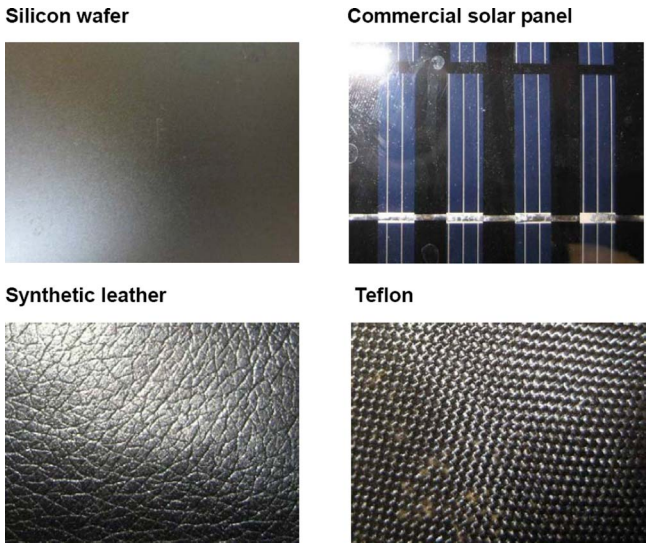


FIG. 3. (Color online) Photos of four reflector-surfaces of 9×9 cm^2 size made of different materials: silicon wafer, commercial solar panel, synthetic leather, and Teflon.

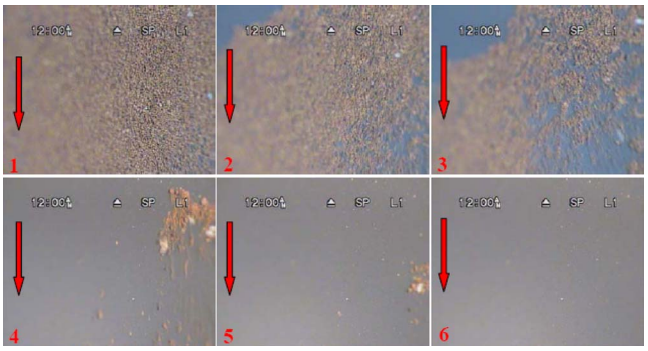


FIG. 4. (Color online) Mars dust simulants lodged on the silicon wafer were removed by airflow after they were levitated in a standing acoustic field. The airflow had no effect on the dust-particles without the acoustic field, as shown in image 1.

distributions of the residual particles after removal on the reflectors were studied by direct counting using a microscope (Nikon LABOPHOT-2, Nikon Instruments Inc., Melville, NY).

Figure 4 contains six images taken in sequence when silicon wafer was chosen as the reflector. Arrow represents the airflow direction. Airflow that was continuously on for all six images was taken, and acoustic signal was continuously turned on starting at image 2. There was no particles' movement at the time when image 1 was taken when only airflow was on. Particles began to be removed as soon as the tweeter was turned on, as shown in images 2–6. Mars dust simulants were removed effectively by the airflow after they were levitated by the standing wave acoustic field.

Figures 5(a) and 5(b) are typical microscopy images of Mars and lunar dust simulants accompanied with the

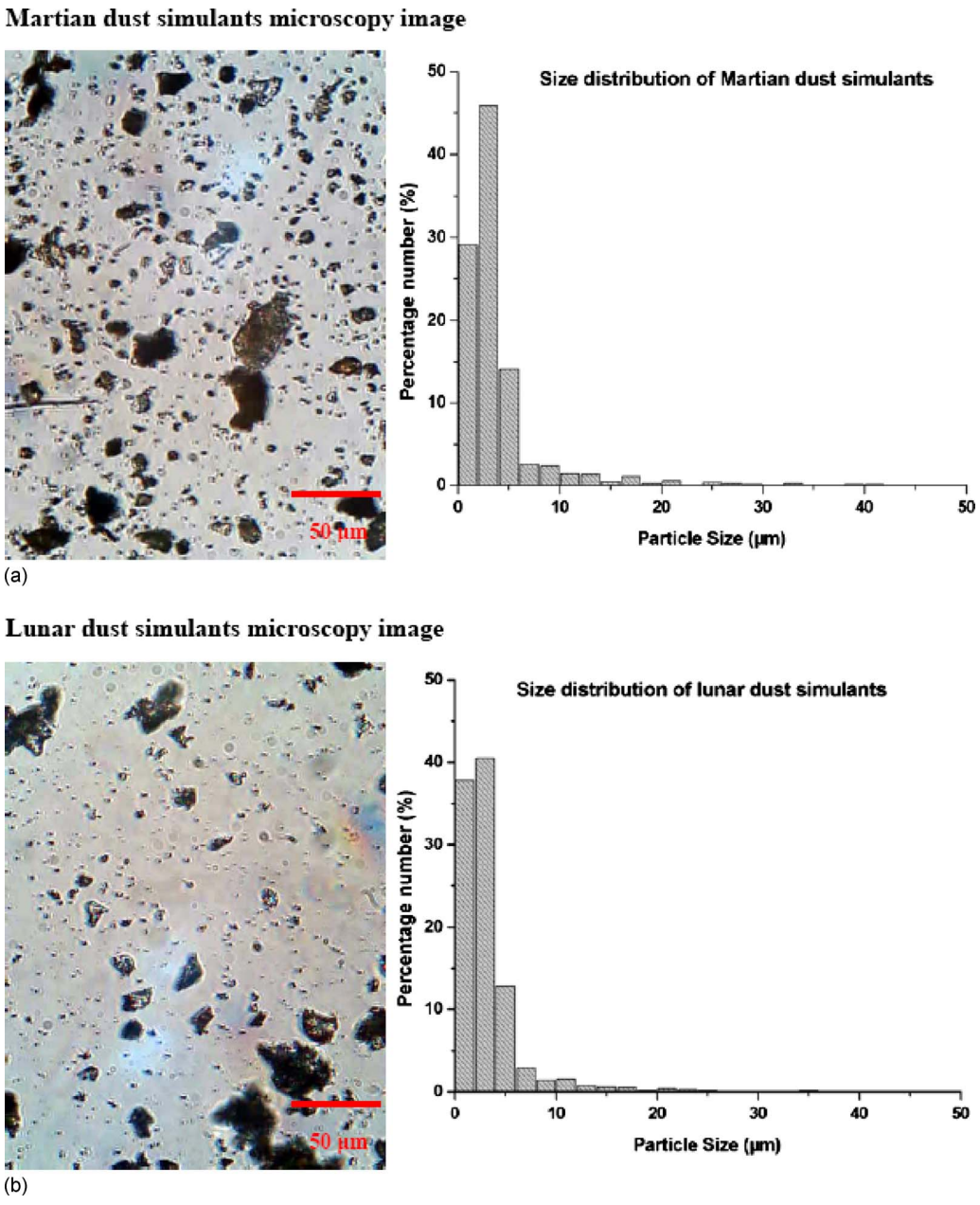


FIG. 5. (Color online) Typical microscopic images of (a) Mars and (b) lunar dust simulants accompanied with the percent-number size-distribution histograms shown on their right.

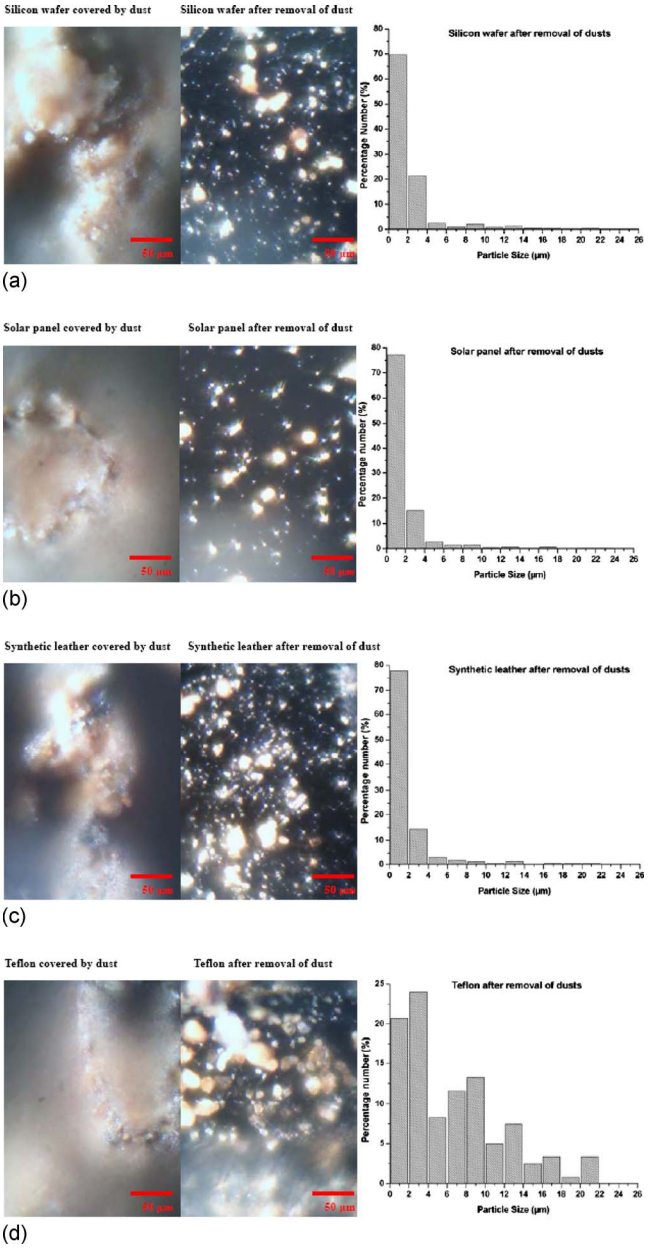


FIG. 6. (Color online) Dust-particle-removal efficiencies for the four target-materials (silicon wafer, commercial solar panel, synthetic leather, and Teflon).

percent-number size-distribution histograms shown on the right. The histograms of Mars and lunar dust simulants indicate that they both have a large component of particles with diameter less than 6 μm, especially in the 2–4 μm range. The percent-number of Mars dust stimulants for >4 μm is about 25%, and that for 2 μm<particle-size<4 μm is about 45%. The lunar dust simulant has a similar percent-number (12% vs 14%) as the Mars dust simulant in the size range >4 μm, a bit lower percent-number (40% vs 45%) in the size range 2–4 μm, and it has more particles in the range <2 μm (38%) than Martian dust simulant has (29%). For further experimental tests, we decided that only Martian dust simulants would be used.

The dust-particle-removal efficiencies for the four target-reflectors are illustrated in Fig. 6. There are three pan-

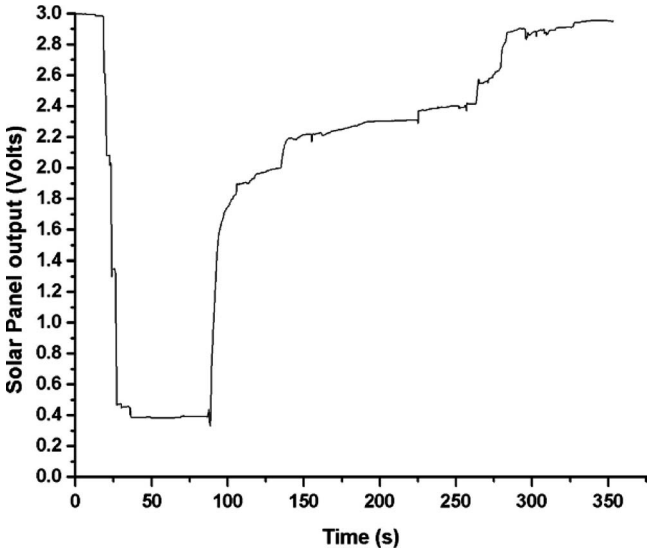


FIG. 7. Solar-panel voltage output vs time. At t=25 s, the Mar dust simulants were spread on the solar panel and output-voltage dropped from 3.0 to 0.4 V. At t=90 s, the tweeter and the airflow were tuned on. As the dust-particles were removed, the output-voltage gradually recovered to 2.95 V. With the airflow alone, no effect was observed.

els for each of Figs. 6(a)–6(d); the first two are photos of a reflector covered by Mars dust simulant taken under the microscope before and after the dust-simulant removal operation by airflow and acoustic standing wave; the third panel on the right is the size-distribution histogram of the particle residual on the reflector after the removal procedure. Few particles larger than 4 μm were left (from more than 25% before the operation to less than 9% after the operation) on the silicon wafer, solar panel, and synthetic leather after the removal operation, as shown in Figs. 6(a)–6(c). Particle-number in the range of 2–4 μm was also dramatically reduced from 46% to around 20% on silicon wafer and 15% on the solar-panel and leather surfaces. Small particles whose size was less than 2 μm dominated in the residuals. When Teflon was used as the target-material, the particle-removal efficiency decreased significantly, as shown in Fig. 6(d). However, we were still able to see Mars dust simulants bouncing around the Teflon material under the excitation of acoustic standing wave. Since the fabric texture is loose and rough, a lot of large and small particles were trapped in the texture.

Figure 7 shows the solar-panel output-voltage measured by a digital multi-meter (model 2000, Keithley Instruments, Inc., Cleveland, OH) as function of time. It reflects that the voltage output changed with time as the Mars dust simulants were sprayed on the surface and gradually removed from it by the combination of acoustic standing wave effect and airflow method. There was a significant drop in solar-panel output-voltage (below 15% of its initial voltage) due to the deposition and coverage of Mars dust simulants. The output-voltage remained constant in time until the airflow and acoustic field were applied at t=90 s. The output-voltage of the solar panel restored quickly to around 65% in the first 20 s and then increased gradually to 98.4% after 4 min. No further improvement could be observed after 4 min.

IV. SUMMARY AND DISCUSSION

Our feasibility study has shown that the majority particles of the dust simulants can be levitated from a surface by the radiation force due to a standing-wave acoustic field. It has been observed that significant reduction in Mars dust simulants lodged on the targeted surfaces of silicon wafers, solar panels, and synthetic leather planes few minutes after using the acoustic technique incorporated with airflow. It has also been observed that the removal efficiency becomes lower for submicron size particles. As shown in our previous work (Wu and Du, 1990), acoustic radiation force is proportional to the a^3 (a is particle radius). For smaller particles, the acoustic radiation force becomes too weak compared to the strong adhesive van der Waals force between the particles and the target-surface.

Equation (1) suggests that the radiation force is proportional to the gradient of acoustic force potential Φ. In order to increase the radiation force, it is advantageous to increase the frequency or decrease the wavelength. On the other hand, the acoustic attenuation increases with frequency. It was found experimentally that the sound source (tweeter) has highest transmitting efficiency near 13.8 kHz. Thus, $f=13.8$ kHz was chosen for the excitation frequency. Another advantage for the higher frequency (>10 kHz) is that human ears are less sensitive above 10 kHz; therefore high amplitude 10 kHz sound wave would cause less annoyance than the same sound level of kilohertz sound waves.

As discussed earlier, the possible modes propagating in the waveguide which can be excited by a sound source with circular symmetry and maximum vibration velocity at the center may be limited to $m=0$ modes. Since experimentally we found that the sound source has the maximum transmitting efficiency at frequency of 13.8 kHz, we chose $f=13.8$ kHz for the excitation frequency between f_{02} and f_{03} to maximize the levitation efficiency. The wave propagating in the waveguide is the superposition of a planar (0, 1) wave and a nonplanar (0, 2) wave. If it were a pure planar wave, $\lambda/2=c/2f=34\,300/(2\times 13\,800)=1.24$ cm. For (0, 2) mode, $\lambda_{02}=c_{p02}/(2f)=2.74$ cm. Experimentally, the location of the reflector ($L=9$ cm) was determined by maximizing levitation effect (maximizing the acoustic radiation force) during the experiment. According to the planar waveguide theory described earlier, the distance between the source and the pressure antinode L is given by $L=\lambda(2n+1)/4$. Thus $\lambda/2=L/(n+1/2)=9\ \text{cm}/(7+0.5)=1.20$ cm, where $n=7$. The fact that the error incurred for the measured wavelength compared with the pure planar wave being quite small, equal to $(1.24-1.20)/1.24\approx 3\%$, implies that the reflector's location by maximizing the levitation force was nearly at one of the pressure antinodes of a standing wave generated by superposition of the pure planar wave and its reflection. It suggests that in the waveguide of our experiment, the planar wave may dominate even one nonplanar wave exists. Strictly speaking, the velocity amplitude being $2u_0 \cos kx$ in a standing wave field from Gorkov's equation was based on a plane-wave assumption. Theoretical modification might be needed to the current experimental situation since the wave was no longer a pure planar wave. On the other hand, as suggested by above argument, the planar wave might still dominate in our experiment, so it might still be approximately considered a planar wave for our experimental situation.

As stated earlier, this study was performed under the Earth atmosphere condition. The technique may be limited to be applied in a space station or other enclosures for habitation. The condition on the moon and Mars is quite different. In order to use it in an open-space, a gas tank is needed to be bought there and the tweeter-waveguide-reflector system needs to be sealed.

ACKNOWLEDGMENTS

This work has been partially supported by NASA under Cooperative Agreement No. NNX08AZ07A. Sincerely thanks to a reviewer for his careful review and constructive suggestions.

Calle, C. I., McFall, J. L., Buhler, C. R., Snyder, S. J., Arens, E. E., Chen, A., Ritz, M. L., Clements, J. S., Fortier, C. R., and Trigwell, S. (**2008**). "Dust particle removal by electrostatic and dielectrophoretic forces with applications to NASA exploration missions," in Proceedings of the ESA Annual Meeting on Electrostatics, Minneapolis, MN, pp. 1–14.

Gorkov, L. P. (**1962**). "On the forces acting on a small particle in an acoustic field in an ideal fluid," Sov. Phys. Dokl. **6**, 773–775.

Horanyi, M. (**1998**). "Electrostatic charging of lunar dust," Lunar Planet. Sci. **29**, 1527–1534.

Kinsler, L. E., Frey, A. R., and Sanders, J. V. (**1982**). *Fundamentals of Acoustics*, 3rd ed. (Wiley, New York).

Schrunk, D. G., Sharpe, B. L., Cooper, B. L., and Thangavelu, M. (**1999**). *The Moon: Resources, Future, Development and Colonization* (Praxis, Chichester).

Stern, S. A. (**1999**). "The lunar atmosphere: History, status, current problems and context," Rev. Geophys. **37**, 453–491.

Wu, J., and Du, G. (**1990**). "Acoustic radiation force on a small compressible sphere in a focused beam," J. Acoust. Soc. Am. **87**, 997–1003.

Finite amplitude method for measuring the nonlinearity parameter B/A in small-volume samples using focused ultrasound

Shigemi Saito[a)]
School of Marine Science and Technology, Tokai University, 3-20-1 Orido, Shimizu-ku, Shizuoka 424-8610, Japan

(Received 13 August 2008; revised 16 October 2009; accepted 28 October 2009)

On the basis of finite amplitude and comparative methods, the acoustic nonlinearity parameter B/A of a liquid sample of as small as 0.1 ml is measured using an 18.6-MHz focused Gaussian beam. The sample fills the space between a polystyrene plate and a tungsten reflector set about 1 mm apart from each other within the focal region. The sound speed c and attenuation coefficient α are determined using the time of flight and the insertion loss of the sound passing through the sample, respectively. The density ρ is estimated from the reflection coefficient at the interface between the polystyrene plate and the sample, where the transformation from longitudinal to transverse waves is considered. To compensate for the effect of velocity dispersion on the second harmonic generation, the relative phase of the second harmonic sound is also measured using dual-frequency sound. By summarizing all the linear properties and amplitude data of the second harmonic component in the sound transmitted through the sample, the B/A value is finally determined. The measurement is validated through the experiments on nondispersive liquids and weakly dispersive biological samples with known B/A values.
 [DOI: 10.1121/1.3268602]

PACS number(s): 43.25.Zx, 43.25.Ba, 43.80.Ev, 43.35.Sx [ROC] Pages: 51–61

I. INTRODUCTION

The acoustic nonlinearity parameter B/A is sensitive to early-stage tissue lesions because it reflects the molecular structure of a material and depends on the concentrations of water and fat particularly in biological tissues.[1–7] According to previously measured results, B/A measurements can exhibit a range of values for a given tissue type in addition to the variation among different types of tissue.[8] Consequently, even a microscopic variation of B/A may exist in biological tissues and its imaging may provide pathological information similar to other linear acoustic imaging techniques. For this purpose, B/A should be evaluated in each small segment of a sample. Although a nonlinear acoustic microscope has been developed for second harmonic imaging, the generated image was not that of B/A.[9] On the basis of the thermodynamic method[10] and the finite amplitude method,[11] various measurement methods have been developed for determining B/A.[8,12–19] Among these, the isentropic phase method, where the sound speed variance during adiabatic pressure change is precisely measured through the phase shift of the transmitted sound (or the frequency shift to keep the phase constant), has the highest accuracy for evaluating B/A and is within 1%.[16] Since a relatively wide sound beam propagating across a relatively large sample is currently employed in those measurements, the averaged value in the region can be well measured. However, a nonuniformly distributed B/A property in a small region cannot be observed unless the measurement method and apparatus are modified. In this paper, a finite amplitude method is investigated that can generate a C-mode image to show the lateral distribution of B/A as well as other linear properties.

In this study, the high lateral resolution of a focused beam within the focal region is utilized as in the acoustic microscope. The author previously presented the finite amplitude method of determining B/A, with the nonlinearly generated second harmonic component in a 1.9-MHz focused sound beam, that was transmitted through a 20-mm-thick sample placed in the focal region.[20,21] This is a comparative method, wherein the acoustic properties are determined by comparing with those of distilled water, whose properties are well known. On the basis of this system, a new acoustic system for B/A measurement has been developed. In addition to reducing the size by employing a narrower beam in a high-frequency (HF) band, the measurement structure is simplified by enabling an ultrasonic transducer to detect the second harmonic by itself.[22,23] Furthermore, a new manner of inserting the sample into the focal region is developed. In the previous method, a simple sound field consisting of a focused Gaussian beam at 1.9 MHz without any oscillatory peaks or dips induced by diffraction and interference was used to facilitate the measurements. This useful property of a Gaussian beam is retained in the new system.

In the finite amplitude method, B/A is estimated with a second harmonic component accumulated with propagation, so the sample thickness equivalent to the propagation distance of the sound is an important item to be measured. In the previous method, a flexible film was employed for the acoustic window to isolate the liquid sample from the water

a)Author to whom correspondence should be addressed. Electronic mail: ssaito@scc.u-tokai.ac.jp

couplant, and the sample thickness was not always constant. Hence, the thickness had to be measured each time particularly if the sample was thin.[24,25] Furthermore, it was almost impossible to flatten the surface of somewhat rigid samples. The premise of a flat interface between the sample and water couplant, on which the theoretical analysis is based, was not always met. If some rigid plate could be used as the acoustic window, such inconveniences would be completely eliminated. A solid plate is used as the acoustic window in the present method. A drawback of the solid acoustic window, namely, that it may significantly degrade the precision of measurements, is addressed in this paper. In addition, the velocity dispersion inherent in biological samples at frequencies in the HF range considerably affects the accumulation of the second harmonic[26] and is addressed here. The method of correcting the estimated B/A in weakly dispersive samples is also presented. Thus, the finite amplitude method for a liquid sample of approximately 1-mm thickness and 0.1-ml volume is developed, where the linear acoustic properties such as the sound speed, density, and attenuation coefficient are simultaneously measured.

The estimation of B/A in liquids has been carried out elsewhere with the second harmonic component generated in a HF focused sound.[27,28] However, in these studies, the B/A values cannot be determined unless the linear properties such as density and attenuation coefficient are already known. An excellent method of measuring B/A by applying a null method to the second harmonic generation in a liquid sample inserted in a wave guide of two silica rods has also been developed using very-high-frequency ultrasound.[29] However, its low lateral resolution as well as the need for samples of variable thickness are not appropriate for the present purpose of visualizing the B/A distribution in biological soft tissues.

II. MEASUREMENT SYSTEM

Figure 1(a) shows the cross-sectional view of the acoustic system used in this work. A 0.19-mm-thick 36°-rotated Y-cut $LiNbO_3$ (LN) transducer, with an inverted polarization layer fabricated by heat treatment, radiates and detects the fundamental signal (18.6 MHz) and the second harmonic (37.2 MHz).[30–32] This transducer has 3-dB bandwidths of 3 and 2 MHz around the center frequencies of 18.6 and 37.2 MHz, respectively. The sensitivity at 37.2 MHz is 85% of that at 18.6 MHz.[33] This LN transducer is bonded on the center of a circular disk of synthetic silica glass of 7.76-mm thickness and 100-mm diameter where a circular electrode of 8 mm diameter has been evaporated. The electrode on the other side of the LN transducer has been fabricated in a twisted star shape of 30 points.[23] The on-source pressure distribution is approximated as Gaussian by $p_0 \exp(-\xi r^2)$ with $\xi = 1.7 \times 10^5$ m^{-2} and $p_{02} \exp(-\xi_2 r^2)$ with $\xi_2 = 9 \times 10^4$ m^{-2} at $f = 18.6$ MHz and $2f = 37.2$ MHz, respectively, where r is the radial distance.[23,34,35] The radiated sound field is a Gaussian beam both at f and $2f$. A concavity with an aperture radius of $a = 5.0$ mm and curvature radius of $b = 8.0$ mm, designed to work as a focusing lens, is provided on the other side of the glass disk, in which the density ρ_0 is 2203 kg/m^3 and the sound speed c_0 is 5953 m/s.[36] The attenuation for the sound transmitting through the lens at its center follows a frequency squared law and is quite low, 2.8×10^{-5} Np at 18.6 MHz.[36]

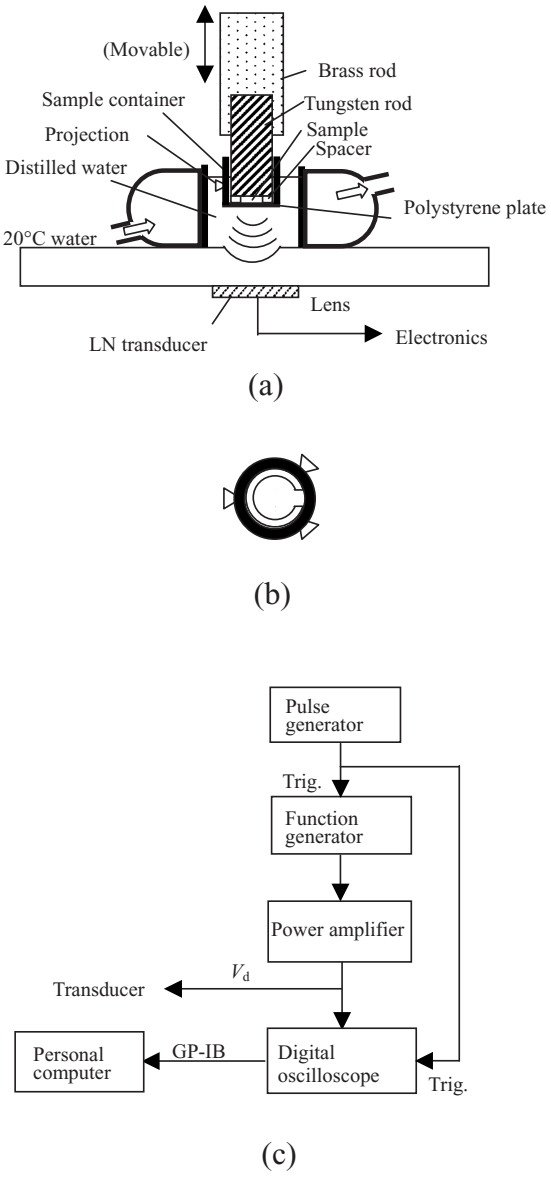


FIG. 1. Experimental setup for measuring nonlinearity parameters of liquid-like samples: (a) cross-sectional view of acoustic system, (b) bottom view of sample container with spacer, and (c) block diagram of electronics.

A cuplike container into which 0.1 ml of sample liquid is poured is made from a brass tube of 11-mm inner diameter and 15-mm length. As a base of the container, an optically transparent polystyrene plate with a thickness d of about 1 mm, in which the density ρ_d, the longitudinal sound speed c_d, and the transversal sound speed c_t are 1049 kg/m^3, 2333 m/s, and 1120 m/s, respectively,[35] has been attached with a double-adhesive tape to function as an acoustic window. A polystyrene plate-container, characterized by the smallest transmission loss and velocity dispersion among many commercially available polymer plates,[35] is used to window the sample. A Teflon-coated stainless-steel ring spacer of 8-mm inner diameter and 1-mm thickness is put at the bottom, as illustrated in Figs. 1(a) and 1(b). A tungsten rod of 10-mm diameter whose end has been optically pol-

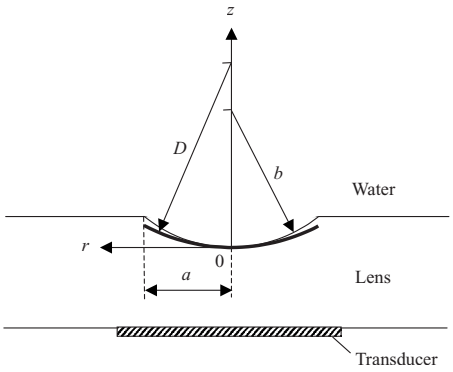


FIG. 2. Analytical model of acoustic system.

ished is immersed into the sample through the upper opening. The sample thickness L is determined by the spacer thickness. In addition to the partial cut-off of the spacer, the high repellency of Teflon against liquid is useful to avoid having bubbles in the sample. The sample container is clamped with three rubber bands suspended from a 16-mm-diameter brass rod and hooked onto three projections provided on the side of the container. For nonliquid and potentially irregular-shaped samples such as biological soft tissues, a 5×5 mm^2 square form slightly thicker than 1 mm is used in the container, which has been filled with about 0.1 ml of 0.9% sodium chloride solution (saline). The sample is pressed between the tungsten rod and polystyrene plate with the tension of rubber bands so that the sample layer has the same thickness as a liquid sample. The volume of the container to be filled with the sample is less than 0.1 ml. The excess liquid sample or saline overflows into the space between the side of the tungsten rod and the inner side of the container. The brass rod sustaining the tungsten rod is attached to a mechanical stage (Suruga Seiki, KS301-20MMS) to move along the z axis. To maintain the temperature of the sample and water couplant at 20 °C, a pump (Sanso Electric, PMD-121B7B1) circulates water cooled by a low-temperature bath (Thomas, T-22) in a cavity around the glass tube.

The block diagram of the electronics employed in the system is illustrated in Fig. 1(c). The function generator (Tektronix, AFG3101), triggered with pulses of 20-kHz repetition rate which are supplied by a pulse generator (Iwatsu, RG-230), generates 7-cycle tone burst signals of frequency f, 15-cycle tone burst signals of $2f$, or dual-frequency tone burst consisting of 7-cycle f and 14-cycle $2f$ signals of the same amplitude. These burst signals drive the transducer via a 12-dB power amplifier (Thamway, T142-4749B excluding the pre-amplifier stage) whose output voltage amplitude is denoted by V_d. A beam with a focal length of $D=b/(1-c_W/c_0)=10.65$ mm is radiated into distilled water (density $\rho_W=1000$ kg/m^3 and sound speed $c_W=1483$ m/s at 20 °C) in a glass tube of 22-mm inner diameter and 30-mm length, which has been bonded with a silicone rubber to the lens. The $1/e$ width of the beam on the focal plane is estimated in water to be $4D\xi^{1/2}/k_W=0.22$ mm at f and $2D\xi_2^{1/2}/k_W=0.08$ mm at $2f$, where $k_W=2\pi f/c_W$. For simplicity, the present transducer can be modeled as a focusing Gaussian source with a curvature radius D, as illustrated in Fig. 2, with a bold curve that coincides with the lens surface at the center. The origin of cylindrical coordinates employed in the following analysis is set at the center of the lens surface.

The burst wave radiated from the transducer takes a roundtrip across the sample and travels back to the transducer. The tungsten rod works as a rigid reflector because of its characteristic impedance of approximately 70 times that of water. The length of the focal region where the strong nonlinear effect takes place is defined as $10\xi D^2/k_W=2.4$ mm, for example, in water.[20] When the reflector is set at the focal plane, a 2-mm-long roundtrip path in the sample is obtained.

In this focusing system, the sound reflected at the focal plane can be received with the highest sensitivity because the whole beam is captured almost in phase. The amplitude of the received signal is then accurately proportional to the magnitude of the reflected wave. Hence, to evaluate the magnitude of a reflected sound from any boundary, the z stage position is adjusted to maximize the amplitude of the received signal so that the boundary locates on the focal plane. Unnecessary waves reflected at other boundaries are temporally separated. The received waveform is observed with a digital oscilloscope (Yokogawa, DL-1740), and the waveform data obtained with a sampling rate of 2 GS/s at the 32-time averaging mode are sent to a personal computer (NEC, MY30V/C-F). The parameters such as amplitudes, time intervals, and second harmonic components of burst waves are derived using LabVIEW programs. The waveform of the burst reflected from the tungsten rod immersed alone in water is first recorded as a reference waveform. Then the maximum value in the cross-correlation function with the reference waveform is taken to be the amplitude of the measured signal. The time interval (whose resolution is 0.5 ns due to the 2-GS/s sampling rate) between two burst waves is also obtained from the time interval between two maximum values of the cross-correlation function with the reference waveform. For estimating the magnitude of second harmonic sound, a Fourier transform is executed for the discrete data of the entire received burst waveform including its transient portions. The phase delay of the second harmonic relative to the fundamental that is introduced later is also obtained with the fast Fourier transform (FFT).

The second harmonic caused by the nonlinear propagation must be detected for the B/A measurement. For this purpose, a second harmonic component directly radiated from the source should preferably be negligibly small. The present 7-cycle burst signal at frequency f driving the transducer is measured to have a $2f$ component of −48 dB relative to the f component. Even if the waveform of each cycle of the driving voltage is purely sinusoidal, the second harmonic exists and is unavoidable. The presence of a −48-dB second harmonic is inevitable and it would be difficult to markedly reduce its level. Since larger harmonic components with relative levels of over −30 dB can be practically obtained in the B/A measurements of the sample to be measured, such a level of the inherent second harmonic is acceptable.

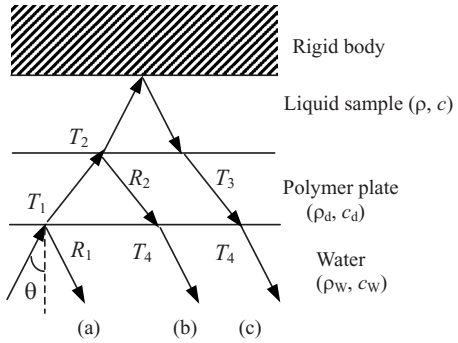


FIG. 3. Reflection and transmission of obliquely incident plane-wave at boundaries of acoustic window.

III. REFLECTION AND TRANSMISSION AT ACOUSTIC WINDOW

One of the features of the present method is to use a solid polymer plate as the acoustic window so as to keep the sample thickness constant. A polymer plate has been used for the acoustic window in an acoustic impedance microscope that was previously developed for observing the microscopic distribution of the characteristic impedance in biological soft tissues.[37] In the present study, the characteristic impedance is measured by the same manner. The calibration method in the acoustic impedance microscope has also been presented based on the approximation using geometrical optics.[38] In the present work, the measurement is realized on the basis of a more rigorous solution.

Provided that the plate is so thick that the roundtrip time within it is longer than the duration of a radiated burst wave, the burst waves returned from the front and rear surfaces can be temporally separated from each other. When such a condition is satisfied, as far as the first arriving burst wave is concerned, the sound wave can be simply treated as a one-way traveling wave without overlap with multiple reflection waves. When a 0.8-mm-thick plate of 2333-m/s sound speed is employed, the roundtrip time is 0.7 μs. If the transient portions are included, the seven-cycle burst with 0.4-μs duration becomes a burst wave of about 0.6-μs duration, so the present system satisfies the no overlap condition.

On the other hand, since a focused beam contains obliquely incident components even at its normal incidence onto the plate, the transmitting property of longitudinal waves may be significantly affected by the transformation to transverse waves at the boundaries.[39] In this section, the linear property of the plate is examined in terms of the reflection and transmission.

Figure 3 shows various propagation paths of the obliquely incident plane-wave with an incidence angle θ. The plane-wave propagates in the layers consisting of a combination of the liquid sample and acoustic window set in front of the reflector. Case (a) represents the path of the wave reflected at the front surface of the acoustic window. The amplitude of the returned sound is proportional to the reflection coefficient R_1 of the pressure amplitude at the water-plate interface. Case (b) represents the path of the wave reflected at the rear surface of the acoustic window. The amplitude of the returning sound is proportional to the product $T_1R_2T_4$. Here, T_1, R_2, and T_4 are the transmission coefficient at the water-plate boundary, the reflection coefficient at the plate-sample boundary, and the transmission coefficient at the plate-water boundary, respectively. Case (c) similarly represents the path of the wave reflected at the tungsten rod. The amplitude of the returning sound is proportional to the product $T_1T_2T_3T_4$, where T_2 and T_3 are the transmission coefficients at the plate-sample and sample-plate boundaries, respectively.

The components of stress and displacement normal to the surface must be continuous at the solid-liquid interface. In addition, the shear stress parallel to the surface must vanish in the solid plate at the interface. Furthermore, wave numbers along the interface must be continuous. By applying these boundary conditions for the plane-wave with an incidence angle θ smaller than the critical angle $\theta_c = \sin^{-1}(c_W/c_d)$, the following coefficients are derived:

$$T_1 = \frac{2\rho_d/\rho_W}{\frac{P_d}{\cos\theta} + \frac{\rho_d}{\rho_W}\left(\frac{c_t}{c_W}\right)^2 Q_1 + \left[\frac{2}{\cos\theta} + \frac{4\rho_d}{\rho_W}\left(\frac{c_t}{c_W}\right)^2 P_t\right]\frac{P_d}{Q_t}\sin^2\theta}, \tag{1}$$

$$T_2 = \frac{Q_t/P_S}{\frac{\sin^2\theta}{P_S} + \frac{2\rho_d}{\rho}\left(\frac{c_t}{c_W}\right)^2 P_t \sin^2\theta + \frac{Q_t}{2P_d}\left[\frac{P_d}{P_S} + \frac{\rho_d}{\rho}\left(\frac{c_t}{c_W}\right)^2 Q_t\right]}, \tag{2}$$

$$T_3 = \frac{2\rho_d/\rho}{\frac{P_d}{P_S} + \frac{\rho_d}{\rho}\left(\frac{c_t}{c_W}\right)^2 Q_t + \left[\frac{2}{\cos\theta} + \frac{4\rho_d}{\rho}\left(\frac{c_t}{c_W}\right)^2 P_t\right]\frac{P_d}{Q_t}\sin^2\theta}, \tag{3}$$

$$T_4 = \frac{Q_t/\cos\theta}{\frac{\sin^2\theta}{\cos\theta} + \frac{2\rho_d}{\rho_W}\left(\frac{c_t}{c_W}\right)^2 P_t \sin^2\theta + \frac{Q_t}{2P_d}\left[\frac{P_d}{\cos\theta} + \frac{\rho_d}{\rho_W}\left(\frac{c_t}{c_W}\right)^2 Q_t\right]}, \tag{4}$$

$$R_1 = 1 - \frac{P_d}{\left(\frac{c_t}{c_W}\right)^2 Q_t \cos\theta} \times \frac{2}{\frac{P_d}{\cos\theta} + \frac{\rho_d}{\rho_W}\left(\frac{c_t}{c_W}\right)^2 Q_t + \left[\frac{2}{\cos\theta} + \frac{4\rho_d}{\rho_W}\left(\frac{c_t}{c_W}\right)^2 P_t\right]\frac{P_d \sin^2\theta}{Q_t}}, \tag{5}$$

$$R_2 = \frac{\frac{P_d}{P_S} - \frac{\rho_d}{\rho}\left(\frac{c_t}{c_W}\right)^2 Q_t + \frac{2P_d}{Q_t}\left[\frac{1}{P_S} + \frac{2\rho_d}{\rho}\left(\frac{c_t}{c_W}\right)^2 P_t\right]\sin^2\theta}{\frac{P_d}{P_S} + \frac{\rho_d}{\rho}\left(\frac{c_t}{c_W}\right)^2 Q_t + \frac{2P_d}{Q_t}\left[\frac{1}{P_S} + \frac{2\rho_d}{\rho}\left(\frac{c_t}{c_W}\right)^2 P_t\right]\sin^2\theta}, \tag{6}$$

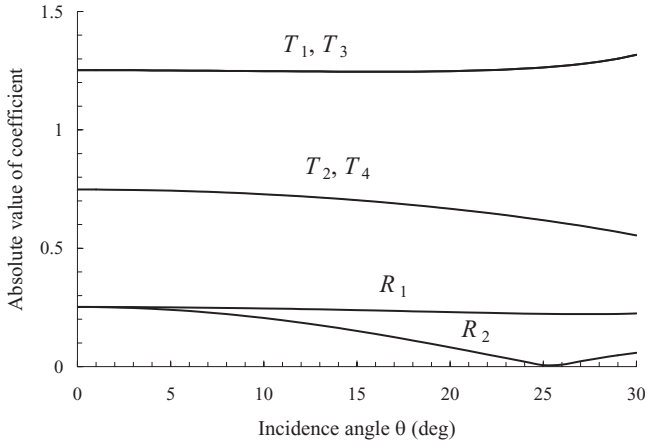


FIG. 4. Dependence of reflection and transmission coefficients on incidence angle θ in the case of $\rho=\rho_W$ and $c=c_W$.

where $P_d=\sqrt{(c_W/c_d)^2-\sin^2\theta}$, $P_t=\sqrt{(c_W/c_t)^2-\sin^2\theta}$, $P_S=\sqrt{(c_W/c)^2-\sin^2\theta}$, and $Q_t=(c_W/c_t)^2-2\sin^2\theta$. It is noted that the values of Eqs. (1)–(6) never depend on the frequency. Assuming an acoustically waterlike sample such as biological soft tissues, ρ and c of the sample in Fig. 3 are considered equal to ρ_W and c_W. Then the θ dependences of the absolute values of T_1, T_2, T_3, T_4, R_1, and R_2 are exemplified, as shown in Fig. 4.

It will be helpful to utilize the Hankel transform of the on-source pressure (with respect to the r coordinate) expressed as $p_0\exp[-\xi r^2-jk_Wr^2/(2D)]$ and $p_{02}\exp(-\xi_2r^2-jk_Wr^2/D)$, which spectrally decomposes the resulting focused beam to cylindrical modes, each of which is the ensemble of plane-waves with the identical offset angle that is defined as the angle of the propagation against the z axis. The offset angle is equivalent to the incidence angle θ, provided the plate is set parallel to the source. The distribution of this source on θ is obtained, as shown in Fig. 5.[40] Almost all the components are involved in θ smaller than θ_c of $\sin^{-1}(c_W/c_d)=39.5°$. Therefore, the coefficients formulated by Eqs. (1)–(6) are available to the entire beam. While the components of $\theta<20°$ are dominant in the radiation at $f=18.6$ MHz, those of $20°<\theta<30°$ are still dominant in the radiation at $2f=37.2$ MHz. Although T_1, T_2, T_3, T_4, and R_1 are almost constant for $\theta<20°$, R_2 varies with θ in Fig. 4.

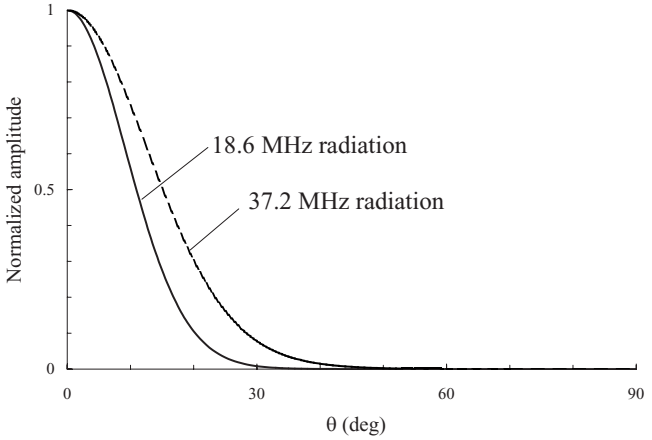


FIG. 5. Distribution of experimental focused Gaussian beam on offset angle.

Because of a higher sound speed in a polystyrene plate compared with that in water, the incidence angle at the plate-sample interface is increased owing to the refraction. Then the reflection at the rear surface of the plate undergoes considerable transformation to transverse waves. The sign of R_2 in Fig. 4 reverses at θ near 25°. This sign reversal occurs at a smaller incidence angle for a sample with a higher sound speed, where the reduction in R_2 at $\theta<20°$ is more significant.

IV. MEASUREMENT OF LINEAR PROPERTIES OF THE SAMPLE

Since the magnitude of the second harmonic generation depends not only on B/A in the medium but also on the linear constants such as density, sound speed, and attenuation coefficient, the measurement of linear properties is necessary in the finite amplitude method. Furthermore, for HF band ultrasound as employed in the present measurement, velocity dispersion may occur. Focused ultrasound has the well-known property that the second harmonic component generated in the focal region is effectively delayed by $\pi/2$ rad relative to the fundamental.[41] The velocity dispersion may compensate this phase delay and actually enhance the second harmonic accumulation.[42] To correctly estimate B/A value, the magnitude of the velocity dispersion in the sample must be measured.

A. Measurement of linear features

Since all the measurements are conducted by comparing with distilled water having well-known properties, distilled water filling the sample container is measured a priori. The time interval τ_W between the burst waves reflected from the rear surface of the acoustic window and the tungsten rod is measured first as well as the amplitude P_{W1} of the wave reflected from the rod for the sound radiation at frequency $f=18.6$ MHz. The sample thickness is then determined as $L=c_W\tau_W/2$. After the frequency is changed to $2f$, the amplitude P_{2W} of the reflected wave from the rod is similarly measured. When the LN transducer is driven with a dual-frequency voltage of f and $2f$, the transmitted sound wave is described by a Fourier series of $\Sigma_{n-1}^2(a_n\cos 2n\pi ft+b_n\sin 2n\pi ft)=\Sigma_{n=1}^2\sqrt{a_n^2+b_n^2}\sin(2n\pi ft+\phi_n)$, where $\phi_n=\mathrm{sgn}(a_n)\cos^{-1}(b_n/\sqrt{a_n^2+b_n^2})$. The phase delay Φ of the zero-crossing portion of the $2f$ component, compared to that of the f component, is defined as $\Phi=2\phi_1-\phi_2$. The phase delay $\Phi(=\Phi_W)$ is measured by executing a FFT for the roundtrip acoustic wave that is reflected at the rod. Under the condition that the spacer has been removed, the amplitude P_{RB} of the reflected wave from the rear surface of the acoustic window pressed tightly on a wet end of the rod is also measured at f.

Then the distilled water in the container will be replaced with the sample. The amplitude P_{SB} of the reflected wave from the rear surface of the acoustic window and the amplitude P_{S1} of the reflected wave from the rod are measured, as well as the time difference τ_S between the reflected waves from the rear surface of the acoustic window and the rod at the same frequency f. The amplitude P_{S2} of the reflected wave from the tungsten rod is also measured at $2f$. In addi-

tion, the phase delay $\Phi(=\Phi_S)$ is also measured for the sound wave that is reflected at the rod using dual-frequency signals.

All these measurements are conducted under an identical (low) excitation level of the transducer. This level is small enough to neglect the energy conversion to higher harmonics yet large enough to obtain a high signal to noise ratio for the received signal of acoustic waves transmitted through considerably dissipative samples. From these values, the linear properties are determined through the following procedures.

B. Determination of sound speed *c*

The sound speed c is determined with the equation

$$c = \frac{2L}{\tau_S}. \tag{7}$$

C. Determination of density *ρ*

The ratio P_{SB}/P_{RB} represents the pressure reflection coefficient R at the boundary between the acoustic window and the sample. In the case of a normal incident plane wave, as is well known,

$$R = \left| \frac{Z - Z_d}{Z + Z_d} \right|, \tag{8}$$

where $Z=\rho c$ and $Z_d=\rho_d c_d$. However, since the focused beam contains obliquely incident components, the above equation may be invalid as described in Sec. III. An appropriate R value can be theoretically derived as follows.

Once a boundary is located on the focal plane, the beam reflected from the boundary returns to the receiver at exactly the same offset angle distribution as that of the sound reflected from a perfect reflector immersed alone in water at the focal distance $z=D$.[43] This is true even if an acoustic window is additionally inserted. The pressure distribution on the receiver has the same profile as that at $z=2D$ in a free field. By neglecting the attenuation in water, the Hankel transform of the pressure in the free field caused by the focusing Gaussian source is given at $z=2D$ as[43]

$$\tilde{p}(s) = \frac{p_0}{2\xi + jk_W/D} \exp\left[-j\frac{Ds^2}{k_W} - \frac{s^2}{2(2\xi + jk_W/D)} \right], \tag{9}$$

where $s=k_W \sin\theta$. In the parabolic approximation of the wave equation, the attenuation is considered only in the propagation along the z axis. Hence, the attenuation of oblique components is quite the same as that of normal components and is therefore omitted in the discussion on the effect of oblique components. By multiplying the value obtained using Eq. (9) with $T_1R_2T_4$, that is, the function of $\theta = \sin^{-1}(s/k_W)$, the θ dependence of the wave returning to the receiver in case (b) of Fig. 3 can be derived. The radial distribution of the sound pressure returning to the plane $z=0$ of Fig. 2 is given as the inverse Hankel transform of the resulting spectral component $T_1R_2T_4\tilde{p}(s)$:

$$p(r) = \int_0^\infty T_1R_2T_4\tilde{p}(s)J_0(rs)sds. \tag{10}$$

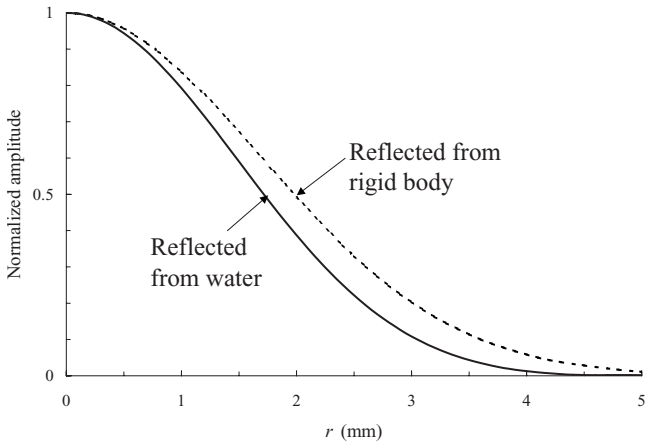


FIG. 6. Calculated radial distribution of pressure on receiver surface.

Assuming a sample with the same ρ and c as those of water, the calculated results for the radial distribution of the reflected pressure amplitude (to be observed at $z=0$) are shown in Fig. 6, in comparison with that of the sound from a rigid reflector of $R_2=1$. The ordinate is normalized by the value at the center in each case. As the R_2 in Fig. 4 decreases at large θ, the beam reflected from the sample becomes slightly narrow. The amplitude of the received signal should be proportional to the surface integral of the pressure multiplied by the receiving sensitivity of Gaussian distribution $\exp(-\xi r^2)$ on the receiver,

$$V = \int_0^a |p(r)| \exp(-\xi r^2) r dr. \tag{11}$$

Since the reflected wave from the focal plane is almost in phase on the curved receiver surface, the integration of the absolute value $|p(r)|$ on the plane $z=0$ is carried out instead of the integration of $p(r)$ on the concave surface for simplicity. The effective reflection coefficient R is thereby theoretically expressed as

$$R = \frac{\int_0^a \left| \int_0^\infty T_1R_2T_4\tilde{p}(s)J_0(rs)sds \right| \exp(-\xi r^2) r dr}{\int_0^a \left| \int_0^\infty T_1T_4\tilde{p}(s)J_0(rs)sds \right| \exp(-\xi r^2) r dr}. \tag{12}$$

The characteristic impedance should be measured using this coefficient. The ratio $R=P_{SB}/P_{RB}$ measured for water, ethylene glycol, and methanol, which are selected from liquid samples with known ρ and c, is compared with the values obtained using Eqs. (8) and (12) in Table I. The experimentally obtained R close to the calculated R using Eq. (12) validates the effective reflection coefficient

$$\frac{P_{SB}}{P_{RB}} = S \left| \frac{Z - Z_d}{Z + Z_d} \right|, \tag{13}$$

where

TABLE I. Comparison of reflection coefficients.

	Water	Ethylene glycol	Methanol
R of Eq. (8)	0.247	0.145	0.476
R of Eq. (12)	0.207	0.102	0.426
P_{SB}/P_{RB}	0.212	0.106	0.438

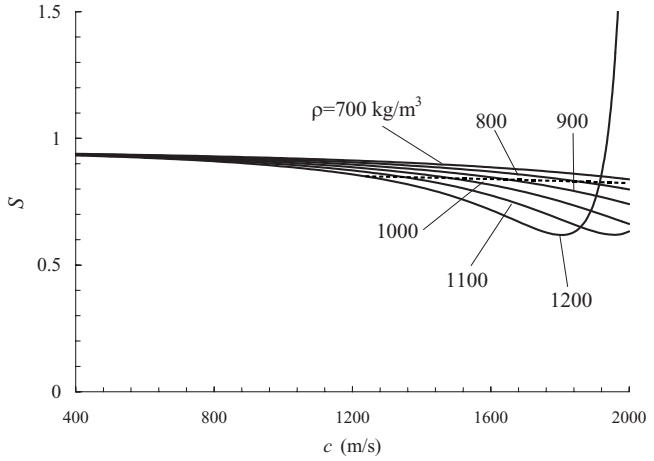


FIG. 7. Correction factor S for reflection coefficient at the interface between polystyrene plate and liquid of various ρ and c.

$$S = \left| \frac{Z + Z_d}{Z - Z_d} \right| \frac{\int_0^a \left| \int_0^\infty T_1 R_2 T_4 \tilde{p}(s) J_0(rs) s ds \right| \exp(-\xi r^2) r dr}{\int_0^a \left| \int_0^\infty T_1 T_4 \tilde{p}(s) J_0(rs) s ds \right| \exp(-\xi r^2) r dr}. \tag{14}$$

Using Eq. (14), the correction factor S for the reflection coefficient R is shown in Fig. 7 in the range of $700 \leq \rho \leq 1200$ kg/m^3 and $400 \leq c \leq 2000$ m/s. As the density and sound speed increase, S decreases. However, as ρ approaches 1200 kg/m^3, S rapidly increases when the sound speed increases to as high as 1900 m/s because the reflection of oblique components never vanishes even though the condition of impedance matching is satisfied with c as high as 2040 m/s. The dotted line in Fig. 7 indicates the value of S assuming various combinations of ρ and c under the restriction of $\rho c = Z_W$, where $Z_W = 1.483 \times 10^6$ kg m^{-2} s^{-1} is the characteristic impedance in water. Even for the sample with the identical characteristic impedance, S varies from 0.839 to 0.823 in the range of $c = 1240$–2000 m/s.

S for $\rho = 1000$ kg/m^3 in Fig. 7 is expressed using the polynomials of $\log_{10} c$, where the range of c is split at 794 m/s to reduce the number of terms in the polynomial. Then the curve of $\log_{10} c$ vs S for any given ρ can be regarded as a horizontally shifted curve for $\rho = 1000$ kg/m^3. Through these procedures, the value of S can be expressed as

$$\left.\begin{aligned} S &= A_0 - A_1 x - A_2 x^2, && c \leq 794 \text{ m/s} \\ &= B_0 - B_1 y - B_2 y^2 - B_3 y^3 - B_4 y^4 - B_5 y^5, && c > 794 \text{ m/s}, \end{aligned}\right\} \tag{15}$$

where $x = \log_{10} c - 3 - 4.377 \times 10^{-2} q - 2.4718 \times 10^{-3} q^2$, $y = \log_{10} c - 3 - 3.7138 \times 10^{-2} q - 1.425 \times 10^{-3} q^2$, and $q = (1000 - \rho)/100$. The coefficients are given as $A_0 = 0.9087$, $A_1 = 0.114$, $A_2 = 0.1181$, $B_0 = 0.9047$, $B_1 = 0.1842$, $B_2 = 0.5079$, $B_3 = 1.2389$, $B_4 = 6.884$, and $B_5 = 21.126$. The value of S calculated with Eq. (15) is shown in Fig. 8. While being poor for characteristic impedances close to $\rho_d c_d$, this approximation well holds for S in the other wide ranges. For presumed biological tissues with characteristic impedances smaller than that of the polystyrene plate, the present expression is useful. By substituting Eq. (15) into Eq. (13), the unknown single quantity ρ is obtained by the Newton method from the experimental P_{SB}/P_{RB}.

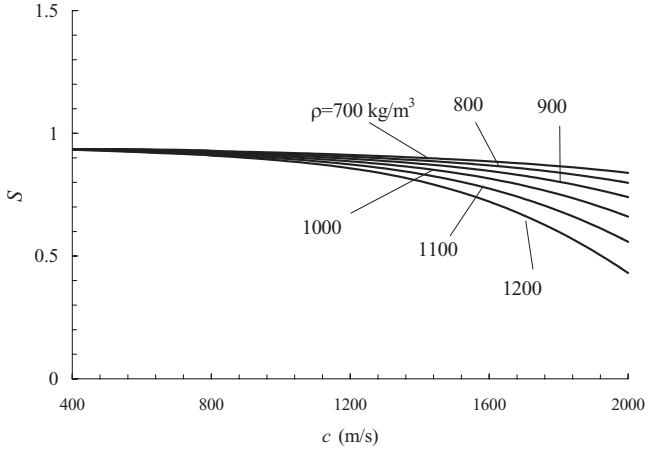


FIG. 8. Values of S approximated with Eq. (15).

D. Determination of attenuation coefficient α

The attenuation coefficients α_1 and α_2 are estimated from the insertion loss due to the sample at the respective frequencies f and $2f$. The insertion loss in the case of distilled water filling the sample container is used as the reference. Since the focal distance changes with the sound speed c in the sample, the propagation distance in the attenuation measurement changes with c. In addition, the magnitude of the impedance mismatching against the polystyrene plate varies with ρc of the sample. We examine the possible effects due to these variables by considering the attenuation coefficients which are given as

$$\alpha_n = n^2 \alpha_W \frac{c}{c_W} + \frac{1}{2L} \ln\left(Q \frac{P_{Wn}}{P_{Sn}} \right) \quad (n = 1,2), \tag{16}$$

where $Q = Z(Z_d + Z_W)^2/[Z_W(Z_d + Z)^2]$ and α_W is the attenuation coefficient at 18.6 MHz in water, namely, 8.65 Np/m.

With the same logic as in the procedure to derive Eq. (14), the correction factor S_T for the transmission coefficient for the roundtrip through the sample can be estimated using the following equation:

$$S_T = \frac{\int_0^a \left| \int_0^\infty T_1 T_2 T_3 T_4 \tilde{p}(s) J_0(rs) s ds \right| \exp(-\xi r^2) r dr}{T_0 \int_0^a \left| \int_0^\infty \tilde{p}(s) J_0(rs) s ds \right| \exp(-\xi r^2) r dr}, \tag{17}$$

where $T_0 = 16 Z_d^2 Z Z_W/[(Z_W + Z_d)(Z + Z_d)]^2$. The calculation at $2f$ employs ξ_2 instead of ξ. When the sample container is filled with water, Eq. (17) determines S_T to be 0.929 and 0.944 at f and $2f$, respectively. When filled with a medium of $\rho = 1113$ kg/m^3 and $c = 1666$ m/s, for instance, S_T becomes 0.930 and 0.945 at f and $2f$, respectively. Hence, the difference in P_{Wn}/P_{Sn} shown in Eq. (16) is only 0.1% as a result of the over 10% differences in both ρ and c from those of water. Therefore, S_T is negligible for determining α_n.

E. Estimation of magnitude of velocity dispersion

Assuming that water is free from velocity dispersion, the variance $\Delta\Phi = \Phi_W - \Phi_S$ of the phase advance is caused by the velocity dispersion in the sample. The quantity Δk

$=\Delta\Phi/(2L)$ to indicate the relative phase advance arising at every 1 m of propagation in the sample is employed as the parameter of velocity dispersion.[35]

V. MEASUREMENT OF NONLINEARITY PARAMETER *B/A*

With the sample container filled with distilled water, the second harmonic component P_{2WN} in the wave reflected from the rod is measured under the sound radiation at frequency f. The second harmonic amplitude P_{2SN} is similarly measured when the container has been filled with the sample. In order to model the measurement, a successive approximation solution of the Khokhlov–Zabolotskaya–Kuznetsov (KZK) equation is derived for continuous waves in each region of water, the acoustic window and the sample, where the effect of velocity dispersion in the sample has been taken into account.[26] By connecting these solutions at all the boundaries, the second harmonic to be received by the transducer has been derived.[42] The phase shift Δk_d due to a slight velocity dispersion existing in the polystyrene plate[35] is also taken into account. By omitting the factors that are common in both expressions, P'_{2WN} and P'_{2SN} to which P_{2WN} and P_{2SN} must be proportional, the following expressions are obtained:

$$
\begin{aligned}
P'_{2WN} = \exp(-8\alpha_W z_W - 8\alpha_W L - 2\alpha_{d2} d)\Bigg\{ & \beta_W T_{AW} \exp(-j2\Delta k_d d) \int_0^{z_W} \frac{\exp(2\alpha_W z')}{h(z')} dz' \\
& + \frac{\beta_d T_{BW} \rho_W c_W^3 \exp(2\alpha_W z_W - j2\Delta k_d d)}{\rho_d c_d^3} \int_0^d \frac{\exp[(\alpha_{d2} - 2\alpha_{d1} + j\Delta k_d) z']}{h(z_W + c_d z'/c_W)} dz' \\
& + \beta_W T_{CW} \exp[2\alpha_W z_W + (\alpha_{d2} - 2\alpha_{d1} - j\Delta k_d) d] \int_0^{2L} \frac{\exp(2\alpha_W z')}{h(z_W + c_d d/c_W + z')} dz' \\
& + \frac{\beta_d T_{DW} \rho_W c_W^3 \exp[2\alpha_W z_W + (\alpha_{d2} - 2\alpha_{d1} - j\Delta k_d) d + 4\alpha_W L]}{\rho_d c_d^3} \int_0^d \frac{\exp[(\alpha_{d2} - 2\alpha_{d1} + j\Delta k_d) z']}{h(z_W + c_d d/c_W + 2L + c_d z'/c_W)} dz' \\
& + \beta_W T_{EW} \exp[2\alpha_W z_W + 2(\alpha_{d2} - 2\alpha_{d1}) d + 4\alpha_W L] \int_0^{z_W} \frac{\exp(2\alpha_W z')}{h(z_W + 2c_d d/c_W + 2L + z')} dz' \Bigg\},
\end{aligned}
\tag{18}
$$

$$
\begin{aligned}
P'_{2SN} = \exp(-8\alpha_W z_S - 2\alpha_2 L - 2\alpha_{d2} d)\Bigg\{ & \beta_W T_{AS} \exp(-j2\Delta k_d d - j2\Delta k L) \int_0^{z_S} \frac{\exp(2\alpha_W z')}{h(z')} dz' \\
& + \frac{\beta_d T_{BS} \rho_W c_W^3 \exp(2\alpha_W z_S - j2\Delta k_d d - j2\Delta k L)}{\rho_d c_d^3} \int_0^d \frac{\exp[(\alpha_{d2} - 2\alpha_{d1} + j\Delta k_d) z']}{h(z_S + c_d z'/c_W)} dz' \\
& + \frac{\left(2 + \frac{B}{A}\right) T_{CS} \rho_W c_W^3 \exp[2\alpha_W z_S + (\alpha_{d2} - 2\alpha_{d1} - j\Delta k_d) d - j2\Delta k L]}{2\rho c^3} \int_0^{2L} \frac{\exp[(\alpha_2 - 2\alpha_1 + j\Delta k) z']}{h(z_S + c_d d/c_W + c z'/c_W)} dz' \\
& + \frac{\beta_d T_{DS} \rho_W c_W^3 \exp[2\alpha_W z_S + (\alpha_{d2} - 2\alpha_{d1} - j\Delta k_d) d + 2(\alpha_2 - 2\alpha_1) L]}{\rho_d c_d^3} \int_0^d \frac{\exp[(\alpha_{d2} - 2\alpha_{d1} + j\Delta k_d) z']}{h(z_S + c_d d/c_W + 2cL/c_W + c_d z'/c_W)} dz' \\
& + \beta_W T_{ES} \exp[2\alpha_W z_S + 2(\alpha_{d2} - 2\alpha_{d1}) d + 2(\alpha_2 - 2\alpha_1) L] \int_0^{z_S} \frac{\exp(2\alpha_W z')}{h(z_S + 2c_d d/c_W + 2cL/c_W + z')} dz' \Bigg\},
\end{aligned}
\tag{19}
$$

where $h(z)=2\xi z - jk_W(1-z/D)$, $T_{AW}=T_{Id}T_{IW}T_{OW}T_{Od}$, $T_{BW}=T_{Id}T_{AW}$, $T_{CW}=T_{IW}T_{BW}$, $T_{DW}=T_{OW}T_{CW}$, $T_{EW}=T_{Od}T_{DW}$, $T_{AS}=T_{Id}T_{IS}T_{OS}T_{Od}$, $T_{BS}=T_{Id}T_{AS}$, $T_{CS}=T_{IS}T_{BS}$, $T_{DS}=T_{OS}T_{CS}$, $T_{ES}=T_{Od}T_{DS}$, $T_{Id}=2Z_d/(Z_d+Z_W)$, $T_{IW}=2Z_W/(Z_d+Z_W)$, $T_{OW}=T_{Id}$, $T_{Od}=T_{IW}$, $T_{IS}=2Z/(Z+Z_d)$, $T_{OS}=2Z_d/(Z+Z_d)$, $z_W=D-L-c_d d/c_W$, and $z_S=D-(c_d d+cL)/c_W$. The second harmonic generation within the lens has been neglected because of its smallness.[33] The nonlinearity coefficient in water is taken to be $\beta_W=3.5$. The attenuation coefficients in the polystyrene plate are given as $\alpha_{d1}=121$ Np/m at 18.6 MHz and $\alpha_{d2}=310$ Np/m at 37.2 MHz.[35] Furthermore, the nonlinearity coefficient β_d and the velocity dispersion Δk_d in the polystyrene are 6.7 and 39.0 rad/m, respectively.[35]

It has theoretically been confirmed that second harmonic component in burst waves employed in the measurement varies with B/A of the sample at the same rate as the second harmonic component in continuous waves.[25] The B/A of the sample is determined by adjusting the B/A value of the third

TABLE II. Measurement results for two liquid samples.

Sample	Ethylene glycol	Methanol
τ_s (μs)	1.170	1.766
P_{SB}/P_{RB}	0.1063	0.4375
P_{W1}/P_{S1}	1.089	1.226
P_{W2}/P_{S2}	1.404	1.260
ρ (kg/m^3)	1101 (1183)	779 (866)
c (m/s)	1670	1106
α_1 (Np/m)	75 (79)	8.8 (33.8)
α_2 (Np/m)	234 (239)	42.2 (67.1)
B/A	9.3	9.3
ρ (literature)	1113	791
c (literature)	1666	1121
α_1 (literature)	65	11.7
α_2 (literature)	242	46.6
B/A (literature)	9.9	9.6

term in the brace of Eq. (19) so that the ratio $|P'_{2SN}/P'_{2WN}|$ fits the observed amplitude ratio P_{2SN}/P_{2WN}, yielding the sought after B/A value, adaptively.

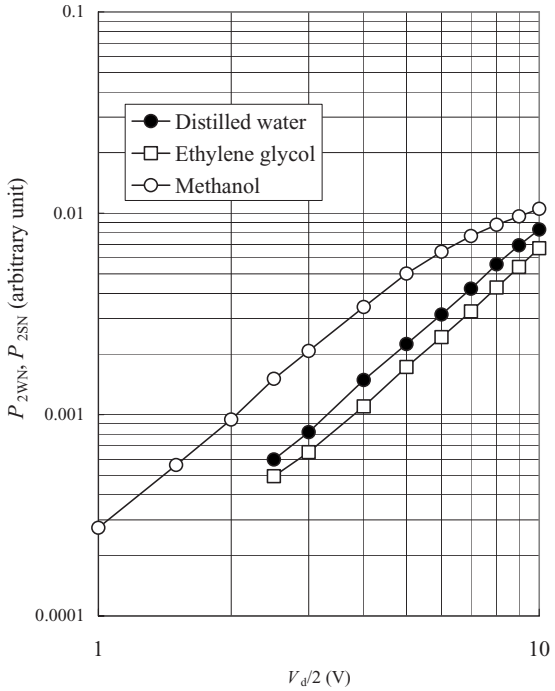


FIG. 9. Observed second harmonic amplitudes P_{2SN} and P_{2WN}.

VI. EXAMPLE OF MEASUREMENT RESULTS

The experimental measurement is carried out for samples whose thickness has been measured by the method described in Sec. IV to be $L=c_W\tau_W/2=0.977$ mm. This value of L is the average of ten measurements that are repeated starting with the pouring of water in the sample container. The standard deviation of the L measurement is estimated to be 1.1 μm. The thickness of the polystyrene plate is measured using a micrometer to be 0.81 mm. Nondispersive liquids with known acoustic properties, ethylene glycol and methanol, which are nonerosive to the polystyrene plate, are measured first.

The determined values for ρ, c, α, and B/A using Eqs. (7), (13), (16), (18), and (19) are shown in Table II together with τ_S, P_{SB}/P_{RB}, P_{W1}/P_{S1}, and P_{W2}/P_{S2} from ten time repetitive measurements after setting the sample. The literature values for ρ, c, and α[24,43] are also shown for comparison. The values of ρ, c, $\exp(-2\alpha_1 L)$, and $\exp(-2\alpha_2 L)$ agree with the literature values within $\pm 2\%$. When the inclusion of oblique incident components is neglected by putting $S=1$ in Eq. (13), the parenthesized values of ρ are obtained. The values of α_1 and α_2 derived from Eq. (16) using these values of ρ are also shown in the parentheses. The agreement with literature values is poor for the values derived under the assumption of $S=1$. Figure 9 shows the average values of $|P_{2SN}|$ and $|P_{2WN}|$ for various driving voltages V_d of the LN transducer. Since the second harmonic components generated in the pre- and post-focal regions cancel each other as previously analyzed,[20,41,44] the second harmonic component significantly varies after replacing the medium in the focal region with a different one. While the ethylene glycol sample shows a weak nonlinear distortion, that is, the second harmonic amplitude $|P_{2SN}|$ increases in proportion to V_d^2, the methanol sample shows sufficiently increasing $|P_{2SN}|$ in a small amplitude range. The trend of saturation is seen at the middle of the measurement range because of significant nonlinear distortions due to low values of α, ρ, and c of methanol. Since the determination of B/A described in Sec. V is based on the analytical result for a small amplitude region where the second harmonic amplitude is proportional to V_d^2, P_{2SN}/P_{2WN} is sampled in such a region. The obtained B/A agrees with the literature value within $\pm 6\%$. It is therefore confirmed that the typical accuracy of the finite amplitude method including an uncertainty of $\pm 10\%$ is obtained. Since the present method is intended for eventually attaining the nonuniform distribution of B/A through C-mode scanning, the reproducibility of the B/A measurement is more important than the accuracy of the absolute value. The ten repetitive measurements of B/A after refilling the sample container yielded relatively small standard deviations of approximately 1%.

The B/A measurement for biological samples is examined. The results are tabulated in Table III where the measured values for Δk are also shown. Similar samples have been measured only in low megahertz ranges.[2,14,45] The measured values for B/A agree well with literature results except for an egg yolk sample where a considerable measurement error is made owing to the attenuation being as high as $\exp(-2\alpha_2 L)=0.1$. When the dispersion in the sample is neglected, B/A values greater than the literature values are obtained as shown in square brackets. A 57% increase in the apparent B/A is found in the case of Δk of 205 rad/m. The increasing rate of $\Delta c/c$ in the sound speed at the frequency $2f$ is approximately equal to $\Delta k/(2k)$. Assuming that $c=1483$ m/s, therefore, $\Delta k=205$ rad/m is equivalent to a slightly higher sound speed with an increase in $\Delta c=1.9$ m/s at $2f$. In spite of such a small velocity change, the effect on the B/A measurement using focused ultrasound is

TABLE III. Measurement results for biological samples.

Sample	Pig liver	Chicken liver	Egg white	Egg yolk	Boiled egg white
τ_S (μs)	1.232	1.251	1.290	1.303	1.291
P_{SB}/P_{RB}	0.1420	0.1620	0.1822	0.1882	0.1832
P_{W1}/P_{S1}	1.393	1.250	1.144	2.173	1.284
P_{W2}/P_{S2}	2.182	1.612	1.344	8.769	1.659
P_{2SN}/P_{2WN}	0.59	0.74	0.77	0.32	0.65
Δk (rad/m)	140	105	1.0	27	54
ρ (kg/m^3)	1074	1045	1012	1030	1031
c (m/s)	1586	1562	1515	1500	1514
α_1 (Np/m)	193.9	134.2	84.1	410.8	142.8
α_2 (Np/m)	451.4	291.7	193.6	1151	300.4
B/A	7.0 [9.4]	6.7 [8.4]	5.8 [5.8]	10.8 [14.0]	5.5 [6.4]
Literature B/A	6.7 ± 1.5[a]	5.7 ± 1.0[b]	5.2 ± 1.3[b]	8.3 ± 1.5[b]	5.2 ± 1.3[b]

[a]Reference 2.
[b]Reference 45.

significant. Such influence of velocity dispersion is a common disadvantage for measurements based on the finite amplitude method.

VII. CONCLUSIONS

By employing a polystyrene plate as an acoustic window capable of keeping the sample thickness constant, a finite amplitude method of measuring the acoustic nonlinearity parameter B/A for liquid samples with volumes as small as 0.1 ml was developed using a HF band focused beam. The transducer had been made so that a Gaussian beam was produced. The fundamental and second harmonic sound, generated in the medium, was detected. The sound speed c was determined from the time interval between the waves reflected from the front and rear interfaces of the sample. By measuring the characteristic impedance of the sample utilizing the pressure reflection coefficient at the boundary between the acoustic window and the sample, the density ρ was determined using the previously determined sound speed c. It was theoretically clarified that this reflection coefficient is considerably smaller than that of a normal incident wave owing to the transformation to transverse waves from longitudinal waves of oblique incident components inherent in focused sound. This effect was corrected for determining the characteristic impedance. On the other hand, the attenuation coefficient α was directly obtained from the insertion loss of the sound due to the sample. Furthermore, by adopting the solution for the KZK equation to the second harmonic component contained in the received sound, the B/A value to make the theoretical amplitude fit the measured one was obtained as the measurement result. The present measurement method was validated by the reasonable agreement of the measured values for such nondispersive samples as ethylene glycol and methanol with their literature values. The values of B/A in weakly dispersive samples including biological tissue were then measured. In this case, the magnitude of the velocity dispersion that is necessary for correcting B/A was also measured. Despite the raw B/A value being greater than the literature value, the corrected B/A value that accounts for the dispersion is in reasonable agreement with the literature value. This fact validates the present measurement method. For generating the C-mode image of the distribution of B/A by scanning the sound beam on a biological tissue, the present method will be automated to enhance the measurement speed, accuracy, and reproducibility.

ACKNOWLEDGMENTS

The author expresses his sincere thanks to Professor Emeritus Kiyoshi Nakamura of Tohoku University for supplying LN plates with an inverted domain layer and for his invaluable suggestions. He is thankful to Mr. Takanori Watanabe and Dr. Jung-Ho Kim for their assistance in the experiment and to Dr. Thomas G. Muir for helping in revising the manuscript. He would also like to thank three anonymous reviewers for their useful comments and advice. This work was supported by a Grant-in-Aid for Scientific Research (C) 19560429 from the Japan Society for the Promotion of Science.

[1]M. Sehgal, G. M. Brown, R. C. Bahn, and J. F. Greenleaf, "Measurement and use of acoustic nonlinearity and sound speed to estimate composition of excised liver," Ultrasound Med. Biol. **12**, 865–874 (1986).
[2]L. Bjørnø, "Characterization of biological media by means of their nonlinearity," Ultrasonics **24**, 254–259 (1986).
[3]K. Yoshizumi, T. Sato, and N. Ichida, "A physiochemical evaluation of the nonlinear parameter B/A for media predominantly composed of water," J. Acoust. Soc. Am. **82**, 302–305 (1987).
[4]P. Sarvazyan, T. V. Chalikian, and F. Dunn, "Acoustic nonlinearity parameter B/A of aqueous solutions of some amino acids and proteins," J. Acoust. Soc. Am. **88**, 1555–1561 (1990).
[5]E. C. Everbach, Z. Zhu, P. Jiang, B. T. Chu, and R. E. Apfel, "A corrected mixture rule for B/A," J. Acoust. Soc. Am. **89**, 446–447 (1991).
[6]P. Jiang, E. C. Everbach, and R. E. Apfel, "Application of mixture laws for predicting the components of tissue phantoms," Ultrasound Med. Biol. **17**, 829–838 (1991).
[7]X. Gong, "Nonlinear ultrasonic parameter in tissue characterization and imaging," Acoust. Imaging **20**, 453–458 (1993).
[8]W. K. Law, L. A. Frizzell, and F. Dunn, "Determination of the nonlinearity parameter B/A of biological media," Ultrasound Med. Biol. **11**, 307–318 (1985).
[9]R. Kompfner and R. A. Lemons, "Nonlinear acoustic microscopy," Appl. Phys. Lett. **28**, 295–297 (1976).
[10]A. B. Coppens, R. T. Beyer, M. B. Seiden, J. Donohue, F. Guepin, R. H. Hodson, and C. Townsend, "Parameter of nonlinearity in fluids," J. Acoust. Soc. Am. **38**, 797–804 (1965).

[11]W. K. Law, L. A. Frizzell, and F. Dunn, "Ultrasonic determination of the nonlinearity parameter B/A for biological media," J. Acoust. Soc. Am. **69**, 1210–1212 (1981).
[12]W. N. Cobb, "Finite amplitude method for the determination of the acoustic nonlinearity parameter B/A," J. Acoust. Soc. Am. **73**, 1525–1531 (1983).
[13]Z. Zhu, M. S. Roos, W. N. Cobb, and K. Jensen, "Determination of the acoustic nonlinearity parameter B/A from phase measurements," J. Acoust. Soc. Am. **74**, 1518–1521 (1983).
[14]C. M. Sehgal, R. C. Bahn, and J. F. Greenleaf, "Measurement of the acoustic nonlinearity parameter B/A in human tissues by a thermodynamic method," J. Acoust. Soc. Am. **76**, 1023–1029 (1984).
[15]X. Gong, Z. Zhu, T. Shi, and J. Huang, "Determination of the acoustic nonlinearity parameter in biological media using FAIS and ITD methods," J. Acoust. Soc. Am. **86**, 1–5 (1989).
[16]E. C. Everbach and R. E. Apfel, "An interferometric technique for B/A measurement," J. Acoust. Soc. Am. **98**, 3428–3438 (1995).
[17]C. Barriere and D. Royer, "Diffraction effects in the parametric interaction of acoustic waves: Application to measurements of the nonlinearity parameter B/A in liquids," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **48**, 1706–1715 (2001).
[18]E. Chavrier, C. Lafon, A. Bier, C. Barriere, X. Jacob, and D. Cathignol, "Determination of the nonlinear parameter by propagating and modeling finite amplitude plane waves," J. Acoust. Soc. Am. **119**, 2639–2644 (2006).
[19]K. D. Wallace, C. W. Lloyd, M. R. Holland, and J. G. Miller, "Finite amplitude measurements of the nonlinear parameter B/A for liquid mixtures spanning a range relevant to tissue harmonic mode," Ultrasound Med. Biol. **33**, 620–629 (2007).
[20]S. Saito, "Measurement of the acoustic nonlinearity parameter in liquid media using focused ultrasound," J. Acoust. Soc. Am. **93**, 162–172 (1993).
[21]S. Saito, "Measurement of acoustic nonlinearity parameter using focused ultrasound detected by a concave receiver," J. Acoust. Soc. Jpn. (E) **16**, 239–246 (1995).
[22]S. Saito, Y. Kameyama, and K. Nakamura, "Ultrasonic focusing Gaussian source to receive nonlinearly generated second harmonic sound by itself," Jpn. J. Appl. Phys., Part 1 **40**, 3664–3667 (2001).
[23]S. Saito, Y. Kameyama, and K. Nakamura, "High frequency focusing Gaussian source intended for B/A imaging," Acoust. Sci. & Tech. **24**, 406–409 (2003).
[24]S. Saito, "Measurement of acoustic property of small sample beyond acoustic window," Jpn. J. Appl. Phys., Part 1 **46**, 4549–4554 (2007).
[25]S. Saito, S. Takahashi, and K. Saida, "Measurement of acoustic nonlinearity parameter of small-volume liquid sample using focused ultrasound," Jpn. J. Appl. Phys. **47**, 3859–3864 (2008).
[26]S. Saito, "Measurement of the nonlinearity parameter in dispersive liquid using focused ultrasound," Jpn. J. Appl. Phys., Part 1 **37**, 3035–3040 (1998).
[27]L. Germain, R. Jacques, and J. D. N. Cheeke, "Acoustic microscopy applied to nonlinear characterization of biological media," J. Acoust. Soc. Am. **86**, 1560–1565 (1989).
[28]J. Banchet and J. D. N. Cheeke, "Measurement of the acoustic nonlinearity parameter B/A in solvents: Dependence on chain length and sound velocity," J. Acoust. Soc. Am. **108**, 2754–2758 (2000).
[29]J. Kushibiki, M. Ishibashi, N. Akashi, T. Sannomiya, N. Chubachi, and F. Dunn, "Transmission line method for the measurement of the acoustic nonlinearity parameter in biological liquids at very high frequencies," J. Acoust. Soc. Am. **102**, 3038–3044 (1997).
[30]K. Nakamura, H. Ando, and H. Shimizu, "Ferroelectric domain inversion caused in $LiNbO_3$ plates by heat treatment," Appl. Phys. Lett. **50**, 1413–1414 (1987).
[31]K. Nakamura, K. Fukazawa, K. Yamada, and S. Saito, "Broadband ultrasonic transducers using a $LiNbO_3$ plate with a ferroelectric inversion layer," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **50**, 1558–1562 (2003).
[32]K. Nakamura, K. Fukazawa, K. Yamada, and S. Saito, "Ultrasonic transducer for second harmonic imaging using a $LiNbO_3$ plate with a local ferroelectric inversion layer," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **53**, 651–655 (2006).
[33]S. Saito, A. Yamamoto, and K. Nakamura, "B/A measurement for liquid media using an LN transducer with inverted domain layer," Jpn. J. Appl. Phys., Part 1 **44**, 4431–4435 (2005).
[34]T. Sakurada and S. Saito, "Production of 20-MHz focusing Gaussian source," J. Acoust. Soc. Jpn. (E) **21**, 115–118 (2000).
[35]S. Saito and Y. Motohashi, "Nonlinearity parameter measurement for polymer plates using focused ultrasound," Jpn. J. Appl. Phys. **47**, 7380–7385 (2008).
[36]J. Kushibiki and M. Arakawa, "Diffraction effects on bulk-wave ultrasonic velocity and attenuation measurements," J. Acoust. Soc. Am. **108**, 564–573 (2000).
[37]N. Hozumi, A. Kimura, S. Terauchi, M. Nagao, S. Yoshida, K. Kobayashi, and Y. Saijo, "Acoustic impedance micro-imaging for biological tissue using a focused acoustic pulse with a frequency range up to 100 MHz," in Proceedings of the IEEE International Ultrasonics Symposium (2005), pp. 170–173.
[38]N. Hozumi, A. Nakano, S. Terauchi, M. Nagao, S. Yoshida, K. Kobayashi, S. Yamamoto, and Y. Saijo, "Precise calibration for biological acoustic impedance microscope," in Proceedings of the IEEE International Ultrasonics Symposium (2007), pp. 801–804.
[39]L. M. Brekhovskikh, *Waves in Layered Media* (Academic, New York, 1980), pp. 41–50.
[40]S. Saito, "Nonlinearly propagating focused sound transmitted through solid plates," J. Acoust. Soc. Jpn. (E) **18**, 37–45 (1997).
[41]S. Saito, B. C. Kim, and T. G. Muir, "Second harmonic component of a nonlinearly distorted wave in a focused sound field," J. Acoust. Soc. Am. **82**, 621–628 (1987).
[42]S. Saito, "Second harmonic component of focused sound observed with a concave receiver," J. Acoust. Soc. Jpn. (E) **16**, 29–36 (1995).
[43]P.-K. Choi, "Data compilation," in *Cho-Onpa Binran (Ultrasonic Handbook)*, edited by Ultrasonic Handbook Editorial Board (Maruzen, Tokyo, Japan, 1999), pp. 714–723 (in Japanese).
[44]S. Saito and B. C. Kim, "Selective detection of second harmonic sound generated at the focal region in a finite amplitude focusing field," J. Acoust. Soc. Jpn. (E) **8**, 167–175 (1987).
[45]N. Endoh, T. Asahina, K. Kojima, R. Mogi, and T. Nishioka, "Experimental investigation of the nonlinearity parameter by time of flight method," in Proceedings of the 12th ISNA, Austin TX (1990), pp. 391–396.

Effective fractional acoustic wave equations in one-dimensional random multiscale media

Josselin Garnier[a)]
Laboratoire de Probabilités et Modèles Aléatoires and Laboratoire Jacques-Louis Lions, Université Paris 7, 2 Place Jussieu, 75251 Paris Cedex 05, France

Knut Sølna
Department of Mathematics, University of California at Irvine, Irvine, California 92697-3875

(Received 27 May 2009; revised 13 October 2009; accepted 14 October 2009)

This paper considers multiple scattering of waves propagating in a non-lossy one-dimensional random medium with short- or long-range correlations. Using stochastic homogenization theory it is possible to show that pulse propagation is described by an effective deterministic fractional wave equation, which corresponds to an effective medium with a frequency-dependent attenuation that obeys a power law with an exponent between 0 and 2. The exponent is related to the Hurst parameter of the medium, which is a characteristic parameter of the correlation properties of the fluctuations of the random medium. Moreover the frequency-dependent attenuation is associated with a special frequency-dependent phase, which ensures that causality and Kramers–Kronig relations are satisfied. In the time domain the effective wave equation has the form of a linear integro-differential equation with a fractional derivative.
 [DOI: 10.1121/1.3263608]

PACS number(s): 43.28.Lv, 43.20.Bi, 43.20.Fn, 43.20.Hq [VEO] Pages: 62–72

I. INTRODUCTION

It is usually difficult to characterize the microscopic properties of heterogeneous materials from direct observations. Any macroscopic measurement that can provide information about the properties of the material is therefore important. In particular, it is of fundamental interest to understand how the microscopic properties of heterogeneous materials, such as composite or porous media, can be extracted from the observation of waves propagating in such media. The idea is that it should be possible to extract information about the variance of the fluctuations of the parameters, their spatial correlation, and their anisotropy, from measurable quantities such as the attenuation or the dispersion of a pulsed wave propagating through the medium. Wave propagation in random medium has been thoroughly analyzed theoretically and numerically in the case in which the random fluctuations have ergodic or mixing properties, that is, when the random values of the parameters of the medium taken at two points become rapidly uncorrelated when the distance between the two points increases. It is then possible to measure statistical properties of the medium from the waves transmitted through or reflected by a heterogeneous medium.[1–3] Applications are numerous; for instance, in medical imaging, techniques have been developed for estimation of the density of bones via ultrasound echography.[4] However, recent analysis of experimental data has shown that the distribution of the elastic moduli of some porous formations follows a self-similar stochastic distribution with long-range correlations.[5,6] Therefore, it becomes important to study the propagation of waves in random media with long-range correlations. Numerical studies have shown that wave propagation experiments can be used to detect important characteristics of the heterogeneities that a medium may contain, and thus help classify heterogeneous media according to their long-range correlation properties.[7]

The analysis set forth in this paper is an asymptotic analysis based on a separation of scale technique and limit theorems in probability theory that results in the macroscopic description of wave propagation in terms of an effective fractional wave equation. This effective wave equation corresponds to an effective medium with a frequency-dependent attenuation. This effective attenuation is apparent in the sense that the random medium is not lossy but scattering, and multiple scattering is responsible for the observed attenuation of the wave front. It turns out that the effective wave equation has the form of equations used in sound propagation in lossy media, which is quite remarkable and could be useful to understand the physical mechanisms responsible for loss in these media. Indeed frequency-dependent attenuation has been observed in a wide range of applications in acoustics,[8,9] and also in other domains, such as seismic wave propagation.[10–12] Experimental observations show that the attenuation of acoustic waves has a frequency dependence of the form $A=A_0\exp(-\alpha(\omega)z)$, where A denote the amplitude of an acoustic variable such as velocity or pressure and ω is the frequency. The damping coefficient has been seen to obey the empirical power law

$$\alpha(\omega)=\alpha_0|\omega|^y, \tag{1}$$

where $\alpha_0\in(0,\infty)$ and $y\in(0,2)$ are parameters that are characteristic of the medium and obtained through a fitting of measured data.[13–16] Different wave equation models have

a)Author to whom correspondence should be addressed. Electronic mail: garnier@math.jussieu.fr

been proposed to reproduce such a power law.[17–21,9,22] One of the problems discussed in these papers is to obtain a causal wave equation in the space-time domain that reproduces the power law. Another problem is to relate such an equation to first principles in physics.

In our paper we propose a derivation of an effective wave equation in an effective medium with a frequency-dependent attenuation with a power law, and we show that this attenuation is accompanied by a special frequency-dependent phase that ensures the causality of the associated approximation and the satisfaction of the Kramers–Kronig relations. The physical model is a one-dimensional acoustic wave equation in a non-lossy medium with random fluctuations that have short- or long-range correlations. The basic phenomenon is multiple scattering, which leads to effective attenuation and dispersion. The relation between the microscopic statistics of the random medium and the parameters of the frequency-dependent attenuation power law is given and discussed. The effective wave equation has the form of a partial differential equation (PDE) with special fractional derivatives.

II. ACOUSTIC WAVE PROPAGATION IN ONE-DIMENSIONAL RANDOM MEDIA

A. Acoustic wave equations

We develop an asymptotic and probabilistic theory for the acoustic wave equations in the presence of random fluctuations of the medium with short- or long-range correlations. The one-dimensional acoustic wave equations are given by

$$\rho(z)\frac{\partial u}{\partial t}+\frac{\partial p}{\partial z}=0, \tag{2}$$

$$\frac{1}{K(z)}\frac{\partial p}{\partial t}+\frac{\partial u}{\partial z}=0, \tag{3}$$

where t is the time variable, z is the one-dimensional space variable, p is the pressure field, and u is the velocity field. For simplicity we assume that the density of the medium ρ is a constant equal to ρ_0. The bulk modulus of the medium K is assumed to be randomly varying in the region $z\in[0,L]$ and we consider the weakly heterogeneous regime,[2,23,24] in which the fluctuations of the bulk modulus are small and rapid (compared to the propagation distance),

$$\frac{1}{K(z)}=\begin{cases}\dfrac{1}{K_0}\left(1+\varepsilon\nu\left(\dfrac{z}{\varepsilon^2}\right)\right), & z\in[0,L]\\[2ex] \dfrac{1}{K_0}, & z\in(-\infty,0)\cup(L,\infty),\end{cases} \tag{4}$$

$$\rho(z)=\rho_0 \quad \text{for} \quad \text{all} \quad z. \tag{5}$$

Here, the dimensionless parameter ε is small and it characterizes the separation of scales. The effective impedance and speed of sound are $\zeta_0=\sqrt{K_0\rho_0}$ and $c_0=\sqrt{K_0/\rho_0}$, respectively. The source located at $z_0<0$ emits a pulse at time z_0/c_0. This pulse impinges on the section $[0,L]$ and hits the boundary at 0 at time 0.

The random process ν is assumed to be stationary and to have mean zero. Its autocorrelation function is denoted by

$$\phi(z)=\mathbb{E}[\nu(x)\nu(x+z)], \tag{6}$$

where $\mathbb{E}$ stands for the expectation with respect to the distribution of the random medium. The function $\phi(z)$ is bounded, even, and maximal at zero [$\phi(0)$ is the variance of the fluctuations]. Its local regularity at zero and its asymptotic behavior at infinity characterize the short- and long-range correlations of the random medium, as we discuss in Sec. II B.

B. Random medium properties

Wave propagation in random media is usually studied in the case in which the driving process ν has mixing properties. This means that the random values $\nu(x+z)$ and $\nu(x)$ taken at two points separated by the distance z become rapidly uncorrelated when $z\to\infty$. More precisely we say that the random process ν is mixing if its autocorrelation function $\phi(z)$ decays fast enough at infinity so that it is absolutely integrable as follows:

$$\int_0^\infty |\phi(z)|dz<\infty. \tag{7}$$

This is the usual assumption for random media, under which the theory is well established. In this case the correlation length can be defined as $l_c=2\int_0^\infty\phi(z)dz/\phi(0)$. The O'Doherty–Anstey theory describes the propagating pulse in this regime. The effective equation for the wave front has been obtained by several authors.[25–27,2,28,29] The pulse propagation is characterized by a random time shift and a deterministic spreading. The random time shift is described in terms of a standard Brownian motion, while the deterministic spreading is described by a pseudo-differential operator that we will describe in Sec. III. If, additionally, the correlation length of the medium is smaller than the typical wavelength, then the pseudo-differential operator can be reduced to a second-order diffusion.[2]

Wave propagation in multiscale and rough media, with short- or long-range fluctuations, has recently attracted a lot of attention, as more and more data collected in real environments confirm that this situation can be encountered in many different contexts, such as in geophysics[5,6] or in laser beam propagation through the atmosphere.[30–32]

Qualitatively, the long-range correlation property means that the random process has long memory (in contrast with a mixing process). This means that the correlation degree between the random values $\nu(x+z)$ and $\nu(x)$ taken at two points separated by the distance z decays only slowly when $z\to\infty$, or equivalently that the autocorrelation function has a slow decay at infinity. More precisely we say that the random process ν has the H-long-range correlation property if its autocorrelation function satisfies

$$\phi(z)\overset{|z|\to\infty}{\simeq} r_H\left|\frac{z}{l_c}\right|^{2H-2}, \tag{8}$$

where $r_H>0$ and $H\in(1/2,1)$. Here the correlation length l_c is the critical length scale beyond which power law behavior (8) is valid. Note that the autocorrelation function is not integrable since $2H-2\in(-1,0)$, which means that a random process with the H-long-range correlation property is not mixing.

Qualitatively the short-range correlation property means that the random process is rough at small scales. This means that the correlation degree between the random values $\nu(x+z)$ and $\nu(x)$ taken at two points separated by the distance z has a sharp decay at zero. It corresponds to the fact that the autocorrelation function decays faster than an affine function at zero. More precisely we say that the random process ν has the H-short-range correlation property if its autocorrelation function satisfies

$$\phi(z) \overset{|z|\to 0}{\simeq} \phi(0)\left(1-d_H\left|\frac{z}{l_c}\right|^{2H}+O\left(\left|\frac{z}{l_c}\right|\right)\right), \tag{9}$$

where $d_H>0$ and $H\in(0,1/2)$. Here the correlation length l_c is the critical length scale below which power law behavior (9) is valid. Additional technical hypotheses on the differentiability of ϕ are necessary for the mathematical proof (see Appendixes B and C). We simply mention here that these technical hypotheses are satisfied by the models presented in Sec. II C.

C. Random medium models with short- or long-range correlations

In this section we present random processes ν that satisfy the conditions that we have imposed on the medium fluctuations. The first model we may think of is based on the fractional Brownian motion (fBM), which is known to have special properties in terms of long-range dependence and in terms of roughness of its trajectories. We remind the reader that the fBM $W_H(z)$ with Hurst index $H\in(0,1)$ is the Gaussian process with mean zero and covariance

$$\mathbb{E}[W_H(x)W_H(x+z)]=\tfrac{1}{2}(|x+z|^{2H}+|x|^{2H}-|z|^{2H}). \tag{10}$$

The fBM is a self-similar process, which means that, for any $a>0$, the statistical distribution of the scaled process $(W_H(az))_{z\geq 0}$ is the same as the one of $(a^H W_H(z))_{z\geq 0}$. The fBM has stationary increments

$$\mathbb{E}[(W_H(x+z)-W_H(x))^2]=|z|^{2H}. \tag{11}$$

For $H=1/2$ the fBM is the standard Brownian motion, which has independent increments and Hölder continuous trajectories of any order strictly less than 1/2.

For $H<1/2$ the fBM is a random process with negatively correlated increments (which is a type of short-range correlation property). The realizations are Hölder continuous trajectories of any order strictly less than H, which means that they are more irregular than the trajectories of the standard Brownian motion.

For $H>1/2$ the fBM is a random process with positively correlated increments (which is a type of long-range correlation property). The realizations are Hölder continuous trajectories of any order strictly less than H, which means that they are more regular than the trajectories of the standard Brownian motion.

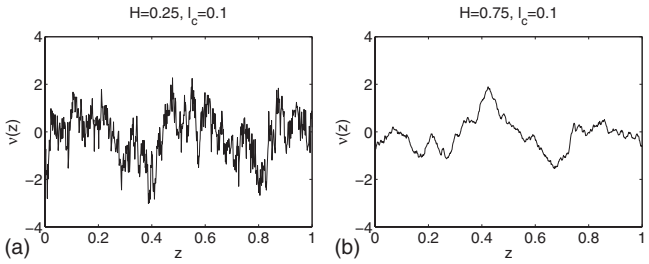


FIG. 1. Realizations of the fractional OU process with Hurst index H and correlation length l_c. The trajectories are more regular when H is larger.

However, the fBM is not stationary. We will first introduce two random processes that have local properties similar to the ones of the fBM but are stationary. We will also introduce two binary medium models that have either the long- or short-range correlation property.

Fractional Ornstein–Uhlenbeck medium. The fractional Ornstein–Uhlenbeck (OU) process $\nu(z)$ is defined by

$$\nu(z)=\frac{\sigma}{\sqrt{H\Gamma(2H)}l_c^H}\left[W_H(z)-\frac{1}{l_c}\int_{-\infty}^{z}e^{(x-z)/l_c}W_H(x)dx\right], \tag{12}$$

where W_H is a fBM with Hurst index $H\in(0,1)$ and $\Gamma(h)=\int_0^\infty e^{-t}t^{h-1}dt$ is the Gamma function. The fractional OU process can be interpreted as a fBM with a restoring force toward zero. The fractional OU process is a zero-mean, stationary, Gaussian process. Its variance is σ^2 and its autocorrelation function is given by

$$\phi(z)=\frac{\sigma^2}{H\Gamma(2H)l_c^{2H}}\left[\frac{1}{4l_c}\int_{-\infty}^{\infty}e^{-|x|/l_c}|z+x|^{2H}dx-\frac{1}{2}|z|^{2H}\right]. \tag{13}$$

If $H=1/2$, then the standard OU process (synthesized with a standard Brownian motion) is a stationary Gaussian Markov process. Its autocorrelation function is

$$\phi(z)=\sigma^2\exp\left(-\frac{|z|}{l_c}\right), \tag{14}$$

which shows that it is a mixing process.

If $H\in(1/2,1)$ then the fractional OU process has the H-long-range correlation property since its autocorrelation function $\phi(z)$ satisfies Eq. (8) with

$$r_H=\frac{\sigma^2(2H-1)}{\Gamma(2H)}. \tag{15}$$

If $H\in(0,1/2)$, then the fractional OU process has the H-short-range correlation property since its autocorrelation function $\phi(z)$ satisfies Eq. (9) with

$$d_H=\frac{1}{2H\Gamma(2H)}. \tag{16}$$

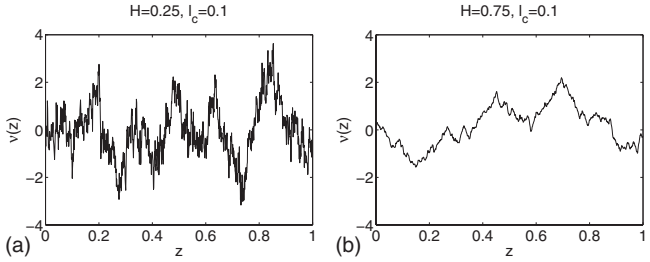


FIG. 2. Realizations of the fractional white noise with Hurst index H and correlation length l_c. The trajectories are more regular when H is larger.

It is possible to simulate paths of the fractional OU process using the Cholesky method (see Fig. 1) or other methods.[33]

Fractional white noise medium. As a second example we consider the model

$$\nu(z) = \frac{\sigma}{l_c^H}[W_H(z+l_c) - W_H(z)], \tag{17}$$

where W_H is a fBM with Hurst index $H \in (0,1)$. The fractional white noise is a zero-mean, stationary, Gaussian process. Its variance is σ^2 and its autocorrelation function is given by

$$\phi(z) = \frac{\sigma^2}{2l_c^{2H}}(|z+l_c|^{2H} + |z-l_c|^{2H} - 2|z|^{2H}). \tag{18}$$

If $H=1/2$, then the standard white noise process (synthesized with a standard Brownian motion) is a stationary Gaussian process. Its autocorrelation function is

$$\phi(z) = \sigma^2\left(1 - \frac{|z|}{l_c}\right)\mathbf{1}_{[0,l_c]}(|z|), \tag{19}$$

which shows that it is a mixing process. Here the indicator function $\mathbf{1}_A(z)$ is equal to 1 if $z \in A$ and 0 otherwise.

If $H \in (1/2,1)$ then the fractional white noise has the H-long-range correlation property since its autocorrelation function $\phi(z)$ satisfies Eq. (8) with $r_H = \sigma^2 H(2H-1)$.

If $H \in (0,1/2)$, then the fractional white noise has the H-short-range correlation property. Indeed its autocorrelation function $\phi(z)$ satisfies Eq. (9) with $d_H = 1$.

Note that the typical trajectories of the fractional white noise are very similar to the ones of the fractional OU process (compare Figs. 1 and 2).

Binary medium with mixing property. Here we construct a process corresponding to a binary medium so that the process ν is stepwise constant and takes values $\pm\sigma$ over intervals with random lengths. We denote by $(l_j)_{j\geq 0}$ the lengths of these intervals and by $(n_j)_{j\geq 0}$ the values taken by the process over each elementary interval. The process $\nu(z)$ is defined by

$$\nu(z) = n_{N_z} \quad \text{where} \quad N_z = \sup\{n \geq 0, L_n \leq z\}, \tag{20}$$

with $L_0=0$ and $L_{n+1}=L_n+l_n$. The random variables n_j are independent and identically distributed (i.i.d.) with the distribution

$$\mathbb{P}(n_1 = \pm\sigma) = \tfrac{1}{2}. \tag{21}$$

The random variables l_j are i.i.d. with the exponential distribution whose probability density function (pdf) is

$$p_{l_1}(z) = \frac{1}{l_c}\exp\left(-\frac{z}{l_c}\right)\mathbf{1}_{[0,\infty)}(z). \tag{22}$$

Note that it is very easy to simulate the random variable l_1, since $-l_c \ln U$ has pdf (22) if U is uniformly distributed over $[0,1]$. The random process $\nu(z)$ is a stationary jump Markov process and its autocorrelation function is

$$\phi(z) = \sigma^2 \exp\left(-\frac{|z|}{l_c}\right), \tag{23}$$

which shows that it is a mixing process. Note that the binary medium process has the same autocorrelation function as the standard OU process, although these two random processes are very different (the first one is a jump process that takes only two values, and the second one is a continuous process). The autocorrelation function is not sufficient to characterize the full statistics of a random medium, but we will see that it is sufficient to characterize wave propagation in the random medium.

Binary medium with long-range correlations. The long-range correlation property for a binary medium corresponds to the existence of intervals much longer than the average interval length. We again consider the process defined by Eq. (20) corresponding to a binary medium where the random variables n_j are i.i.d. with distribution (21) and the random variables l_j are i.i.d. with the distribution with the pdf

$$p_{l_1}(z) = (3-2H)\frac{l_c^{3-2H}}{z^{4-2H}}\mathbf{1}_{[l_c,\infty)}(z), \tag{24}$$

where $H \in (1/2,1)$. Note that it is very easy to simulate the random variable l_1, since $l_c U^{-1/(3-2H)}$ has pdf (24) if U is uniformly distributed over $[0,1]$. The average length of the random interval is

$$\mathbb{E}[l_1] = \frac{3-2H}{2-2H}l_c, \tag{25}$$

while the variance of l_1 is infinite. A salient aspect of this model is that very long intervals (i.e., much longer than $\mathbb{E}[l_1]$) can be generated, which are responsible for the infinite variance of the length of the interval and for the long-range correlation property of the random medium. The process ν is bounded, it has mean zero and variance σ^2, but it is not stationary. However, using renewal theory,[34] the distribution of the process $(\nu(x+z))_{z\geq 0}$ converges to a stationary distribution when $x \to \infty$ and the autocorrelation function of ν satisfies

$$\mathbb{E}[\nu(x)\nu(x+z)] \overset{x\to\infty}{\to} \phi(z) = \sigma^2 \int_z^\infty \frac{\mathbb{P}(l_1 > s)}{\mathbb{E}[l_1]} ds.$$

The autocorrelation function

$$\phi(z) = \sigma^2 \left[\frac{1}{3-2H} \frac{l_c^{2-2H}}{|z|^{2-2H}} \mathbf{1}_{[l_c,\infty)}(|z|) + \left(1 - \frac{2-2H}{3-2H} \frac{|z|}{l_c} \right) \mathbf{1}_{[0,l_c)}(|z|) \right] \tag{26}$$

satisfies H-long-range correlation property (8) with $r_H = \sigma^2/(3-2H)$.

It is also possible to make the process stationary by simply modifying the statistical distribution of the length of the first interval: If the random lengths $(l_j)_{j\geq 1}$ are i.i.d. according to the distribution with pdf (24), and if l_0 is independent of the $(l_j)_{j\geq 1}$ and has the distribution with the pdf

$$p_{l_0}(z) = \frac{\mathbb{P}(l_1 > z)}{\mathbb{E}[l_1]} = \frac{2-2H}{3-2H} \frac{1}{l_c} \mathbf{1}_{[0,l_c)}(z) + \frac{2-2H}{3-2H} \frac{l_c^{2-2H}}{z^{3-2H}} \mathbf{1}_{[l_c,\infty)}(z), \tag{27}$$

then the process ν is bounded, zero-mean, and stationary, and it has the H-long-range correlation property since its autocorrelation function $\mathbb{E}[\nu(x)\nu(x+z)]$ is given by Eq. (26) for any x.

Binary medium with short-range correlations. The short-range correlation property for a binary medium corresponds to the accumulation of intervals with lengths much smaller than the average length. We again consider the process defined by Eq. (20) corresponding to a binary medium where the random variables n_j are i.i.d. with distribution (21) and the random variables l_j are i.i.d. with the distribution with the pdf

$$p_{l_1}(z) = \frac{1-2H}{(l_i/l_c)^{2H-1} - 1} \frac{l_c^{1-2H}}{z^{2-2H}} \mathbf{1}_{[l_i,l_c]}(z), \tag{28}$$

where $H \in (0,1/2)$ and $0 < l_i < l_c$. Here the inner scale l_i is introduced in order to obtain a well-defined and normalized pdf , and it will be taken to be much smaller than the correlation length l_c. Note that it is very easy to simulate the random variable l_1, since $l_i[1-(1-(l_i/l_c)^{1-2H})U]^{-1/(1-2H)}$ has pdf (28) if U is uniformly distributed over $[0,1]$. The average length of the random interval is

$$\mathbb{E}[l_1] = \frac{1-2H}{2H} \frac{l_c^{2H} - l_i^{2H}}{l_i^{2H-1} - l_c^{2H-1}}, \tag{29}$$

and its variance is finite. A salient aspect of this model is that it exhibits an accumulation of very small intervals (i.e., much smaller than $\mathbb{E}[l_1]$), which corresponds to very rapid changes in the medium properties. Using renewal theory the distribution of the process $(\nu(x+z))_{z\geq 0}$ converges to a stationary distribution when $x \to \infty$ and the autocorrelation function of ν satisfies

$$\mathbb{E}[\nu(x)\nu(x+z)] \overset{x\to\infty}{\to} \phi(z), \tag{30}$$

where

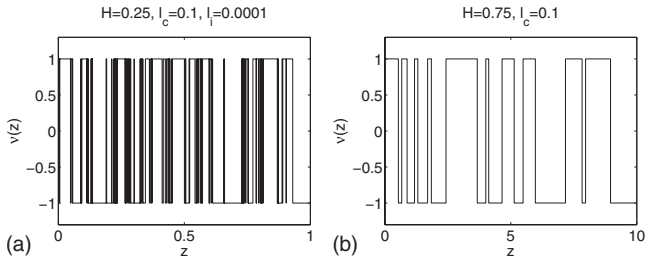


FIG. 3. Realizations of a binary medium with the index H. When $H<1/2$ we can observe an accumulation of very small intervals responsible for the H-short-range correlation property (a). When $H>1/2$ we can observe the existence of some intervals longer than the average responsible for the H-long-range correlation property (b).

$$\phi(z) = \frac{\sigma^2}{1-(l_i/l_c)^{2H}} \left(1 - \frac{1}{1-2H} \frac{|z|^{2H}}{l_c^{2H}} + \frac{2H}{1-2H} \frac{|z|}{l_c} \right) \times \mathbf{1}_{(l_i,l_c]}(|z|) + \sigma^2 \left(1 - \frac{2H}{1-2H} \frac{((l_i/l_c)^{2H} - (l_i/l_c))}{1-(l_i/l_c)^{2H}} \frac{|z|}{l_i} \right) \times \mathbf{1}_{[0,l_i]}(|z|). \tag{31}$$

If $l_i \ll l_c$, then

$$\phi(z) \simeq \sigma^2 \left(1 - \frac{1}{1-2H} \frac{|z|^{2H}}{l_c^{2H}} + \frac{2H}{1-2H} \frac{|z|}{l_c} \right) \mathbf{1}_{(0,l_c]}(|z|), \tag{32}$$

which satisfies H-short-range correlation property (9) with $d_H = 1/(1-2H)$.

Note that the trajectories of the process $\nu(z)$ for a binary medium (see Fig. 3) are very different from the ones of the fractional OU process or the fractional white noise. However, we will see that wave propagation in these random media can be described in the same effective way.

III. ANALYSIS OF THE WAVE FRONT

A. Mode decomposition

We consider the right- and left-going waves defined in terms of the local impedance and moving with the local sound speed

$$\begin{bmatrix} A^\varepsilon(t,z) \\ B^\varepsilon(t,z) \end{bmatrix} = \begin{bmatrix} \zeta^{\varepsilon-1/2}(z) p(t,z) + \zeta^{\varepsilon 1/2}(z) u(t,z) \\ -\zeta^{\varepsilon-1/2}(z) p(t,z) + \zeta^{\varepsilon 1/2}(z) u(t,z) \end{bmatrix}. \tag{33}$$

From Eqs. (4) and (5) the local impedance and sound speed are given by

$$\zeta^\varepsilon(z) = \sqrt{K(z)\rho(z)} = \frac{\zeta_0}{\sqrt{1+\varepsilon\nu(z/\varepsilon^2)}}, \tag{34}$$

$$c^\varepsilon(z) = \sqrt{K(z)/\rho(z)} = \frac{c_0}{\sqrt{1+\varepsilon\nu(z/\varepsilon^2)}}. \tag{35}$$

The mode amplitudes satisfy

$$\frac{\partial}{\partial z} \begin{bmatrix} A^\varepsilon \\ B^\varepsilon \end{bmatrix} = -\frac{1}{c^\varepsilon(z)} \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \frac{\partial}{\partial t} \begin{bmatrix} A^\varepsilon \\ B^\varepsilon \end{bmatrix} + \frac{\partial_z \zeta^\varepsilon(z)}{2\zeta^\varepsilon(z)} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} A^\varepsilon \\ B^\varepsilon \end{bmatrix}. \tag{36}$$

This system is completed with an initial condition corre-

sponding to a right-going pulse wave that is incoming from the homogeneous half-space $z<0$ and is impinging on the random medium in $[0,L]$,

$$A^\varepsilon(t,z)=f\left(\frac{t-z/c_0}{\varepsilon^2}\right),\quad B^\varepsilon(t,z)=0,\quad t<0. \tag{37}$$

Equation (36) clearly exhibits the two important aspects of the propagation mechanisms. The first term on the right describes transport along the random characteristics with the local sound speed $c^\varepsilon(z)$. The second term on the right describes coupling between the right- and left-going modes, which is proportional to the derivative of the logarithmic impedance. The asymptotic analysis of this system is given in the mixing case in Ref. 2 and in the long-range case in Ref. 35. The main steps are the following ones. We first perform a series of transformations to rewrite the evolution equations of the modes by centering along the characteristic of the right-going mode. We then obtain an upper-triangular system that can be integrated more easily. In a second step we apply a limit (homogenization) theorem to this system in order to establish an effective equation for the wave front. The result of this asymptotic theory is given in Sec. III B.

B. Wave front transmission

We now state the fundamental results that characterize the wave front transmitted through a random medium. Let us introduce the random travel time

$$\tau_0^\varepsilon(z)=\frac{z}{c_0}+\frac{\varepsilon}{2c_0}\int_0^z \nu\left(\frac{x}{\varepsilon^2}\right)dx. \tag{38}$$

(1) If the random process is mixing or if it satisfies the H-long-range correlation property, $H\in(1/2,1)$, then the wave front observed in the random frame moving with the random travel time

$$A^\varepsilon(\tau_0^\varepsilon(z)+\varepsilon^2\tau,z),\quad z>0 \tag{39}$$

converges in distribution as $\varepsilon\to 0$ to the deterministic profile

$$a(\tau,z)=\frac{1}{2\pi}\int \exp\left(-i\omega\tau-\frac{\gamma_c(\omega)\omega^2}{8c_0^2}z-i\frac{\gamma_s(\omega)\omega^2}{8c_0^2}z\right)\times\hat f(\omega)d\omega, \tag{40}$$

where $\hat f(\omega)$ is the Fourier transform of the initial pulse and

$$\gamma_c(\omega)=2\int_0^\infty \phi(z)\cos\left(\frac{2\omega z}{c_0}\right)dz, \tag{41}$$

$$\gamma_s(\omega)=2\int_0^\infty \phi(z)\sin\left(\frac{2\omega z}{c_0}\right)dz. \tag{42}$$

(2) If the random process is mixing then the expectation of the random travel time $\tau_0^\varepsilon(z)$ is z/c_0 and its variance is

$$\mathrm{Var}(\tau_0^\varepsilon(z))=\frac{\varepsilon^4}{c_0^2}\frac{\gamma_c(0)}{4}z+o(\varepsilon^4), \tag{43}$$

as $\varepsilon\to 0$. The random travel time $\tau_0^\varepsilon(z)$ has the distribution of

$$\frac{z}{c_0}+\frac{\varepsilon^2}{c_0}\frac{\sqrt{\gamma_c(0)}}{2}W(z)+o(\varepsilon^2), \tag{44}$$

as $\varepsilon\to 0$, where $W(z)$ is a standard Brownian motion.

(3) Under the H-long-range correlation property $H\in(1/2,1)$, the expectation of the random travel time $\tau_0^\varepsilon(z)$ is z/c_0 and its variance is

$$\mathrm{Var}(\tau_0^\varepsilon(z))=\frac{\varepsilon^{2(3-2H)}l_c^{2-2H}}{c_0^2}\frac{r_H}{4H(2H-1)}z^{2H}+o(\varepsilon^{2(3-2H)}), \tag{45}$$

as $\varepsilon\to 0$. The random travel time $\tau_0^\varepsilon(z)$ has the distribution of

$$\frac{z}{c_0}+\frac{\varepsilon^{3-2H}l_c^{1-H}}{c_0}\sqrt{\frac{r_H}{4H(2H-1)}}W_H(z)+o(\varepsilon^{3-2H}), \tag{46}$$

as $\varepsilon\to 0$, where $W_H(z)$ is a fBM with Hurst index H.

The first point in the mixing case and the second point are classical results known as the O'Doherty–Anstey theory[36] and they can be found in Refs. 1, 26, and 2. The third point was established in Ref. 23 for a certain class of subordinated Gaussian processes and in Ref. 35 for the processes under consideration in this paper. The first point in the long-range case was proved in Ref. 35.

We see from the first point that the frequency-dependent attenuation

$$\frac{\gamma_c(\omega)\omega^2}{8c_0^2} \tag{47}$$

of the wave front in Eq. (40) is always non-negative because $\gamma_c(\omega)$ is the power spectral density of the stationary fluctuations $\nu(z)$ of the random medium.

The term $\exp[-i\gamma_s(\omega)\omega^2 z/(8c_0^2)]$ in Eq. (40) is a frequency-dependent phase modulation and $\gamma_s(\omega)$ is conjugate to $\gamma_c(\omega)$. This shows that the transmitted wave front when centered with respect to the random travel time correction propagates in a dispersive effective medium with the frequency-dependent wavenumber given by

$$k(\omega)=\frac{\omega}{c_0}-\varepsilon^2\frac{\gamma_s(\omega)\omega^2}{8c_0^2}, \tag{48}$$

up to smaller terms in ε.

The fundamental results stated above show that the transmitted wave front in the random medium is modified in two ways compared to propagation in a homogeneous one.

First, its arrival time at the end of the slab $z=L$ has a small random component. In the usual case in which ϕ is integrable, the random time shift is of order ε^2 and its statistical distribution is described in terms of a standard Brownian motion. In the long-range correlation case, the random time shift is of order ε^{3-2H} and its statistical distribution is

described in terms of a fBM.[23] Remember, however, that the pulse width is of order ε^2, which means that the random time delay is large compared to the pulse width; moreover, it becomes relatively larger as H is closer to 1.

Second, if we observe the wave front near its random arrival time, then we see a pulse profile that, to leading order, is deterministic and is the original pulse shape convolved with a deterministic kernel that depends only on the second-order statistics of the medium (i.e., the autocorrelation function of ν)

$$a(\tau,z) = [\mathcal{H}(\cdot,z) * f](\tau). \tag{49}$$

The convolution kernel is given by

$$\mathcal{H}(\tau,z) = \frac{1}{2\pi}\int \exp\left(-i\omega\tau - \frac{\gamma_c(\omega)\omega^2}{8c_0^2}z - i\frac{\gamma_s(\omega)\omega^2}{8c_0^2}z\right)d\omega. \tag{50}$$

We describe the effective pulse attenuation and dispersion in Sec. III C.

C. Deterministic pulse deformation

In this section we analyze the main properties of the effective equation for the wave front: The important function affecting the dynamics is Fourier transforms (41) and (42) of the positive lag part of the autocorrelation function of the random fluctuations of the medium. We have stated that $A^\varepsilon(\tau_0^\varepsilon(z)+\varepsilon^2\tau,z)$ converges to a given by Eq. (40). By taking an inverse Fourier transform, it is possible to identify the PDE satisfied by a as follows:

$$\frac{\partial a}{\partial z} = \mathcal{L}a, \tag{51}$$

where $\mathcal{L}$ is a pseudo-differential operator that describes the deterministic pulse deformation

$$\mathcal{L} = \mathcal{L}_c + \mathcal{L}_s, \tag{52}$$

$$\int_{-\infty}^{\infty} \mathcal{L}_c a(\tau)e^{i\omega\tau}d\tau = -\frac{\gamma_c(\omega)\omega^2}{8c_0^2}\int_{-\infty}^{\infty} a(\tau)e^{i\omega\tau}d\tau, \tag{53}$$

$$\int_{-\infty}^{\infty} \mathcal{L}_s a(\tau)e^{i\omega\tau}d\tau = -\frac{i\gamma_s(\omega)\omega^2}{8c_0^2}\int_{-\infty}^{\infty} a(\tau)e^{i\omega\tau}d\tau. \tag{54}$$

Note that the function $\gamma_c(\omega)$ is even and the function $\gamma_s(\omega)$ is odd, so that $\mathcal{L}_c$ and $\mathcal{L}_s$ are two real operators [i.e., if $a(\tau)$ is a real-valued function, then $\mathcal{L}_{c,s}a(\tau)$ is real-valued]. PDE (51) is completed with the initial condition $a(\tau,z=0)=f(\tau)$.

The first qualitative property satisfied by effective PDE (51) is that it preserves the causality. Indeed, in the time domain, we can write

$$\mathcal{L}a(\tau) = \frac{1}{8c_0}\int_0^\infty \phi\left(\frac{c_0 s}{2}\right)\frac{\partial^2 a}{\partial\tau^2}(\tau-s)ds. \tag{55}$$

This shows that, if a vanishes for $\tau<0$, then $\mathcal{L}a$ also vanishes for $\tau<0$, which is a manifestation of causality.

The pseudo-differential operator also satisfies the Kramers–Kronig relations.[37,38] Indeed the function

$$\omega^2[\gamma_c(\omega) + i\gamma_s(\omega)] = 2\omega^2\int_0^\infty \phi(z)\exp\left(i\frac{2\omega z}{c_0}\right)dz \tag{56}$$

is analytic in the upper complex half plane and vanishes as $|\omega|\to\infty$. Note that our effective equation is obtained via an asymptotic analysis of a system of equations that are causal, so that there is no surprise that causality and Kramers–Kronig relations are satisfied by the effective equation.

Equation (51) gives the effective evolution of the front wave $a(\tau,z)$ in the frame moving with the random travel time $\tau_0^\varepsilon(z)$. If we ignore the small random time shift and focus our attention to the deterministic pulse deformation, then we can write the effective equation in the form of a wave equation for the pressure in the original frame and at the scale $z=\varepsilon^2\xi$ and $t=\varepsilon^2\tau$ as

$$\frac{\partial p}{\partial\xi} + \frac{1}{c_0}\frac{\partial p}{\partial\tau} = \varepsilon^2\mathcal{L}p,$$

$$p(\tau,\xi=0) = \zeta_0^{1/2}f(\tau). \tag{57}$$

This PDE is valid up to $\xi=L/\varepsilon^2$ in the asymptotic framework $\varepsilon\to 0$ and it is equivalent to the following effective wave equation:

$$\frac{\partial^2 p}{\partial\xi^2} - \frac{1}{c_0^2}\frac{\partial^2 p}{\partial\tau^2} = \frac{2\varepsilon^2}{c_0}(\partial_\tau\mathcal{L})p,$$

$$p(\tau,\xi=0) = \zeta_0^{1/2}f(\tau), \quad \partial_\xi p(\tau,\xi=0) = -\frac{\zeta_0^{1/2}}{c_0}f'(\tau). \tag{58}$$

The pseudo-spectral operator $\mathcal{L}$ can be divided into two parts as in Eq. (52). The first component $\mathcal{L}_c$, as pointed out in Sec. III B after Eq. (47), is a frequency-dependent attenuation that can be interpreted as an effective diffusion operator. The second component $\mathcal{L}_s$ is an effective dispersion operator, since it preserves the energy. Equation (58) can be written in the form of an integro-differential equation

$$\frac{\partial^2 p}{\partial\xi^2} - \frac{1}{c_0^2}\frac{\partial^2 p}{\partial\tau^2} = \frac{\varepsilon^2}{4c_0^2}\int_0^\infty \phi\left(\frac{c_0 s}{2}\right)\frac{\partial^3 p}{\partial\tau^3}(\tau-s)ds.$$

D. Mixing random media

In this subsection we consider the case of a mixing random medium. We first assume that the typical wavenumber ω/c_0 of the input pulse is such that $|\omega|l_c/c_0 \ll 1$. This condition means that the typical wavelength is longer than l_c and in this case we find that

$$\frac{\gamma_c(\omega)\omega^2}{c_0^2} = \frac{\gamma_c(0)\omega^2}{c_0^2}, \quad \frac{\gamma_s(\omega)\omega^2}{c_0^2} = 0, \tag{59}$$

which shows that we have an effective second-order diffusion [i.e., a quadratic frequency dependence of the attenuation of the form of Eq. (1) with $y=2$] and no effective dispersion.

We next consider the case of a mixing random medium with the affine decay behavior at zero as follows:

$$\phi(z) = \phi(0)\left(1 - d_{1/2}\frac{|z|}{l_c} + o\left(\frac{|z|}{l_c}\right)\right), \tag{60}$$

with $d_{1/2} \geq 0$. Affine decay (60) of the autocorrelation function is typical of a Markov process, such as the standard OU process synthesized with a standard Brownian motion or the binary medium process in the case in which the lengths of the intervals have an exponential distribution. This behavior is, in fact, fairly general. For instance, affine decay (60) holds for the white noise process synthesized with a standard Brownian motion or the binary medium process in the case in which the lengths of the intervals have an arbitrary distribution with positive finite expectation. If we assume that the typical wavenumber ω/c_0 of the input pulse is such that $|\omega| l_c/c_0 \gg 1$, which means that the typical wavelength is smaller than l_c, then we have (see Appendix A)

$$\frac{\gamma_c(\omega)\omega^2}{c_0^2} = \frac{\phi(0) d_{1/2}}{2 l_c}, \quad \frac{\gamma_s(\omega)\omega^2}{c_0^2} = \frac{\phi(0)\omega}{c_0}, \tag{61}$$

which shows that we have an effective constant attenuation [of the form of Eq. (1) with $y=0$] and no effective dispersion.

These two cases (quadratic and constant attenuations) are the ones observed with standard models of random media, that are mixing and not rough, and they have been extensively studied.[1,2] As we will see in Secs. III E and III F, the picture becomes more interesting when non-mixing or rough random media are considered.

E. Random media with long-range correlations

This is an interesting regime that leads to explicit formulas. This is the regime in which the random medium has the H-long-range correlation property $H \in (1/2, 1)$ and the typical wavenumber ω/c_0 of the input pulse is such that

$$\frac{|\omega| l_c}{c_0} \ll 1.$$

This second condition means that the typical wavelength is longer than l_c and therefore the pulse probes the long-range properties of the medium. In this case we find by using Eq. (8) and formula 3.761 in Ref. 39 that

$$\frac{\gamma_c(\omega)\omega^2}{c_0^2} = r_H \frac{\Gamma(2H-1)}{2^{2H-2}} \cos\left(\left(H - \frac{1}{2}\right)\pi\right) \frac{1}{l_c}\left(\frac{|\omega| l_c}{c_0}\right)^{3-2H}, \tag{62}$$

$$\frac{\gamma_s(\omega)\omega^2}{c_0^2} = r_H \frac{\Gamma(2H-1)}{2^{2H-2}} \sin\left(\left(H - \frac{1}{2}\right)\pi\right) \left(\frac{|\omega| l_c}{c_0}\right)^{2-2H} \frac{\omega}{c_0}. \tag{63}$$

This shows that the wave propagation in random media with long-range correlations exhibits frequency-dependent attenuation that is characterized by a power law of the form of Eq. (1) with the exponent $y = 3-2H$ ranging from 1 to 2. This exponent is related to the power decay rate at infinity of the autocorrelation function of the medium fluctuations. The frequency-dependent attenuation is associated with a frequency-dependent phase and the ratio of the effective dispersion coefficient over the effective diffusion coefficient is

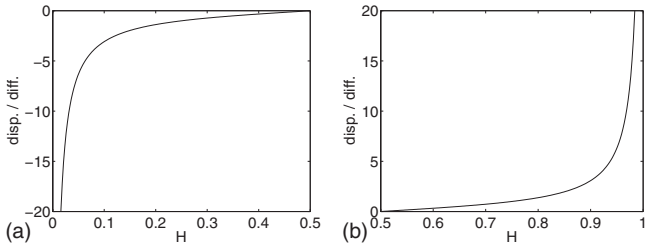


FIG. 4. Ratios of the effective dispersion coefficient over the effective diffusion coefficient. Note the symmetry of the ratio with respect to $H=1/2$.

$$R_{\text{disp/diff}}(H) = \frac{\sin\left(\left(H - \frac{1}{2}\right)\pi\right)}{\cos\left(\left(H - \frac{1}{2}\right)\pi\right)}, \tag{64}$$

which shows that the effective dispersion is stronger than the effective diffusion when H is close to 1 and that it is weaker when H is close to 1/2 (see Fig. 4). Note also that we recover the standard mixing case [formula (59)] when $H \to 1/2$.

In the time domain, we can write

$$\mathcal{L}a(\tau) = \frac{r_H l_c^{2-2H}}{2^{1+2H} c_0^{3-2H}} \int_0^\infty \frac{1}{s^{2-2H}} \frac{\partial^2 a}{\partial \tau^2}(\tau - s) ds. \tag{65}$$

If we go back to the original frame and at the scale $z = \varepsilon^2 \xi$ and $t = \varepsilon^2 \tau$ and if we substitute expression (65) of the pseudo-differential operator into Eq. (58), then we obtain the effective fractional wave equation

$$\frac{\partial^2 p}{\partial \xi^2} - \frac{1}{c_0^2}\frac{\partial^2 p}{\partial \tau^2} = \frac{r_H l_c^{2-2H} \varepsilon^2}{2^{2H} c_0^{4-2H}} \int_0^\infty \frac{1}{s^{2-2H}} \frac{\partial^3 p}{\partial \tau^3}(\tau - s) ds. \tag{66}$$

This effective wave equation has the form of the wave equation used in sound propagation in lossy media,[9] but not exactly. Indeed, in Ref. 9 Szabo proposed to use the Riemann–Liouville fractional derivative[40] and to write the fractional wave equation for lossy media in the form

$$\frac{\partial^2 p}{\partial \xi^2} - \frac{1}{c_0^2}\frac{\partial^2 p}{\partial \tau^2} = c_{\varepsilon,H} \int_0^\infty \frac{1}{s^{5-2H}} p(\tau - s) ds. \tag{67}$$

By carrying three integrations by parts steps one can see that the Caputo derivative in Eq. (66) and the Riemann–Liouville derivative in Eq. (67) are related to each other but there are differences in the form of local terms produced by the integration by parts formula.[17] In particular, as noted in Refs. 10, 11, and 17, the Riemann–Liouville fractional derivative has the fundamental disadvantage that the derivative of a constant is not zero so that attenuation does not vanish for a system at equilibrium.[41] The Caputo derivative has been introduced to overcome this drawback and has been discussed in detail in the context of wave propagation in lossy media in Ref. 17. In the context of effective attenuation produced by multiple scattering our analysis shows that the fractional derivative has the Caputo form given in Eq. (66).

F. Random media with short-range correlations

This is the regime in which the random medium possesses the H-short-range correlation property $H \in (0,1/2)$, and the typical wavenumber ω/c_0 of the input pulse is such that

$$\frac{|\omega| l_c}{c_0} \gg 1.$$

This second condition means that the typical wavelength is smaller than l_c and therefore the pulse probes the short-range properties of the medium. In this case we find by using Eq. (9) that

$$\frac{\gamma_c(\omega)\omega^2}{c_0^2} = \phi(0) d_H \frac{\Gamma(1+2H)}{2^{2H}} \sin(H\pi) \frac{1}{l_c} \left(\frac{|\omega| l_c}{c_0} \right)^{1-2H}, \tag{68}$$

$$\frac{\gamma_s(\omega)\omega^2}{c_0^2} = \phi(0)\frac{\omega}{c_0} - \phi(0) d_H \frac{\Gamma(1+2H)}{2^{2H}} \cos(H\pi) \times \left(\frac{|\omega| l_c}{c_0} \right)^{-2H} \frac{\omega}{c_0}. \tag{69}$$

This result is proved in Appendix B in the general case (including the fractional OU process and the binary medium) and in Appendix C for the fractional white noise (which does not fulfill a technical hypothesis used in Appendix B and requires a separate treatment). This result shows that wave propagation in random media with short-range correlations exhibits frequency-dependent attenuation that is characterized by a power law of the form of Eq. (1) with the exponent $y=1-2H$ ranging from 0 to 1. This exponent is related to the power decay rate at zero of the autocorrelation function of the medium fluctuations. Moreover the ratio of the effective dispersion coefficient over the effective diffusion coefficient is

$$R_{\mathrm{disp/diff}}(H) = -\frac{\cos(H\pi)}{\sin(H\pi)}, \tag{70}$$

which shows that the effective dispersion is stronger than the effective diffusion when H is close to 0 and that it is weaker when H is close to 1/2 (see Fig. 4). Note also that we recover the standard mixing case [formula (61)] when $H \to 1/2$.

In the time domain, we can write

$$\mathcal{L}a(\tau) = \frac{\phi(0)}{8c_0}\frac{\partial a}{\partial \tau} - \frac{\phi(0) d_H}{2^{3+2H} c_0^{1-2H} l_c^{2H}} \int_0^\infty s^{2H} \frac{\partial^2 a}{\partial \tau^2}(\tau - s) ds = \frac{\phi(0)}{8c_0}\frac{\partial a}{\partial \tau} + \frac{\phi(0) H d_H}{2^{2+2H} c_0^{1-2H} l_c^{2H}} \int_0^\infty \frac{1}{s^{1-2H}} \frac{\partial a}{\partial \tau}(\tau - s) ds. \tag{71}$$

If we go back to the original frame, neglect the random time shift, and substitute expression (71) of the pseudo-differential operator into Eq. (57), we obtain exactly the form of the integro-differential wave equation proposed by Hanyga and Rok[20] in the context of wave propagation in lossy media

$$\frac{\partial p}{\partial \xi} + \frac{1}{c_0}\frac{\partial p}{\partial \tau} = \frac{\varepsilon^2 \phi(0)}{8c_0}\frac{\partial p}{\partial \tau} + \frac{\phi(0) H d_H \varepsilon^2}{2^{2+2H} c_0^{1-2H} l_c^{2H}} \int_0^\infty \frac{1}{s^{1-2H}} \frac{\partial p}{\partial \tau}(\tau - s) ds. \tag{72}$$

If we substitute expression (71) into Eq. (58) we obtain the effective fractional wave equation

$$\frac{\partial^2 p}{\partial \xi^2} - \frac{1}{c_0^2}\left(1 + \frac{\varepsilon^2 \phi(0)}{4}\right)\frac{\partial^2 p}{\partial \tau^2} = \frac{\phi(0) H d_H \varepsilon^2}{2^{1+2H} c_0^{2-2H} l_c^{2H}} \int_0^\infty \frac{1}{s^{1-2H}} \frac{\partial^2 p}{\partial \tau^2}(\tau - s) ds. \tag{73}$$

As in the long-range correlation case this effective wave equation has the form of the fractional wave equations proposed for the modeling of wave propagation in lossy media.

IV. CONCLUSION

In this paper we have clarified the relation between the effective attenuation and dispersion of a wave propagating through a randomly scattering medium and the statistics of the random medium, in particular, its short- or long-range correlation properties. We have given explicit formulas between the power decay rate of the autocorrelation function of the random medium and the exponents of the power law frequency dependences of the effective attenuation and dispersion. This gives a practical way to measure the Hurst parameter of the medium, which is a characteristic parameter of the correlation function, from the observation of the attenuation or the dispersion of a pulsed wave propagating through the medium. The main two results are the following ones.

When a long-wavelength pulse propagates in a random medium with an autocorrelation function that decays at infinity as $|z|^{2H-2}$, $H \in (1/2,1)$, then the attenuation has a power law frequency dependence of the form $\alpha(\omega) = \alpha_0 |\omega|^y$ with $y = 3-2H \in (1,2)$.

When a short-wavelength pulse propagates in a random medium with an autocorrelation function that behaves at zero like $1 - d_H |z|^{2H}$, $H \in (0,1/2)$, then the attenuation has a power law frequency dependence of the form $\alpha(\omega) = \alpha_0 |\omega|^y$ with $y = 1-2H \in (0,1)$.

In both cases a special frequency-dependent phase is associated with the frequency-dependent attenuation and it ensures that causality and Kramers–Kronig relations are respected. Effective fractional wave equations can be written that have the form of equations studied in literature in the context of wave propagation in lossy media.[17,20,9]

These results were derived in the case of one-dimensional wave equations and the mathematical tools that have been used are restricted to this case. The generalization of the theory to three-dimensional random media remains an open question.

ACKNOWLEDGMENTS

We would like to thank Mathias Fink, Sverre Holm, and Mickaël Tanter for useful and stimulating discussions during the Cargèse Workshop 2009.

APPENDIX A: PROOF OF EQUATION (61)

Let us assume that the process is mixing and the autocorrelation function $\phi(z)$ satisfies Eq. (60). The function $\phi(z)$ can then be written as

$$\phi(z) = \phi(0)\left[\phi_{1/2}\left(\frac{d_{1/2}}{2}\frac{|z|}{l_c}\right) + \tilde{\phi}\left(\frac{|z|}{l_c}\right)\right], \tag{A1}$$

where $\phi_{1/2}$ is defined by

$$\phi_{1/2}(u) = (1-u)^2 \mathbf{1}_{[0,1]}(u). \tag{A2}$$

We also assume that ϕ is continuously differentiable and piecewise twice continuously differentiable, and that ϕ, ϕ', and ϕ'' are absolutely integrable over $(0,\infty)$. The function $\tilde{\phi}$ is continuously differentiable and piecewise twice differentiable over $(0,\infty)$, because ϕ and $\phi_{1/2}$ are continuously differentiable and piecewise twice differentiable over $(0,\infty)$. [For $\phi_{1/2}$, one can check that $\phi_{1/2}(1^-)=0=\phi_{1/2}(1^+)$, $\phi'_{1/2}(1^-)=0=\phi'_{1/2}(1^+)$, and $\phi''_{1/2}(1^-)=2$, $\phi''_{1/2}(1^+)=0$.] Moreover, $\tilde{\phi}$, $\tilde{\phi}'$, and $\tilde{\phi}''$ are absolutely integrable over $(0,\infty)$, $\tilde{\phi}(0)=0$, and $\tilde{\phi}'(0^+)=0$. By performing a double integration by parts and using Riemann–Lebesgue's lemma, we obtain

$$\int_0^\infty \tilde{\phi}(u)e^{iNu}du = -\frac{1}{N^2}\int_0^\infty \tilde{\phi}''(u)e^{iNu}du \overset{N\to\infty}{\simeq} o\left(\frac{1}{N^2}\right). \tag{A3}$$

Moreover, by performing a double integration by parts, we obtain

$$\int_0^\infty \phi_{1/2}(u)e^{iNu}du = \frac{i}{N} + \frac{2}{N^2} + \frac{2i}{N^3}(1-e^{iN}) \overset{N\to\infty}{\simeq} \frac{i}{N} + \frac{2}{N^2} + o\left(\frac{1}{N^2}\right). \tag{A4}$$

Using Eqs. (A3) and (A4) in Eq. (A1) gives formula (61).

APPENDIX B: PROOF OF EQUATIONS (68) and (69)

Let us assume that the autocorrelation function $\phi(z)$ satisfies the H-short-range property $H \in (0,1/2)$. For technical reasons we also assume that $\phi(z)$ is continuously differentiable and piecewise twice continuously differentiable on $(0,\infty)$ [note that $\phi'(z)$ blows up at 0^+; that $\phi(z)$, $\phi'(z)$, and $\phi''(z)$ are absolutely integrable at infinity, say, on (l_c,∞); and that $\phi''(z)+2H(2H-1)\phi(0)d_H|z/l_c|^{2H-2}/l_c^2$ is absolutely integrable at zero, say, on $(0,l_c]$. The fractional OU process and the binary medium satisfy these technical hypotheses, but not the fractional white noise, which we will consider separately in Appendix C. The function $\phi(z)$ can then be written as

$$\phi(z) = \phi(0)\left[\phi_H\left([d_H(1-2H)]^{1/(2H)}\frac{|z|}{l_c}\right) + \tilde{\phi}\left(\frac{|z|}{l_c}\right)\right], \tag{B1}$$

where ϕ_H is defined by

$$\phi_H(u) = \left(1 - \frac{1}{1-2H}u^{2H} + \frac{2H}{1-2H}u\right)\mathbf{1}_{[0,1]}(u). \tag{B2}$$

The function $\tilde{\phi}$ is continuously differentiable and piecewise twice differentiable over $(0,\infty)$, because ϕ and ϕ_H are continuously differentiable and piecewise twice differentiable over $(0,\infty)$ [for ϕ_H, one can check that $\phi_H(1^-)=0=\phi_H(1^+)$, $\phi'_H(1^-)=0=\phi'_H(1^+)$, and $\phi''_H(1^-)=2H$, $\phi''_H(1^+)=0$]. Moreover, $\tilde{\phi}$, $\tilde{\phi}'$, and $\tilde{\phi}''$ are absolutely integrable over $(0,\infty)$ and $\tilde{\phi}(0)=0$, and $\tilde{\phi}'(0^+)$ is well defined (since $\tilde{\phi}''$ is absolutely integrable at 0). Consequently a double integration by parts yields

$$\int_0^\infty \tilde{\phi}(u)e^{iNu}du = -\frac{\tilde{\phi}'(0^+)}{N^2} - \frac{1}{N^2}\int_0^\infty \tilde{\phi}''(u)e^{iNu}du \overset{N\to\infty}{=} O\left(\frac{1}{N^2}\right). \tag{B3}$$

Moreover, by performing an integration by parts step, we obtain

$$\int_0^\infty \phi_H(u)e^{iNu}du \overset{N\to\infty}{\simeq} \frac{i}{N} - i\frac{2H}{1-2H}\frac{1}{N^{1+2H}}\int_0^\infty v^{2H-1}e^{iv}dv + o\left(\frac{1}{N^{1+2H}}\right). \tag{B4}$$

The computation of the definite integral $\int_0^\infty v^{2H-1}e^{iv}dv = \Gamma(2H)e^{iH\pi}$ (see formula 3.761 in Ref. 39) gives

$$\int_0^\infty \phi_H(u)e^{iNu}du \overset{N\to\infty}{\simeq} \frac{i}{N} - \frac{i}{N^{1+2H}}\frac{\Gamma(1+2H)}{1-2H}\cos(H\pi) + \frac{1}{N^{1+2H}}\frac{\Gamma(1+2H)}{1-2H}\sin(H\pi) + o\left(\frac{1}{N^{1+2H}}\right). \tag{B5}$$

Using Eqs. (B3) and (B5) in Eq. (B1) gives Eqs. (68) and (69).

APPENDIX C: THE FRACTIONAL WHITE NOISE WITH $H \in (0,1/2)$

As we have already noticed the fractional white noise with the Hurst index $H \in (0,1/2)$ satisfies the H-short-range property, but fails to meet one of the technical hypotheses listed in Appendix B. Indeed its autocorrelation function $\phi(z)$ is not differentiable at $z=l_c$. However, we show in this appendix that this model has "almost" the effective attenuation and dispersion predicted in the general case in Appendix B. Here $\phi(z)$ is explicit so we can carry out the computation and we find that, if $|\omega|l_c/c_0 \gg 1$, then

$$\frac{\gamma_c(\omega)\omega^2}{c_0^2} = \sigma^2 \frac{\Gamma(1+2H)}{2^{2H}} \sin(H\pi)\left[1-\cos\left(\frac{2\omega l_c}{c_0}\right)\right] \times \frac{1}{l_c}\left(\frac{|\omega| l_c}{c_0}\right)^{1-2H}, \tag{C1}$$

$$\frac{\gamma_s(\omega)\omega^2}{c_0^2} = \sigma^2 \frac{\omega}{c_0} - \sigma^2 \frac{\Gamma(1+2H)}{2^{2H}} \times \left[\cos(H\pi) + \sin(H\pi)\sin\left(\frac{2|\omega| l_c}{c_0}\right)\right] \times \left(\frac{|\omega| l_c}{c_0}\right)^{-2H} \frac{\omega}{c_0}. \tag{C2}$$

The singularity of $\phi'(z)$ at l_c is responsible for the additional terms that have a fast periodic modulation in ω. In this paper we are interested in pulse propagation, and as shown by Eq. (40) we integrate over ω. Using the identity (formula 3.936 in Ref. 39)

$$\frac{1}{2\pi}\int_0^{2\pi} \exp(x\cos(\theta) + ix\sin(\theta))d\theta = 1, \quad \forall x \in \mathbb{R}, \tag{C3}$$

we find that the additional modulated terms vanish when expressions (C1) and (C2) are substituted into Eq. (40). Therefore, we can consider that the effective attenuation and dispersion terms for the fractional white noise model are given by general formulas (68) and (69).

[1] M. Asch, W. Kohler, G. Papanicolaou, M. Postel, and B. White, "Frequency content of randomly scattered signals," SIAM Rev. **33**, 519–626 (1991).
[2] J.-P. Fouque, J. Garnier, G. Papanicolaou, and K. Sølna, *Wave Propagation and Time Reversal in Randomly Layered Media* (Springer, New York, 2007).
[3] A. Ishimaru, *Wave Propagation and Scattering in Random Media* (Academic, New York, 1978).
[4] A. Aubry, A. Derode, and F. Padilla, "Local measurements of the diffusion constant in multiple scattering media: Application to human trabecular bone imaging," Appl. Phys. Lett. **92**, 124101 (2008).
[5] S. Dolan, C. Bean, and B. Riollet, "The broad-band fractal nature of heterogeneity in the upper crust from petrophysical logs," Geophys. J. Int. **132**, 489–507 (1998).
[6] M. Sahimi and S. E. Tajer, Phys. Rev. E **71**, 046301 (2005).
[7] S. M. V. Allaei, M. Sahimi, and M. R. R. Tabar, "Propagation of acoustic waves as a probe for distinguishing heterogeneous media with short-range and long-range correlations," J. Stat. Mech.: Theory Exp. **2008**, P03016/1–P03016/28.
[8] D. T. Blackstock, "Generalized Burgers equations for plane waves," J. Acoust. Soc. Am. **77**, 2050–2053 (1985).
[9] T. L. Szabo, "Time domain wave equations for lossy media obeying a frequency power law," J. Acoust. Soc. Am. **96**, 491–500 (1994).
[10] M. Caputo, "Linear models of dissipation whose Q is almost frequency independent II," Geophys. J. R. Astron. Soc. **1**, 529–539 (1967).
[11] M. Caputo and F. Mainardi, "A new dissipation model based on memory mechanism," Pure Appl. Geophys. **91**, 134–147 (1971).
[12] T. L. Szabo and J. R. Wu, "A model for longitudinal and shear wave propagation in viscoelastic media," J. Acoust. Soc. Am. **107**, 2437–2446 (2000).
[13] K. Aki and P. G. Richards, *Quantitative Seismology* (Freeman, San Francisco, CA, 1980), Vol. **1**, Chap. 5.
[14] T. E. Gómez Álvarez-Arenas, F. R. Montero de Espinosa, M. Moner-Girona, E. Rodriguez, A. Roig, and E. Molins, "Viscoelasticity of silica aerogels at ultrasonic frequencies," Appl. Phys. Lett. **81**, 1198–1200 (2002).
[15] *Physical Principles of Medical Ultrasonics*, edited by C. R. Hill, J. C. Bamber, and G. R. ter Haar (Wiley, Chichester, 1986), Chaps. 4 and 5.
[16] A. C. Kibblewhite, "Attenuation of sound in marine sediments: A review with emphasis on new low-frequency data," J. Acoust. Soc. Am. **86**, 716–738 (1989).
[17] W. Chen and S. Holm, "Modified Szabo's wave equation models for lossy media obeying frequency power law," J. Acoust. Soc. Am. **114**, 2570–2574 (2003).
[18] W. Chen and S. Holm, "Fractional Laplacian time-space models for linear and nonlinear lossy media exhibiting arbitrary frequency power-law dependency," J. Acoust. Soc. Am. **115**, 1424–1430 (2004).
[19] S. Gelinsky, S. A. Shapiro, and T. Müller, "Dynamic poroelasticity of thinly layered structure," Int. J. Solids Struct. **35**, 4739–4751 (1998).
[20] A. Hanyga and V. E. Rok, "Wave propagation in micro-heterogeneous porous media: A model based on an integro-differential wave equation," J. Acoust. Soc. Am. **107**, 2965–2972 (2000).
[21] N. V. Sushilov and R. S. C. Cobbold, "Frequency domain wave equation and its time-domain solutions in attenuating media," J. Acoust. Soc. Am. **115**, 1431–1435 (2004).
[22] M. G. Wismer, "Finite element analysis of broadband acoustic pulses through inhomogeneous media with power law attenuation," J. Acoust. Soc. Am. **120**, 3493–3502 (2006).
[23] R. Marty and K. Sølna, "Acoustic waves in long-range random media," SIAM J. Appl. Math. **69**, 1065–1083 (2009).
[24] K. Sølna, "Acoustic pulse spreading in a random fractal," SIAM J. Appl. Math. **63**, 1764–1788 (2003).
[25] R. Burridge and H. W. Chang, "Multimode one-dimensional wave propagation in a highly discontinuous medium," Wave Motion **11**, 231–249 (1989).
[26] R. Burridge, G. Papanicolaou, and B. White, "One-dimensional wave propagation in a highly discontinuous medium," Wave Motion **10**, 19–44 (1988).
[27] J.-F. Clouet and J.-P. Fouque, "Spreading of a pulse traveling in random media," Ann. Appl. Probab. **4**, 1083–1097 (1994).
[28] A. Nachbin and K. Sølna, "Apparent diffusion due to topographic microstructure in shallow waters," Phys. Fluids **15**, 66–77 (2003).
[29] K. Sølna and G. Papanicolaou, "Ray theory for a locally layered medium," Waves Random Media **10**, 151–198 (2000).
[30] A. Fannjiang and K. Sølna, "Scaling limits for wave beams in atmospheric turbulence," Stochastics Dyn. **4**, 135–151 (2004).
[31] A. E. Gargett, "The scaling of turbulence in the presence of stable stratification," J. Geophys. Res. **93**, 5021–5036 (1988).
[32] C. Sidi and F. Dalaudier, "Turbulence in the stratified atmosphere: Recent theoretical developments and experimental results," Adv. Space Res. **10**, 25–36 (1990).
[33] J.-M. Bardet, G. Lang, G. Oppenheim, A. Philippe, and M. S. Taqqu, in *Theory and Applications of Long-Range Dependence*, edited by P. Doukhan, G. Oppenheim, and M. S. Taqqu (Birkhauser, Boston, 2003), pp. 579–624.
[34] W. Feller, *An Introduction to Probability Theory and Its Applications* (Wiley, New York, 1971), Vol. **2**, Chap. 11.
[35] J. Garnier and K. Sølna, "Pulse propagation in random media with long-range correlation," Multiscale Model. Simul. **7**, 1302–1324 (2009).
[36] R. F. O'Doherty and N. A. Anstey, "Reflections on amplitudes," Geophys. Prospect. **19**, 430–458 (1971).
[37] H. A. Kramers, "La diffusion de la lumière par les atomes (Diffusion of the light by the atoms)," Atti. Congr. Intern. Fisica Como **2**, 545–557 (1927).
[38] R. de L. Kronig, "On the theory of the dispersion of x-rays," J. Opt. Soc. Am. **12**, 547–557 (1926).
[39] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products* (Academic, San Diego, CA, 1980).
[40] S. G. Samko, A. A. Kilbas, and O. I. Marichev, *Fractional Integrals and Derivatives: Theory and Applications* (Gordon and Breach Science, New York, 1987).
[41] M. Seredynska and A. Hanyga, "Nonlinear Hamiltonian equations with fractional damping," J. Math. Phys. **41**, 2135–2156 (2000).

Broadband source localization using horizontal-beam acoustic intensity striations

Altan Turgut[a] and Marshall Orr
Acoustics Division, Naval Research Laboratory, Washington, DC 20375

Daniel Rouseff
Applied Physics Laboratory, College of Ocean and Fishery Sciences, University of Washington, Seattle, Washington 98105

(Received 1 December 2008; revised 20 September 2009; accepted 29 September 2009)

Waveguide invariant theory is applied to horizontal line array (HLA) beamformer output to localize moving broadband noise sources from measured acoustic intensity striation patterns. Acoustic signals emitted by ships of opportunity (merchant ships) were simultaneously recorded on a HLA and three hydrophones separated by 10 km during the RAGS03 (relationship between array gain and shelf-break fluid processes) experiment. Hough transforms are used to estimate both the waveguide invariant parameter "beta" and the ratio of source range at the closest point of approach to source speed from the observed striation patterns. Broadband (50–150-Hz) acoustic data-sets are used to demonstrate source localization capability as well as inversion capability of waveguide invariant parameter beta. Special attention is paid to bathymetric variability since the acoustic intensity striation patterns seem to be influenced by range-dependent bathymetry of the experimental area. The Hough transform method is also applied to the HLA beam-time record data and to the acoustic intensity data from three distant receivers to validate the estimation results from HLA beamformer output. Good agreement of the results from all three approaches suggests the feasibility of locating broadband noise sources and estimating waveguide invariant parameter beta in shallow waters. © *2010 Acoustical Society of America.* [DOI: 10.1121/1.3257211]

PACS number(s): 43.30.Bp, 43.30.Wi, 43.30.Zk, 43.60.Lq [RAS] Pages: 73–83

I. INTRODUCTION

In a shallow-water waveguide, when plotted in time, pressure spectra of acoustic signals generated by a moving broadband source often exhibit sloped striation patterns due to constructive and destructive interferences of propagating acoustic modes. Chuprov[1] derived a simple formula for the so-called *waveguide invariant* that characterizes the dispersive nature of the waveguide and relates the observed slope of the striations to source range and frequency. The waveguide invariant, usually designated as beta, has the canonical value of 1.0 for an isovelocity waveguide with a perfectly reflecting bottom. However, for an arbitrary sound-speed profile, different combinations of modes at different frequencies will provide slightly different values of beta. The value of beta might also be range-dependent if the waveguide has a range-dependent bathymetry or sound-speed profile.

An attractive feature of the waveguide invariant concept is its simplicity. The effect can be observed with incoherent processing of acoustic intensity without any information about the phase. D'Spain and Kuperman[2] used the waveguide invariant to interpret observed single hydrophone spectrograms in shallow water environments that vary in range and azimuth. Heaney and Cox[3] proposed an incoherent method for geoacoustic characterization. Several investigators have used the waveguide invariant when applying coherent processing to vertical array data. Thode[4] used a vertical receiving array and a guide source to construct a virtual receiver. Hodgkiss *et al.*[5] exploited the concept to move the focal range in a time-reversal experiment. Thode *et al.*[6] examined the sidelobe structure of a matched field processor. Yang[7] used the waveguide invariant theory for source motion compensation.

Striation patterns can also be observed in the output of a coherent processor such as a horizontal line array (HLA) beamformer.[8,9] The analysis of HLA beamformer data provides an opportunity to identify individual striation patterns when multiple striation patterns occur due to multiple sources at different azimuths. Also, array gain in the beamformer output provides higher intensity levels so that the desired dynamic range of striation patterns can be still be observed for low signal to noise ratios (SNRs).

In the present paper, the waveguide invariant is used to study the output of a coherent processor that uses a bottom-mounted horizontal array. We show how the beamformer output, displayed as a low-frequency analysis and recording (LOFARgram), exhibits striations, the trajectories of which can be described using the waveguide invariant. The theory is developed in Sec. II with analytical results compared to numerical simulations. In Sec. III, the waveguide invariant beta and the ratio of range at closest point of approach (CPA) to source speed is estimated from the trajectories of striations by using the Hough transform[10] method. In Sec. IV, the RAGS03 New Jersey Shelf experiment is briefly described, and localization results of several merchant ships and the R/V Oceanus are presented. Summary and discussion are given in Sec. V.

[a]Author to whom correspondence should be addressed. Electronic mail: turgut@wave.nrl.navy.mil

II. THEORY

In this section, we outline various processing schemes in which the waveguide invariant arises. We first consider the simplest case where the observable is the acoustic intensity from a distant source measured as a function of range and frequency in a range-independent waveguide. We derive a specific form of the waveguide invariant that makes explicit its dependence on the propagating acoustic modes. For later use on data from the RAGS03 experiment, we then outline a processing scheme that treats the waveguide invariant as a range-dependent parameter. We then show how the same waveguide invariant arises when performing coherent processing of horizontal array data.

A. Incoherent processing in a range-independent waveguide

Acoustic intensity, when mapped versus range and frequency, often exhibits striation patterns that can be represented by level or contour curves. To analyze the properties of the level curves, let I be the observed intensity as a function of range r from the source and frequency ω. The level curves satisfy

$$\frac{\partial I}{\partial \omega} d\omega + \frac{\partial I}{\partial r} dr = 0. \tag{1}$$

Solving for the slope $d\omega/dr$ of the striations yields

$$\frac{d\omega}{dr} = -\frac{\partial I/\partial r}{\partial I/\partial \omega}. \tag{2}$$

To analyze Eq. (2), it is convenient to use a normal mode representation for the intensity. For source and receiver depths z_s and z, respectively, the associated pressure p is

$$p = \sum_m (\xi_m r)^{-1/2} \Psi_m(z_s) \Psi_m(z) e^{i(\xi_m + i\alpha_m)r}, \tag{3}$$

where Ψ_m are the mode functions and ξ_m the corresponding horizontal wavenumbers. The modal attenuation α_m accounts for bottom loss. Certain unimportant scaling terms have been suppressed. It follows that the intensity, $I=|p|^2$, is

$$I = \sum_{m,n} A_{mn} \exp(i\Delta\xi_{mn} r). \tag{4}$$

The amplitudes A_{mn} can be derived from Eq. (3). Since the striation patterns are usually determined by rapidly oscillating phase terms, the weak range and frequency dependence in A_{mn} is neglected, and the derivatives are calculated as

$$\frac{\partial I}{\partial r} \approx i\omega \sum_{m,n} A_{mn} \Delta Q_{mn} \exp(i\Delta\xi_{mn} r) \tag{5a}$$

and

$$\frac{\partial I}{\partial \omega} \approx ir \sum_{m,n} A_{mn} \Delta S_{mn} \exp(i\Delta\xi_{mn} r), \tag{5b}$$

where $\Delta Q_{mn} = (1/\nu_m - 1/\nu_n)$ is the mode phase slowness difference with modal phase velocity $\nu = \omega/\xi$, and $\Delta S_{mn} = (1/u_m - 1/u_n)$ is the mode group slowness difference with modal group velocity $u = (\partial\xi/\partial\omega)^{-1}$.

Brekhovskikh and Lysanov[11] gave a definition for the waveguide invariant as a derivative of phase slowness with respect to group slowness. They cautioned, however, that this expression applies only to a group of modes, not necessarily all the propagating modes. Making this dependence on mode index explicit, we define the local invariant β for a particular pair of modes as

$$\beta = -\frac{\Delta Q_{mn}}{\Delta S_{mn}} = -\frac{\Delta\xi_{mn}/\omega}{\partial\Delta\xi_{mn}/\partial\omega}. \tag{6}$$

Equation (6) can be used in Eq. (5a) to eliminate the term related to the phase velocities. Combining Eqs. (2)–(4), (5a), (5b), and (6) yields for the slope of the striations

$$\frac{d\omega}{dr} = \beta \frac{\omega}{r}. \tag{7}$$

The waveguide invariant β is assumed to be independent of the mode indices m and n. For a Pekeris waveguide, the assumption is justified and $\beta \approx 1$, the canonical value for shallow water. For more complicated scenarios, as when there is a significant thermocline in the water column, the numerical values for the various β in Eq. (7) can differ by an order of magnitude and might depend on the source and receiver depths.[12] In addition to depth-dependence, β might also be range and azimuth dependent if there is significant range and azimuth variability in the waveguide. Focusing on the bathymetric variability, the range and azimuth dependency of β is discussed next.

B. Incoherent processing in a waveguide with range and azimuth dependency

In this section, properties of acoustic striations are analyzed for a weakly range-dependent waveguide. First, a brief derivation of level curves of intensity (interference structure) is given for a single receiver in a waveguide with range and azimuth dependency. Detailed analysis of acoustic intensity striations for a range-dependent case can be found in Ref. 13 and for a range- and azimuth-dependent case in Ref. 2. In a weakly range-dependent waveguide with adiabatic mode propagation, acoustic intensity in Eq. (4) can be modified as

$$I = \sum_{m,n} A_{mn} \exp\left(i \int \Delta\xi_{mn} dr \right), \tag{8}$$

where wavenumber differences $\Delta\xi_{mn}$ are both range and azimuth dependent. In this case, the condition along the acoustic intensity level curves is

$$\delta I = \frac{\partial I}{\partial \omega}\delta\omega + \frac{\partial I}{\partial r}\delta r + \frac{\partial I}{\partial \theta}\delta\theta = 0. \tag{9}$$

Ignoring range and azimuth dependence of mode amplitudes in Eq. (8), Eq. (9) can be expressed as[2]

$$\left(\int \frac{\Delta\xi_{mn}}{\partial\omega} dr \right) \delta\omega + \Delta\xi_{mn}^R \delta r + \frac{\partial}{\partial\theta}\left(\int \Delta\xi_{mn} dr \right) \delta\theta = 0, \tag{10}$$

where $\Delta\xi_{mn}^R$ are the modal wavenumber differences at the receiver location. Assuming adiabatic mode propagation in a

weakly range-dependent waveguide, Chuprov[1] derived a range-dependent waveguide invariant $\beta(r)$ as

$$\beta(r) = -\frac{\frac{\Delta\xi_{mn}^R}{\omega}}{\frac{1}{r}\int \frac{\partial \Delta\xi_{mn}}{\partial\omega} dr} = -\frac{\Delta Q_{mn}^R}{\Delta \bar{S}_{mn}}, \tag{11}$$

where $\Delta\bar{S}_{mn}=(1/r)\int \Delta S_{mn}(x)dx$ is the average modal group slowness difference over the waveguide and ΔQ_{mn}^R is the modal phase slowness difference at the receiver location. Using Eq. (11) and including the azimuthal dependency of the waveguide invariant, Eq. (10) can be expressed as

$$r\Delta\bar{S}_{mn}\delta\omega + \omega\Delta Q_{mn}^R\delta r + \left(-r\int \frac{\Delta\xi_{mn}^R}{\omega}\frac{\partial}{\partial\theta}(1/\beta(r,\theta))d\omega\right)\delta\theta = 0. \tag{12}$$

The frequency dependency of the modal wavenumber differences at the receiver can be obtained as $\Delta\xi_{mn}^R(\omega)=C\omega^{-1/\beta^R}$ by integrating Eq. (6) in frequency, where C is the integration constant. Using this frequency dependency in the last term of Eq. (12), we obtain

$$r\Delta\bar{S}_{mn}\delta\omega + \omega\Delta Q_{mn}^R\delta r + r\omega\beta^R\Delta Q_{mn}^R\frac{\partial}{\partial\theta}(1/\beta(r,\theta))\delta\theta = 0. \tag{13}$$

Using the property given in Eq. (11), Eq. (13) can be rewritten in terms of the range- and azimuth-dependent waveguide invariant as

$$\frac{\delta\omega}{\omega} - \beta(r,\theta)\frac{\delta r}{r} + \beta^R\beta(r,\theta)\frac{\partial}{\partial\theta}(1/\beta(r,\theta))\delta\theta = 0. \tag{14}$$

Integration of Eq. (14) yields

$$\omega(t) = \omega_0\left(\frac{\beta(r,\theta)}{\beta(r_0,\theta_0)}\right)^{-\beta^R}\exp\left(\int_0^{(t)}\frac{\beta(r,\theta)}{r}dr\right), \tag{15}$$

which describes a time-frequency dependence of level curves with an integration constant ω_0 as a broadband source moves through a range- and azimuth-dependent waveguide. If we assume that the broadband source moves through a range-independent environment over the range interval r_0 and $r(t)$, Eq. (15) can be approximated as

$$\omega = \omega_0\left(\frac{r}{r_0}\right)^{\beta}\left(\frac{\beta}{\beta_0}\right)^{-\beta^R}. \tag{16}$$

Considering a broadband source moving along a straight track with a constant velocity V, the range to the receiver can be written as

$$r = \sqrt{r_0^2 + V^2(t-t_0)^2}, \tag{17}$$

where r_0 is the range at the CPA and t_0 is the time when the source is at the CPA. Then, Eq. (16) for this case can be written as

$$\omega = \omega_0(1+(V/r_0)^2(t-t_0)^2)^{\beta/2}(\beta/\beta_0)^{-\beta^R}. \tag{18}$$

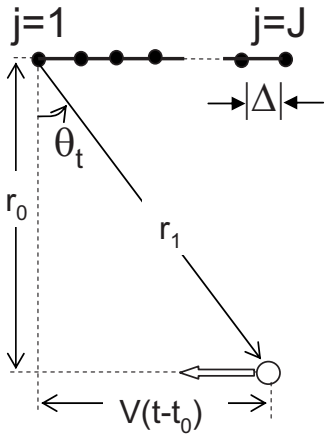


FIG. 1. Geometry of a moving source and a fixed horizontal receiving array considered in the coherent processing.

The above equation is in Sec. III to estimate range-dependent β as well as the ratio of source speed to range at the CPA V/r_0 and time at the CPA t_0.

C. Coherent processing

The analysis to this point has considered only incoherent processing with the observable being the acoustic intensity. The striation patterns can also be observed in the output of a coherent processor such as a beamformer. Consider a simple beamformer and let p_j be the acoustic pressure at a particular frequency as observed at the jth sensor in an array, and let $b_j(\theta)$ be the corresponding component of the beamformer steering vector for the look direction θ:

$$B(\theta) = \left|\sum_j b_j(\theta)p_j\right|^2. \tag{19}$$

For passive sonar systems, the beamformed data $B(\theta)$ are often displayed as a LOFARgram. A LOFARgram plots the temporal evolution of the beamformer's spectrum for assorted look directions. Striations in a LOFARgram can be discerned using the same technique as for the incoherent processor.

As an example illustrating LOFARgram striations, consider the geometry shown in Fig. 1. A source moves at constant velocity V along a track parallel to a fixed horizontal receiving array. Let θ be the time-dependent bearing of the source referenced to the array broadside. Assume that the array has J sensors equally spaced Δ apart, and the range from the source to the reference $j=1$ element is r_1. In the far field, the range to the jth array sensor is

$$r_j \approx r_1 - (j-1)\Delta\sin\theta_t. \tag{20}$$

Using Eq. (20) in the modal expansion for pressure [Eq. (3)] yields

$$p_j \approx \sum_m \{(\xi_m r_1)^{-1/2}\Psi_m(z_s)\Psi_m(z)e^{-i(\xi_m+i\alpha_m)r_1}\} \times e^{-i\xi_m(j-1)\Delta\sin\theta_t}. \tag{21}$$

For a conventional beamformer

$$b_j(\theta) = \exp[ik_0(j-1)\Delta\sin\theta], \tag{22}$$

where $k_0=\omega/c_0$ and c_0 is the reference sound speed. Combining Eqs. (20)–(22) in Eq. (19) and rearranging the order of summations yield

$$B(\theta)=\sum_{m,n} d_m^* A_{mn} d_n \exp(i\Delta\xi_{mn} r_1), \tag{23}$$

where

$$d_n=\sum_j \exp[i(j-1)\Delta(\xi_n \sin\theta_t - k_0 \sin\theta)]. \tag{24}$$

The modal weighting A_{mn} evaluated at $r=r_1$ is the same as in Eq. (4). Equation (4) for incoherent processing and Eq. (23) for coherent processing differ primarily by the presence of terms describing the interference between the steering vector and the propagating acoustic modes, d_m and d_n. For the special case of the beamformer tracking the source, $\theta=\theta_t$, the results for incoherent processing can be exploited. When $\theta=\theta_t$ and $k_0\approx\xi_n$ in Eq. (24), each term in the summation can be approximated by unity and $d_n\approx J$. For this special case,

$$B(\theta_t)\approx J^2 I. \tag{25}$$

The above approximate relation indicates that striation patterns would be similar for both incoherent processing and coherent processing at signal look direction. However, the beamformer output provides additional array gain at a signal look direction so that localization of multiple sources at longer ranges and different azimuths is possible. Then, Eq. (18) should be also valid for the beamformer output LOFARgram striations. For a range-independent waveguide Eq. (18) reduces to

$$\omega=\omega_0[1+(V/r_0)^2(t-t_0)^2]^{\beta/2}. \tag{26}$$

If the term in the square brackets is close to unity, Eq. (26) can be expanded yielding to first order

$$\omega-\omega_0=\frac{\omega_0}{2}\beta(V/r_0)^2(t-t_0)^2. \tag{27}$$

The above equation describes a family of nested parabolas, each at a minimum at the time of closest approach, $t=t_0$. Note that the width of the parabolas is related to beta. If the second term in the square bracket is large ($(V/r_0)|t-t_0|\gg 1$), Eq. (26) can be approximated as

$$\omega=\omega_0[(V/r_0)|t-t_0|]^{\beta}, \tag{28}$$

which describes linear lines when $\beta\sim 1$, crossing zero frequency ($\omega=0$) at $t=t_0$.

Figure 2 shows three simulated LOFARgrams compared to the trajectories defined by Eq. (26). Range-independent environmental parameters were used in the calculations simulating the RAGS03 experimental site. The sound-speed profile represents a well-mixed winter profile with sound speeds of 1496 m/s at the surface and 1502 m/s at the bottom. The bottom parameters are inferred from previous acoustic probe[14] and chirp sonar measurements.[15] The water is 90 m deep above an acoustically fast seabed (1650 m/s) with moderate loss (0.1 dB/wavelength). The scenario is consistent with Fig. 1 with $V=8$ m/s and the CPA range 4.9 km except that the source track is not parallel to the fixed

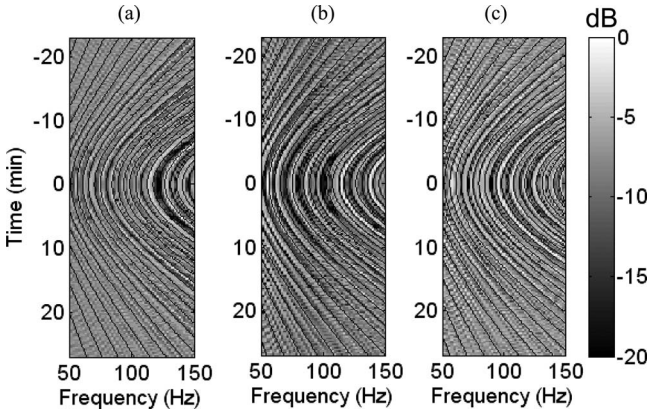


FIG. 2. Simulated LOFARgrams for range-independent bathymetry compared to theoretical predictions made using Eq. (26) for different depths of source and receiving array: (a) source depth 7 m and receiver depth 89 m, (b) source depth 45 m and receiver depth 45 m, and (c) source depth 45 m and receiver depth 89 m.

HLA ($\theta_{t_0}=-10$ deg). The $J=32$ element bottom-mounted receiving array has a spacing of $\Delta=15$ m. The pressure was calculated using a normal mode solver FEMODE (Ref. 16) and used as input to the beamformer in Eq. (19). The calculation was repeated at ~1.9-Hz intervals. The beamformer output, evaluated at $\theta=\theta_t$, is shown over a time interval bracketing the CPA time $t_0=0$. The Hough transform method, described in Sec. III is used to estimate the two parameters (β and r_0/V) in Eq. (26) that provided best fit to the simulation results. In Fig. 2(a), the source depth is 7 m and the HLA depth is 89 m. The estimated parameters $\beta=1.05$ and $r_0/V=625$ provide the best fit between the simulated striation patterns and the estimated trajectories (solid lines). In Fig. 2(b), the source depth is 45 m, the HLA depth is 45 m, and the best fit is obtained using $\beta=1.1$ and $r_0/V=625$ s. In Fig. 2(c), the source is at depth 45 m, the HLA depth is 89 m, and the best fit is obtained using $\beta=1.15$ and $r_0/V=625$ s. Figures 2(a)–2(c) indicate weak source-depth-dependence of β for the RAGS03 experimental site. Previous experimental[5] and modeling[17] results have also suggested some depth-dependence on the numerical value for beta especially in the presence of a strong thermocline in the water column.

III. PARAMETER ESTIMATION FROM LOFARgrams USING GENERALIZED RADON TRANSFORMS

Radon transforms provide a mapping from an image domain to a parameter domain, where the parameters characterize the curves to be detected. As a special case of generalized Radon transforms (GRTs), the Hough transform method[10] has been extensively used in the computer image processing community for pattern recognition as well as object detection.[18] A difficult global detection problem in the image domain is converted into a simpler local peak detection problem by means of the Hough transform. In Eq. (26), the parameters β, r_0/V, ω_0, and t_0 describe the curves to be detected in LOFARgrams. Each point $(\beta, r_0/V, \omega_0, t_0)$ in the parameter space is mapped to the integral of intensity variation in the LOFARgrams along the curves described in Eq. (26),

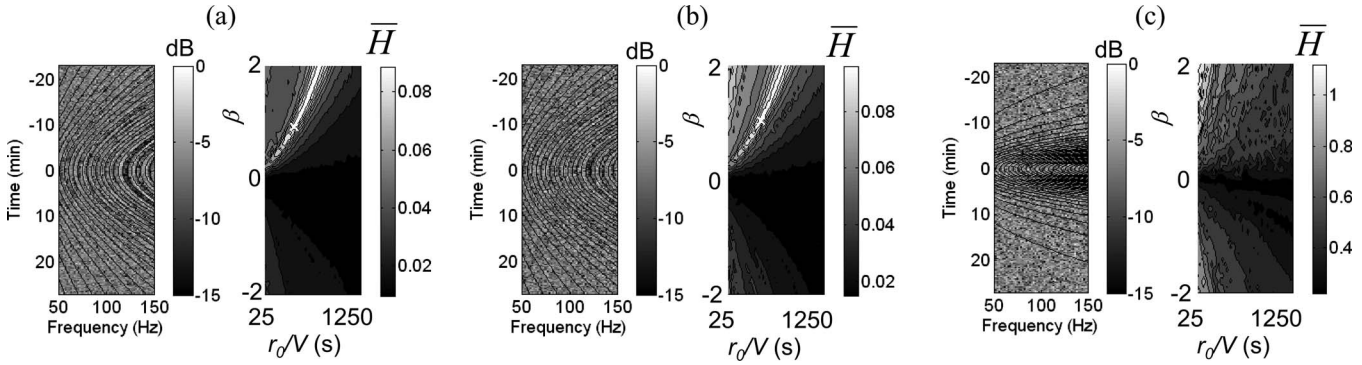


FIG. 3. Simulated LOFARgrams with different SNRs and calculated total variances in parameter space. Estimated trajectories corresponding to the parameters with maximum total variance are also shown on the LOFARgrams (solid lines): (a) SNR=0 dB (same as [Fig. 2(a)]), (b) SNR=−6 dB, and (c) SNR=−32 dB.

$$H(\omega_0,t_0,\beta,r_0/V)=\int L(\omega,t)ds, \tag{29}$$

where $L(\omega,t)$ is the intensity variation $(L(\omega,t)=I(\omega,t)-I_0(t))$ at point (ω,t) in the LOFARgram and ds is the incremental length along the curves described in Eq. (26). Considering the slowly varying nature of the intensity as a function of range (or time), a time-dependent (only frequency-averaged) mean intensity $I_0(t)$ is used to remove the overall trend of transmission loss as a function of range (or time). In Eq. (29), $H(\omega_0,t_0,\beta,r_0/V)$ will be a sinusoidal function of ω_0 if the parameters β, r_0/V, and t_0 represent an interference pattern in the LOFARgram. The variance (energy) of this sinusoidal function determines how well the interference patterns are represented by a family of curves described by parameters β, r_0/V, and t_0 in Eq. (26). The total intensity along each curve is squared and normalized by the curve length, and total variance is obtained by summation over ω_0 (representing a family of curves) as

$$\bar{H}(t_0,\beta,r_0/V)=\sum_{\omega_0}\frac{(\int L(\omega,t)ds)^2}{\int ds}. \tag{30}$$

Note that the calculated total variance in three-dimensional parameter space is expected to have a global maximum when the interference pattern is best represented by Eq. (26) with the inverted values of parameters β, r_0/V, and t_0. In Fig. 3(a) the total variance is calculated for simulated LOFARgrams obtained from a bottom-mounted HLA beamformer output given in Fig. 2(a). Estimation of t_0 can also be included in a search for a global maximum in three-dimensional parameter space, but omitted here for practical purposes. In Figs. 3(b) and 3(c), Gaussian noise is added to simulated LOFARgrams, yielding SNRs of −6 dB for the HLA beamformer output [Fig. 3(b)] and −32 dB for the single hydrophone (mid phone) of the HLA [Fig. 3(c)]. The corresponding total variances of intensity calculated in two-dimensional parameter space are also shown in Figs. 3(b) and 3(c). The values of the two parameters corresponding to the maximum total variance are $\beta=1.05$, $r_0/V=625$ s for the bottom-mounted array and erroneous values of $\beta=1.95$, $r_0/V=204$ s for a single receiver. Note that the maximum total variance provides accurate parameter estimation for −6 dB SNR (bottom-mounted array case) and erroneous parameter estimation for −32 dB SNR ratio (single-receiver case). Also, a distinct curve of local maxima around the global maximum indicates that the two estimated parameters are correlated, and finding the location of true maximum might be challenging for noisy data.

When acoustic signals are emitted by two different noise sources and received by a single receiver, the GRT analysis is expected to provide two maxima with different r_0/V and t_0 values. This case is demonstrated in Fig. 4, showing the LOFARgram [Fig. 4(a)] and calculated total variance in two-dimensional parameter space [Fig. 4(b)]. The values of three parameters corresponding to the maximum total variance are $\beta=1.05$, $r_0/V=435$ s, and $t_0=0$. Note that the second peak at $\beta=1.05$, $r_0/V=222$ s, and $t_0=-400$ s corresponds to the second noise source.

Adiabatic propagation of 100–150-Hz broadband signals over a sloping bottom is also simulated to demonstrate the effects of bathymetry on the acoustic striation patterns. As in calculating Figs. 2–4, all the other parameters are assumed to be range-independent simulating the RAGS03 experimental site. Figure 5 shows the HLA beamformer output and calculated total variance for a 3-min moving window to depict the asymmetric nature of intensity level curves and range-dependency of waveguide invariant parameter beta. Range at CPA to source speed ratio of $r_0/V=300$ s and bottom slopes of $\alpha=0.000\,17$ [Fig. 5(a)] and $\alpha=0.0035$ [Fig. 5(b)] are used in the numerical simulations. Assuming range-independent bathymetry between the receiver and source at

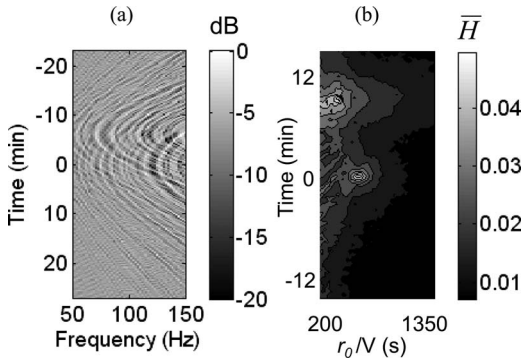


FIG. 4. (a) Simulated LOFARgram for two moving source and (b) calculated total variances assuming that β is known *a priori*.

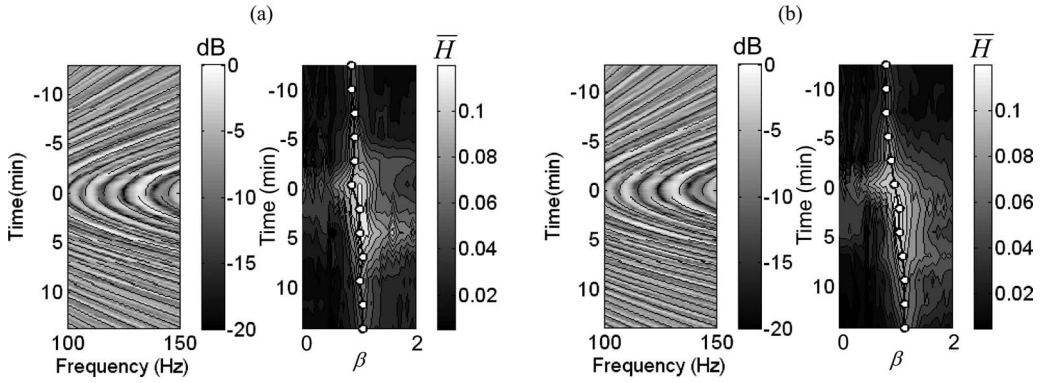


FIG. 5. Simulated LOFARgrams with different bottom slope and calculated total variances in parameter space. Estimated trajectories for 3-min time window corresponding to $\beta(r)$ value with maximum total variance are also shown on the LOFARgrams (solid lines). (a) Bottom slope $\alpha=0.000\,17$ and (b) bottom slope $\alpha=0.000\,35$.

the CPA, Eq. (26) is used as a Hough transform template when the moving source is near the CPA. Then, Eq. (18) is used as a Hough transform template as the range is increased for each moving window.

IV. EXPERIMENT

The RAGS03 experiment was conducted on the New Jersey Shelf between December 12, 2003 and January 4, 2004. The primary objective was to study the interaction of sound with shelf-break fluid processes in the ocean waveguide. The experiment featured moored sources transmitting 16-, 60-, and 100-Hz bandwidth acoustic signals at respective center frequencies of 224, 300, and 500 Hz. Three vertical receiving arrays, each having 32 elements, were located at ranges of 10, 20, and 30 km from the source (see Fig. 6). A 465-m-aperture, 96-element, bottom-mounted horizontal array was also deployed at 20-km range. Heavy shipping traffic near the experimental site provided a vast amount of broadband (20–500-Hz) ship noise data that were used to simultaneously estimate the waveguide invariant "beta" and range at the CPA to ship speed ratio. Figure 7 shows the 100-Hz ambient noise level at one element of the HLA measured during the experiment. Peaks in the noise level curve are usually associated with the CPAs of transiting ships. Two particular data sets, Ship-A and Ship-B, recorded on YD348 and YD351, were used to demonstrate the capability of the GRT method used for estimating the waveguide invariant beta and the ratio of range at the CPA to ship speed.

Figures 8(a) and 8(b) show "bathtub" striation patterns generated by Ship-A and Ship-B in the HLA beamformer output. Note in Fig. 8(a) the deviations from the regular striation patterns around $t=25$ min. Also note that the bathtub curves are slightly asymmetric in respect to time at $t=0$ in Fig. 8(b). These anomalies are mainly attributed to bathymetric variability at the experimental site (i.e., sand ridges and overall shelf slope). To verify that these anomalies are not due to a change in ship speed or bearing, the ratio of range at the CPA to ship speed and ship bearing is also estimated from the bearing-time records (BTRs) by using the Hough transform method. Figures 8(c) and 8(d) show the BTRs calculated using conventional broadband beamforming for Ship-A and Ship-B, respectively. Assuming a straight ship track, beam-time curve is described as

$$B(t) = \tan^{-1}\left(\frac{\frac{r_0}{V}\cos\theta + (t-t_0)\sin\theta}{-\frac{r_0}{V}\sin\theta + (t-t_0)\cos\theta}\right), \tag{31}$$

where t_0 is the time at the CPA, r_0/V is the ratio of range at the CPA to ship speed, and θ is the bearing angle measured from the endfire [see Fig. 8(e)]. The BTR outputs are integrated along the curves described by Eq. (31), normalized by

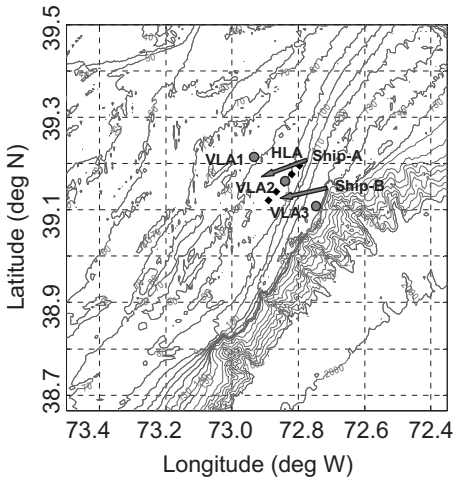


FIG. 6. Geometry of fixed vertical and horizontal receiver arrays and the bathymetry map of the experimental area.

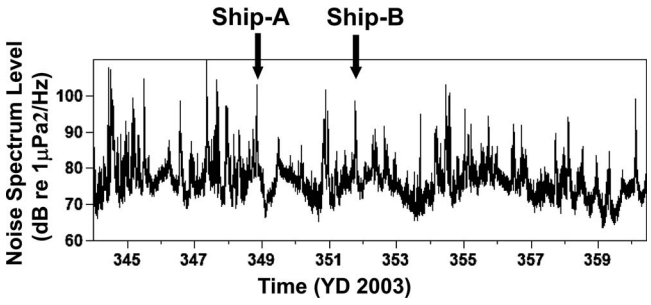


FIG. 7. Measured ambient noise level (100 Hz) at one element of the HLA. Two noise peaks corresponding to two merchant ships (Ship-A and Ship-B) are also shown.

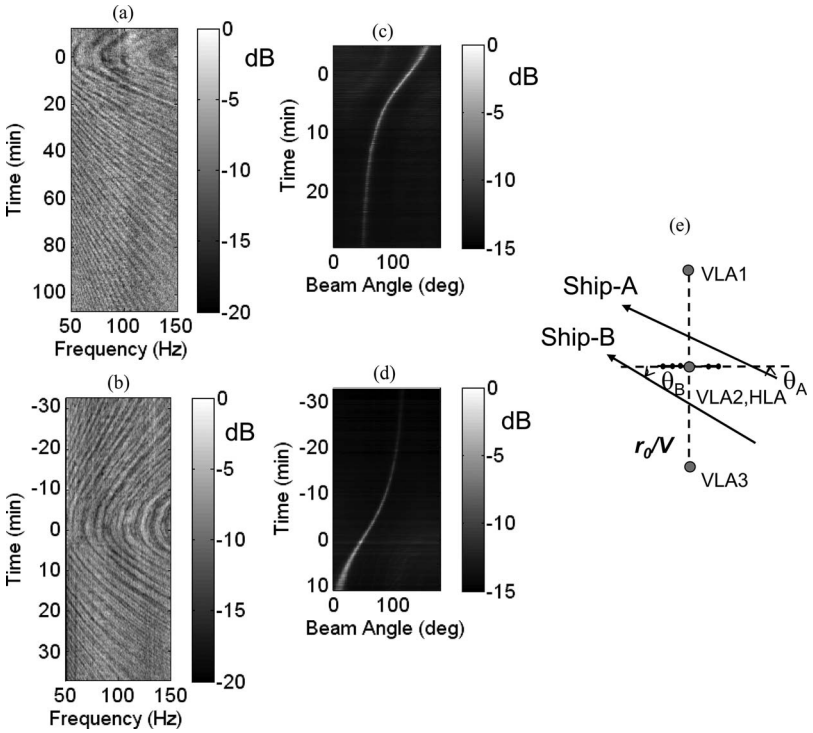


FIG. 8. LOFARgrams calculated from the HLA beam outputs: (a) Ship-A, (b) Ship-B, and corresponding beam-time records: (c) Ship-A, (d) Ship-B, and (e) bearing angles of two moving ships with respect to one horizontal and three vertical fixed receivers.

the curve length, and represented in the parameter space as

$$P(t_0,\theta,r_0/V)=\int B(t)ds \Big/ \int ds. \tag{32}$$

Figure 9(a) shows the estimated beam-time curve for Ship-A overlaid on the BTR data with the parameters $\theta=36$ deg and $r_0/V_A=450$ s from the GRT analysis [see Fig. 9(b)]. Figure 9(c) shows the estimated beam-time curve for Ship-B overlaid on the BTR data with the parameters $\theta=-42$ deg and $r_0/V_B=700$ s from the GRT analysis [see Fig. 9(d)]. Note the perfect match between the BTR outputs and the beam-time curves with the estimated (constant) parameters, indicating straight ship tracks and constant ship speeds.

Next, the range-dependency of waveguide invariant parameter beta is calculated from the HLA beamform output for Ship-A using $r_0/V_0=450$ s (see Fig. 10). Equation (26) was used for the Hough transform near time $t_0=0$ by assuming range-independent bathymetry between the receiver and source near the CPA. The waveguide invariant parameter is estimated as $\beta_0=1.05$ corresponding to the maximum total variance in Fig. 10(b). Then, Eq. (18) is used as a Hough transform template as the range is increased for each 10-min moving window. Note that as Ship-A moved to the shallower regions, the range-dependent waveguide invariant parametert varied from 1.05 to 1.45 (thick solid line with circles). Similar range-dependency of beta was also observed when

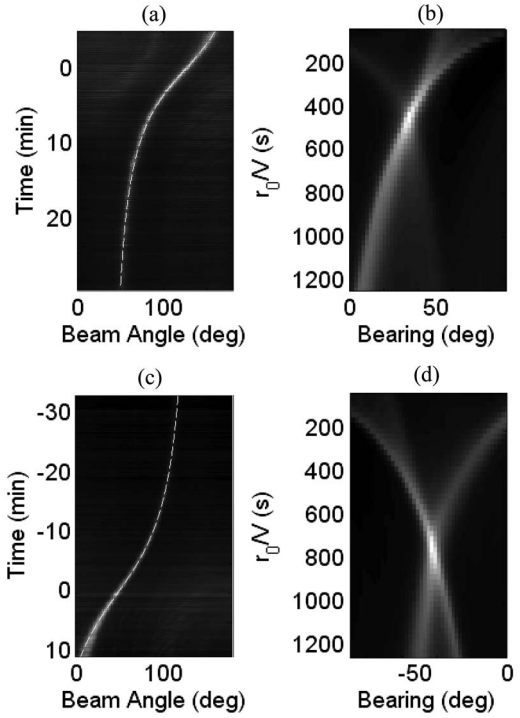


FIG. 9. Measured beam-time records: (a) Ship-A, (c) Ship-B, and corresponding Hough transform outputs: (b) Ship-A and (d) Ship-B. Estimated beam-time curves corresponding to the parameters with maximum Hough transform output also shown on the BTRs (white dashed lines).

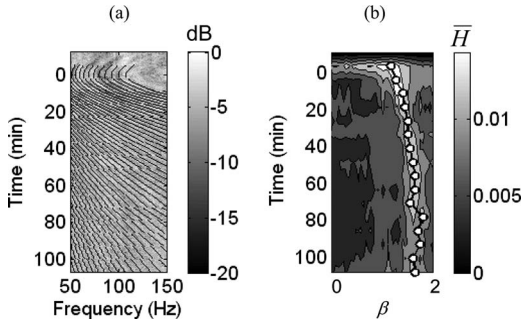


FIG. 10. (a) Ship-A LOFARgram calculated from the HLA beam output and (b) corresponding total variances in parameter space. Estimated trajectories for 7-min time window (solid lines) and corresponding to $\beta(r)$ values with maximum total variance (solid line with circle symbols) are also shown.

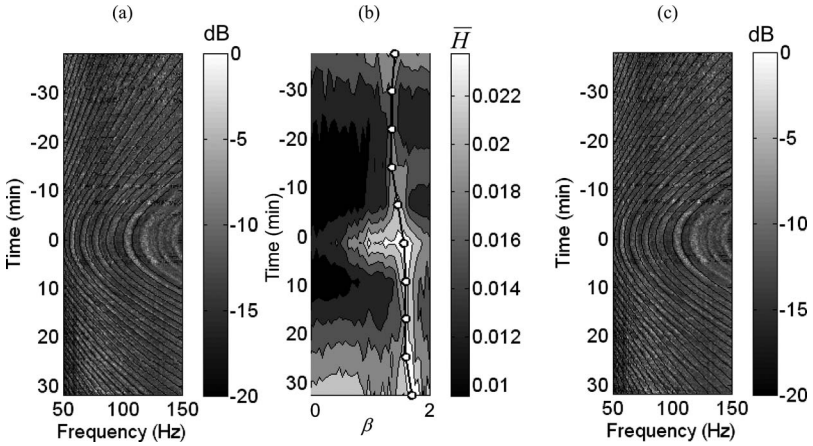


FIG. 11. (a) Ship-B LOFARgram calculated from the HLA beam output and (b) corresponding total variances in parameter space. The estimated trajectories for 7-min time window (solid lines) and corresponding $\beta(r)$ values with maximum total variance (dashed lines) are also shown. (c) Ship-B LOFARgram compared with the estimated trajectories (solid lines).

Ship-B moved from a deep to shallow region (see Fig. 11). If a range-independent bathymetry with a constant value of β is assumed, level curves are expected to be symmetric about the CPA point at $t_0=0$. Figure 11(c) shows the discrepancy between the level curves of the HLA beamform output and those calculated using a constant waveguide invariant parameter ($\beta=1.15$).

Since the ship locations were not independently recorded during the experiment, the parameter estimation from two distant receivers are used to verify the results of the parameter estimations from the HLA beamformer output. Note that if the ratio r_0/V is known for two single receivers, separated by distance R [see Fig. 12(a)], the ship speed can be calculated from the geometric relation

$$V=(R/t_{12})\sin\phi, \tag{33}$$

where t_{12} is the time during the ship travels between two CPA points, and ϕ is the CPA angle. The CPA angle is calculated as

$$\phi=\cot^{-1}[(1/t_{12})(r_1/V_1\pm r_2/V_2)], \tag{34}$$

where r_1/V_1, r_2/V_2, and t_{12} are estimated from the GRT analysis. Figures 12(b)–12(d) show the LOFARgrams recorded on mid phones of vertical line arrays (VLAs) VLA1, VLA2, and VLA3, respectively. Note that the level curves of the LOFARgram in Fig. 12(c) are similar to those of the HLA beamformer output depicted in Fig. 11(a) since the VLA2 and the HLA were deployed at a distance less than 500 m apart. Travel time differences between each CPA point are calculated from Fig. 12 as $t_{12}=780$ s for Ship-A and $t_{23}=952$ s for Ship-B. Corresponding CPA ranges to ship speed ratios are estimated from the GRT analysis as $r_1/V_A=625$ s and $r_2/V_A=450$ s for Ship-A and $r_2/V_B=625$ s and $r_3/V_B=450$ s for Ship-B. Then, the CPA angles and ship speeds are calculated from Eqs. (34) and (35) as $\phi_A=54$ deg, $\phi_B=48$ deg, $V_A=10$ m/s, and $V_B=8$ m/s. Next, the simultaneous estimation of waveguide invariant parameter beta and the ratio of the range at CPA to ship speed are demonstrated from the LOFARgram recorded when the ship is near the CPA. Equation (26) is used as a Hough transform template with the assumption of range-independent bathymetry between the receiver and ship near the CPA point. Figure 13(a) shows a 16-min window of a LOFARgram recorded on a single receiver at the VLA1 location. Figure 13(b) shows calculated total variance in the parameter space with a maximum at $r_1/V_A=625$ s and $\beta=1.05$. The solid lines in Fig. 13(a) are the level curves calculated using the estimated parameters. Figure 14(a) shows a 16-min window of LOFARgram recorded on a single receiver at the VLA2 location. Figure 14(b) shows calculated total variance in the parameter space with a maximum at $r_2/V_A=450$ s and $\beta=1.05$. The solid lines in Fig. 14(a) are the level curves calculated using the estimated parameters.

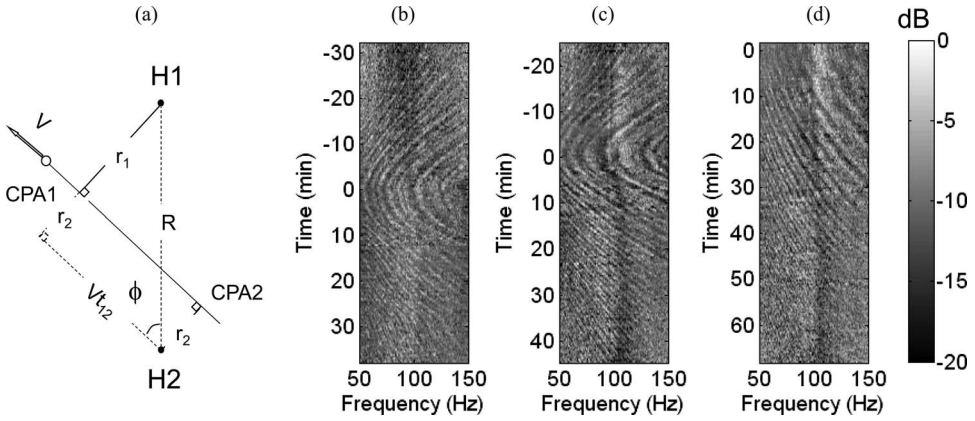


FIG. 12. (a) Geometry of a moving source and two fixed receivers for ship speed and bearing estimation. Measured Ship-A (single receiver) LOFARgrams at (b) VLA1, (c) VLA2, and (d) VLA3 locations.

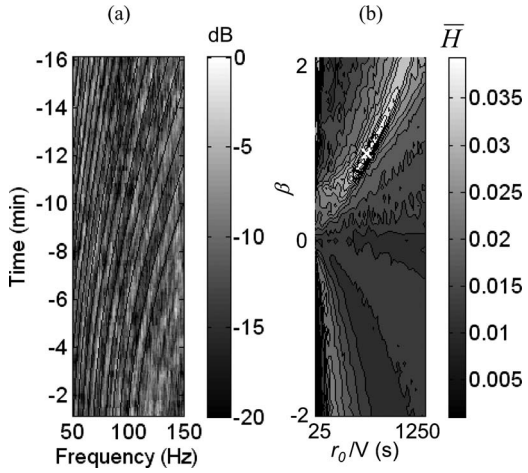


FIG. 13. (a) Ship-A LOFARgram recorded at the VLA1 and (b) calculated total variances in the parameter space. Estimated trajectories corresponding to the parameters with maximum total variance are also shown on the LOFARgram (solid lines).

Bathymetry dependency of waveguide invariant parameter beta is demonstrated using broadband acoustic signature of R/V Oceanus. Figure 15(a) shows the HLA beamform output for the R/V Oceanus as it moved along a cross-shelf track with a speed of 5.5 m/s. Figure 15(b) shows the measured water depth along the ship track. Near the CPA, the waveguide invariant parameter is estimated as $\beta_0=1.15$ using Eq. (28) for the Hough transform. For comparison, level curves calculated using a constant value of $\beta_0=1.15$ are also shown in Fig. 15(a) as solid lines. The discrepancy between the calculated level curves and those of the HLA beamform output suggests a bathymetry-dependent waveguide invariant parameter. Measured striation levels for HLA beamform output, single-receiver spectrum in winter conditions, and single-receiver spectrum in summer conditions are also compared. Figures 16(a)–16(c) show the striation levels for HLA beamform output, single-receiver data in winter conditions, and single-receiver data in summer conditions, respectively. The data in summer conditions were obtained during the

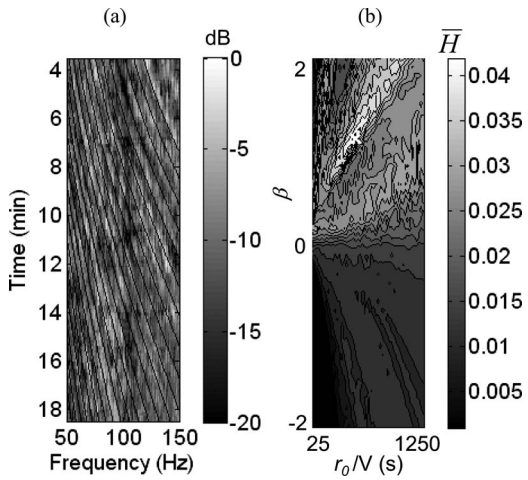


FIG. 14. (a) Ship-B LOFARgram recorded at the VLA2 and (b) calculated total variances in parameter space. Estimated trajectories corresponding to the parameters with maximum total variance are also shown on the LOFARgram (solid lines).

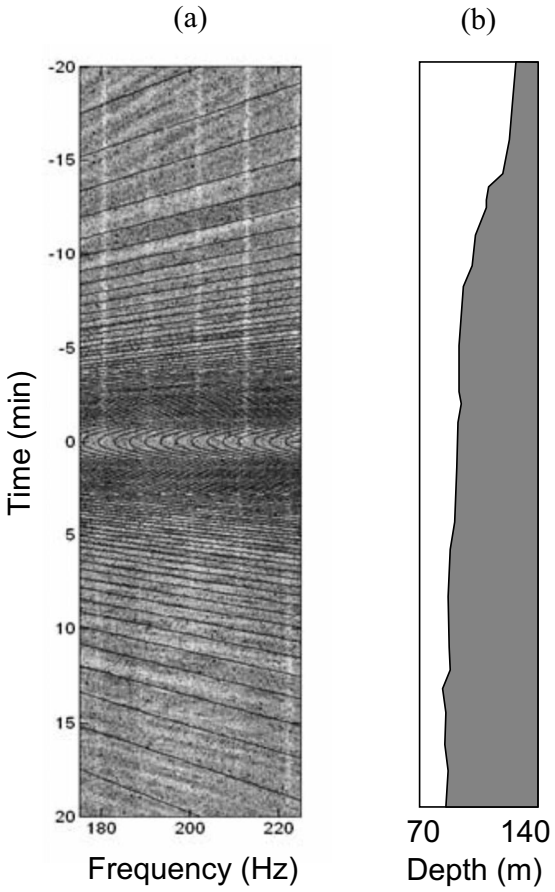


FIG. 15. (a) R/V Oceanus LOFARgram calculated from the HLA beamform output and estimated level curves (solid lines) using $\beta_0=1.15$. (b) Measured bathymetry along the ship track.

Boundary 2007 Experiment that was conducted in June of 2007 at Malta Plateau. The same source (R/V Oceanus), moving at 5.5-m/s ship speed, was used in both experiments. Figure 16(d) shows the bathymetry and source range near the CPA of the R/V Oceanus. The striations were observable up to 60-, 30-, and 5-min time lags from the time at CPA for the beamform winter data, single-phone winter data, and single-phone summer data, respectively. Considering the 5.5-m/s ship speed, these time lags correspond to ~20-, ~10-, and ~1.7-km ranges. The short range of observable striations in the single-phone summer data might be mainly due to the surface ducting and downward refracting effects of the summer profile introducing excess scattering and bottom loss.

V. SUMMARY

We have shown that HLA beamformer output exhibits striation patterns similar to those of acoustic intensity measured at a single receiver. However, during the RAGS03 winter experiment, HLA beamformer output provided the capability of tracking and ranging multiple sources at long ranges (up to 40 km) and various azimuths. Results from the Boundary 2007 summer experiment showed that this range might be drastically reduced by the surface ducting and downward refracting effects of the summer sound-speed profile. Numerical simulations and data from the RAGS03 experiment indicated that it is feasible to simultaneously esti-

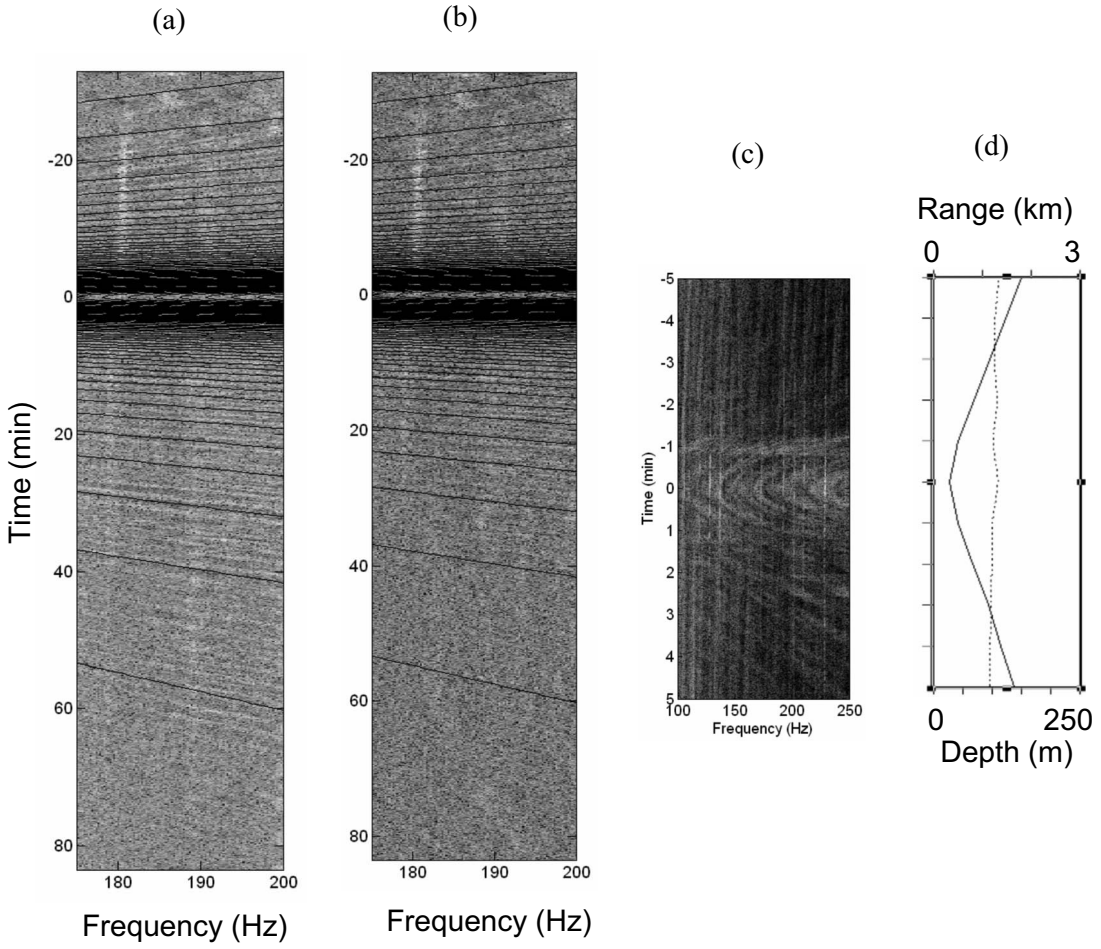


FIG. 16. Comparison of R/V Oceanus LOFARgrams by using (a) winter beamformed data, (b) winter single-receiver data, and (c) summer single-receiver data. Estimated level curves (solid lines) by using $\beta_0=1.15$ are also shown for the winter data. (d) Bathymetry along the ship track and source-receiver range measured during the Boundary 2007 experiment.

mate the waveguide invariant parameter beta and the ratio of range at CPA to ship speed from observed striation patterns in the single-receiver acoustic intensity data or HLA beamformer output.

High-quality LOFARgrams were recorded during December 2003 at the New Jersey Shelf when a uniform and range-independent winter sound-speed profile existed. Within the 50–150-Hz frequency band considered in this study, the acoustic intensity striation patterns seem to be mainly influenced by range-dependent bathymetry in the experimental area. The Hough transform method was applied to the HLA beamformer output, to the HLA beam-time-record data, and to the single-receiver acoustic intensity data. All three independent methods/data-sets provided comparable values of the ratio of range at CPA to ship speed. In addition, ship speed and bearing are extracted using the estimated time at CPA and ratio of range at CPA to ship speed for two distant receivers.

ACKNOWLEDGMENTS

This work was supported by the Office of Naval Research. We thank Peter Mignerey David Walsh, Steve Wales, Jeff Schindall, Earl Carey, Michael McCord, and Mark Hulbert of the Naval Research Laboratory for their scientific contribution to the RAGS03 experiment. We also thank the captain and crew of the R/V Endeavor and R/V Oceanus for the excellent support throughout the RAGS03 experiment.

[1]S. D. Chuprov, "Interference structure of sound field in the layered ocean," in *Ocean Acoustics. Modern State*, edited by L. M. Brekhovskikh and I. B. Andreeva (Nauka, Moscow, 1982), pp. 71–91.
[2]G. L. D'Spain and W. A. Kuperman, "Application of waveguide invariants to analysis of spectrograms from shallow water environments that vary in range and azimuth," J. Acoust. Soc. Am. **106**, 2454–2468 (1999).
[3]K. Heaney and H. Cox, "Rapid geoacoustic characterization for limiting environmental uncertainty for sonar system performance prediction," in *Impact of Littoral Environmental Variability on Acoustic Predictions and Sonar Performance*, edited by N. Pace and F. B. Jensen (Kluwer Academic, Dordrecht, 2002), pp. 123–130.
[4]A. M. Thode, "Source ranging with minimal environmental information using a virtual receiver and waveguide invariant theory," J. Acoust. Soc. Am. **108**, 1582–1594 (2000).
[5]W. Hodgkiss, H. Song, W. Kuperman, T. Akal, C. Ferla, and D. Jackson, "A long-range and variable focus phase-conjugation experiment in a shallow water," J. Acoust. Soc. Am. **105**, 1597–1604 (1999).
[6]A. M. Thode, W. A. Kuperman, G. L. D'Spain, and W. S. Hodgkiss, "Localization using Bartlett matched-field processor sidelobes," J. Acoust. Soc. Am. **107**, 278–286 (2000).
[7]T. C. Yang, "Motion compensation for adaptive horizontal line array processing," J. Acoust. Soc. Am. **113**, 245–260 (2003).
[8]D. Rouseff and C. V. Leigh, "Using the waveguide invariant to analyze Lofargrams," in Proceedings of the Oceans '02 MTS/IEEE (2002), Vol. **4**, pp. 2239–2243.
[9]T. C. Yang, "Beam intensity striations and applications," J. Acoust. Soc. Am. **113**, 1342–1352 (2003).
[10]P. V. C. Hough, "Method and means for recognizing complex patterns,"

U.S. Patent No. 3,069,654 (December 18, 1962).
[11]L. M. Brekhovskikh and Y. P. Lysanov, *Fundamentals of Ocean Acoustics*, 2nd ed. (Springer, New York, 1991).
[12]D. Rouseff, "Effects of shallow water internal waves on ocean acoustic striation patterns," Waves Random Media **11**, 377–393 (2001).
[13]S. V. Burenkov, "Distinctive features of the interference structure of a sound field in a two dimensionally inhomogeneous waveguide," Sov. Phys. Acoust. **35**, 465–467 (1989).
[14]B. J. Kraft, L. A. Mayer, P. Simpkin, P. Lavoie, E. Jabs, E. Lynskey, and J. Goff, "Calculation of *in situ* acoustic wave properties in marine sediments," in *Impact of Littoral Environmental Variability on Acoustic Predictions and Sonar Performance*, edited by N. Pace and F. B. Jensen (Kluwer Academic, Dordrecht, 2002), pp. 63–170.
[15]A. Turgut, D. L. Lavoie, D. J. Walter, and W. B. Sawyer, "Measurements of bottom variability during SWAT New Jersey Shelf experiments," in *Impact of Littoral Environmental Variability on Acoustic Predictions and Sonar Performance*, edited by N. Pace and F. B. Jensen (Kluwer Academic, Dordrecht, 2002), pp. 91–98.
[16]M. D. Collins, "FEPE user's guide," Naval Ocean Research and Development Activity, Stennis Space Center, MS, NORDA Technical Note 365 (1988).
[17]D. Rouseff and R. C. Spindel, "Modeling the waveguide invariant as a distribution," AIP Conf. Proc. **621**, 137–148 (2002).
[18]K. V. Hansen and P. A. Toft, "Fast curve estimation using preconditioned generalized Radon transform," IEEE Trans. Image Process. **5**, 1651–1661 (1996).

Geoacoustic inversions and localizations with adaptively beamformed data from a surface ship of opportunity source

Steven A. Stotts[a] and Robert A. Koch
Applied Research Laboratories, The University of Texas at Austin, P.O. Box 8029, Austin, Texas 78713-8029

(Received 20 April 2009; revised 28 September 2009; accepted 28 September 2009)

A technique has been developed for incorporating adaptively beamformed (ABF) data into geoacoustic inversions and source localizations based on matched-field processing. Rather than adaptively adjusting matched-field weights, for this paper ABF processing is applied to construct subaperture beam data from which the inversions and localizations are derived. This application of ABF processing to inversion and localization is successfully demonstrated with data collected on a horizontal line array from a surface ship of opportunity source (SSOS). When measured data representing noise from one or more interfering sources at various levels are injected into the data collected from the SSOS, the inversion processing with ABF subaperture data is successful at interferer noise levels exceeding the SSOS levels by 10–13 dB and for which the inversion processing with conventionally beamformed data fails.
 [DOI: 10.1121/1.3257207]

PACS number(s): 43.30.Pc, 43.60.Kx, 43.60.Mn, 43.30.Bp [AIT] Pages: 84–95

I. INTRODUCTION

The purpose of this paper is to establish that accurate geoacoustic inversions and source localizations can be performed, either separately or simultaneously, from adaptively beamformed (ABF) subaperture data. Processing of horizontal line array (HLA) data with ABF techniques has been applied successfully over several decades to improve source detection by reducing the effects of noise from strong interferers, relative to processing with conventional beamforming (CBF) techniques, for many array configurations in a variety of ocean environments.[1–3] A number of results on the application of adaptive techniques in matched-field processing have been presented over the past several years.[4–9] Thus far, most previous treatments have applied adaptive processing to adjust the matched-field weights for element level data.[4–7] Adaptive matched-field processing methods are perceived to be sensitive to environmental mismatch and insufficiently robust for geoacoustic inversions.[5–8] Several alternative methods have been developed to deal with the lack of robustness.[7,9]

The approach in this paper avoids issues associated with the sensitivity of adaptive matched-field processing to environmental mismatch. Instead, the adaptive processing is applied to element level plane wave weights to produce subaperture beam data and is robust. These ABF subaperture data are then treated as data from single receivers located at the subaperture phase centers, and the matched-field processing replicas are computed only for the positions of the phase centers. This approach has the advantage that the computational effort required to produce the matched fields at the subaperture phase centers is reduced relative to that required for element level data.[10] A significant requirement for the success of this approach, at least for CBF data, is that the beams are broad enough, i.e., the subapertures be sufficiently short, to encompass the dominant multipaths in the model replicas. Thus, this paper addresses a significant issue for applying this technique to inversion and localization processing, namely, whether the information from the important multipaths will be preserved still by the ABF processing required to produce subaperture beam data.

Geoacoustic inversion processing with CBF subaperture data from a surface ship of opportunity source (SSOS) has been demonstrated.[10,11] In the present paper, accurate localizations are produced with ABF subaperture data from the same SSOS for which the inversions of CBF subaperture data were performed. In addition, an analysis of localizations with ABF subaperture data examines the effect of injecting simulated noise from interferers into the HLA data from the SSOS.[12] Then, simultaneous geoacoustic and source track parameter value inversion results are computed from ABF and CBF subaperture data formed from element data with real source interferers injected into the data.

In Sec. II the experimental geometry and the acoustic environment for the SSOS data are described, and details are provided about the processing applied to produce the subaperture ABF data. The geoacoustic and source track models and inversion processing details, including definitions for the cost function and cost distribution envelopes, are described in Sec. III. The estimation of the resulting signal to noise ratio (SNR) levels when interferers are injected into the original measured data is discussed in Sec. IV. Section V contains a summary of the results of source localizations with a simulated interferer included in the CBF and ABF subaperture data. Section VI demonstrates that successful inversions are possible with ABF data. For Sec. VII interferer data are constructed from measured SSOS data outside the

[a]Author to whom correspondence should be addressed. Electronic mail: stotts@arlut.utexas.edu

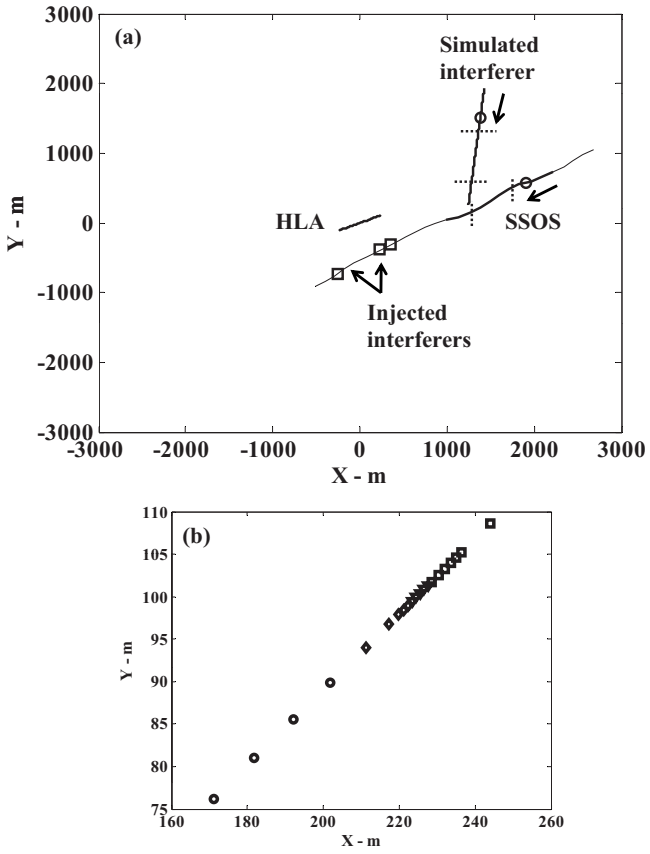


FIG. 1. (a) HLA location, SSOS track reconstruction, simulated interferer track, and source positions for which data for injected interferers were extracted. North is in the positive y direction. Bold line on source track identifies data time segment for localizations. Dotted lines bracket track segment during inversion data. Circles identify source and simulated interferer position at sequence 65 for ambiguity surface construction. (b) HLA geometry for subaperture 1 (circles), subaperture 2 (diamonds), subaperture 3 (triangles), and subaperture 4 (squares).

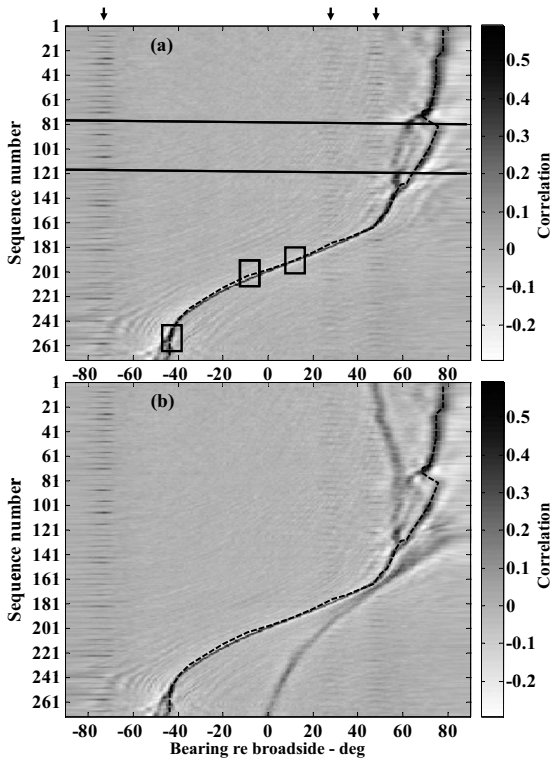


FIG. 2. Bearing-time display of R/V Longhorn passage: (a) measured data only and (b) with simulated interferer. The time segment bracketed in (a) by horizontal lines was selected for inversion data. Arrows at the top of (a) indicate bearings with seismic profiler activity. Time samples from which data were injected into the inversion data to represent an interferer are indicated by boxes in (a). The dashed lines in (a) and (b) show the source bearing track for which beam spectra were constructed.

time interval spanning the inversion data and are injected into the inversion data. Section VII A contains comparisons of the results from inversions with ABF and CBF SSOS data with two interferers injected, and comparisons of inversion results from ABF and CBF SSOS data with three interferers injected are discussed in Sec. VII B. Section VII C contains a discussion of the performance of ABF in suppressing intermittent noise sources (seismic profilers) from SSOS inversion data as compared to the performance of CBF processing.

II. EXPERIMENT DATA AND HLA BEAM PROCESSING

The processed data for the analyses presented in this paper were collected on a bottom mounted HLA from a SSOS, the R/V Longhorn, in the Gulf of Mexico in February 1998.[10] Figure 1(a) shows the array position and the Longhorn track reconstruction from recorded GPS. The array line of bearing (LOB) (66° east of north) was approximately parallel to the bathymetric contours.[13] The R/V Longhorn approached endfire of the array moving at approximately 8 kts and traversed from northeast to southwest along an approximately range-independent track. Figure 1(b) shows the positions of the elements in four subapertures from the northeastern half of the array. The data from these elements were selected for inversion processing. The total array length is 534.2 m, and the four subaperture lengths are 33.68, 12.16, 4.54, and 16.98 m. The data were sampled at 2457.6 Hz, and complex spectra were formed with 8192 point (8k) fast Fourier transforms (FFTs).

The bearing-time display (BTD) in Fig. 2(a) comprises approximately 15 min of the data collected. The BTD was obtained by cross-correlating CBF beam data in the 10–250 Hz band from two seven-element subapertures in the northeastern half of the array. A subset of the data encompassing FFT sequences 77–120 was processed for the inversions in this paper. There are also emissions from several distant seismic profilers evident in Fig. 2(a), but the FFT sequences with these emissions were deleted from the inversion data as much as possible, except for the analysis discussed in Sec. VII C. The range from the SSOS to the center of the array varied from 1.0 to 1.7 km (approximately 9–15 water depths) for the inversion data sample.

For inversions from CBF and ABF data, track beam spectra along the SSOS bearing track shown in Fig. 2 were formed for all four subapertures. (Data generated from more precise bearing tracks for each subaperture produced little change in the inversion solutions, in part because the short subapertures chosen for the inversions produce wide main lobe beam widths containing all the dominant multipath arrivals.) The data chosen for the inversions contain samples with strong multipath effects that produce several arrivals, as seen between bearings 55° and 75° in Fig. 2. In the data

segments with strong multipath effects, the beam spectra follow the bearing track of the multipath with the highest correlation. In these data segments, the multipath arrivals from the water column are supplemented by arrivals along sub-bottom propagated (refracted) paths that are possible because the top sediment layer has a sound speed ratio (sound speed at the top of the sediment layer divided by the ocean sound speed at the water-sediment interface) <1.0.[14]

This dataset was chosen for several reasons. First, the soft-sediment environment makes a simultaneous inversion of the source track and geoacoustic parameters necessary.[11] Second, the spread in multipath arrivals in the inversion data tests the effects of the ABF beam processing on the inversion solution. Third, the successful inversions performed in Ref. 10 provide a good benchmark result. Finally, the source track provides good data for the studies with injected interferers discussed in Secs. V and VII.

Numerous references give detailed descriptions and applications of ABF to plane wave beamforming of HLA subaperture data,[1–3] and only the highlights will be discussed here. To produce ABF inversion data, a minimum variance distortionless response plane wave beamformer was applied to the measured element data to construct the plane wave weights from cross-spectral matrices (CSMs) averaged over frequency and/or time for each subaperture.[1,12,15] In addition, a −3 dB white noise gain constraint was applied in constructing the ABF weights. For the inversion data, the CSMs formed to construct the ABF weights were averaged over approximately 5 Hz. However, the inversion solutions were approximately the same if any frequency averaging over 2–7 Hz bands was performed. The inversion results discussed in this paper were obtained from ABF data produced from CSMs constructed without sequence averaging. Constructing the CSMs with two or three sequence averages did not substantially alter the inversion solution. Constructing CSMs by averaging over more than three sequences, or over frequency bands exceeding 7 Hz, produced some degradation in received levels (RLs) in the ABF subaperture data. Seismic profiler signals, when present, occur mostly in single FFT sequences; therefore, no additional suppression of the profiler signals was realized with ABF data from CSMs constructed with sequence averaging.

The ABF and CBF track beam spectrograms from the 33.68 m length subaperture are shown in Fig. 3. The RL interference pattern produced by frequency dependent propagation effects is well represented in the spectrogram. The data contain both tonal and broadband data components that can be processed in the SSOS inversions. The seismic profiler emissions at frequencies below 50 Hz are identifiable in Fig. 3 by the numerous horizontal stripes apparent in the spectrum. The ABF spectrogram in Fig. 3(a) shows that the profiler levels for many of the data sequences are suppressed in the ABF data for frequencies in the 50–100 Hz band, compared to the CBF spectrogram for the same sequences and frequencies, as shown in Fig. 3(b). For other frequencies, the spectrogram from the CBF track beam data contains similar features to the ABF spectrogram, but the SNR for the SSOS is slightly lower in Fig. 3(b) than in Fig. 3(a). This is a potential advantage to performing inversions from subaperture ABF data in that greater decibel differences between the RL peaks and nulls would provide an increase in the inversion processor performance with the hybrid cost function described in Sec. III C. Also, a comparison of the inversions from the ABF and CBF data in the 50–100 Hz band is presented in Sec. III C.

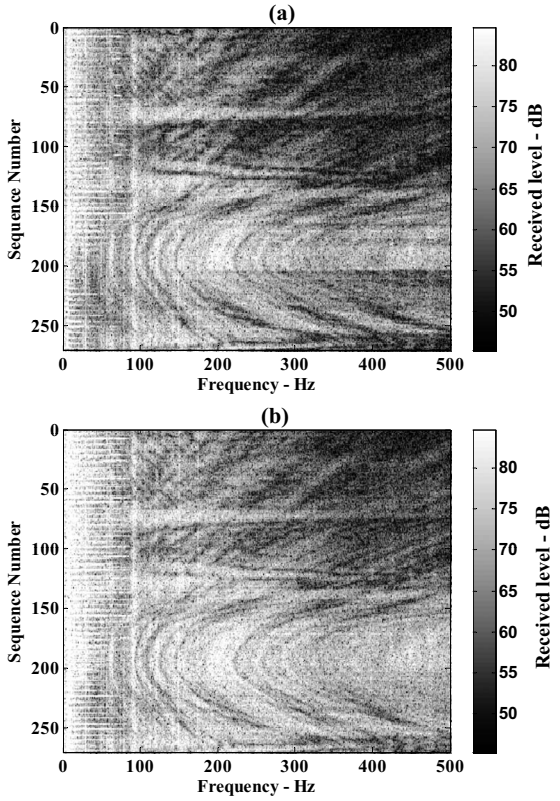


FIG. 3. Track beam spectrogram from a four element subaperture for the 15 min data segment depicted in Fig. 2(a) for (a) ABF data and (b) CBF data.

III. LOCALIZATION AND INVERSION PROCESSING

A. Geoacoustic and source track models

The sediment is described by a three layer model consisting of two fluid layers overlying a fluid half-space, i.e., the substrate. Perhaps the most important inversion parameter for each layer is the layer sound speed ratio $C_{tj}/C_{b(j-1)}$ (speed at the top of the layer divided by the speed at the bottom of the next layer up), where t_j (b_j) refers to the depth at the top (bottom) of layer j. For the top of the sediment $C_{b0} \equiv C_w$, where C_w is the water sound speed at the water-sediment interface. For the top two sediment layers, the thickness H_j of layer j, the layer density ρ_j, the layer sound speed gradient C_{gj}, the attenuation α_{tj} at 1 kHz at the top of the layer, the layer attenuation gradient α_{gj}, and the layer frequency exponent e_j are the other possible inversion parameters. For the half-space, the sound speed ratio C_{sub}/C_{b2}, the density ρ_{sub}, the attenuation α_{sub}, and the attenuation exponent e_{sub} are possible inversion parameters. For the analysis in this paper the sediment layer and half-space attenuation exponents were all set to 1.0, because attenuations for soft sediments follow this power law form.

In addition to the geoacoustics, the initial source range and bearing and a constant course and speed were inversion parameters. The source depth was fixed at 2.0 m for the R/V Longhorn.

B. Inversion overview

The inversions in this paper were performed using simulated annealing.[16,17] At each inversion iteration in the algorithm, the value of just 1 parameter is varied, and modeled received fields at subaperture phase centers are calculated and correlated with the measured data.[18] The normal mode propagation model is based on NAUTILUS.[19–22] At each iteration a cost value, representing the discrepancy of the modeled fields from the data, is computed. For each iteration, if the cost value decreases from the previous iteration, the parameter value that was varied is retained for the next iteration in which a different parameter value is varied. If the cost value for some iteration is greater than the cost value from the previous iteration, the parameter value that was changed has some probability, calculated from the Metropolis criterion,[16] of being accepted for the next iteration instead of being replaced by the parameter value from the previous iteration.

There are several variables that define the annealing schedule. At each stage in the annealing search algorithm, the temperature determines the sizes of the parameter value perturbations and the probability of accepting changes in parameter values that produce an increase in the cost. A maximum number of iterations and the number of perturbations for each parameter per temperature cycle are specified along with a temperature reduction rate. A high initial temperature and a large number of iterations are chosen for the inversion studies presented here to allow a sufficient sampling of the cost value distributions and to ensure that the inversion solutions obtained are independent of the initial parameter values.

C. Cost function

The cost function is constructed from the mean squared difference of model and data cross-spectra in multiple FFT sequences and multiple frequency bands.[18] The final correlation form of the cost function is

$$C = 1 - \sum_{\substack{\text{center}\\ \text{frequencies}}} \frac{\left| \sum_{\substack{\text{element pairs,}\\ \text{sequences}}} \frac{\langle D_i D_j^* \rangle M_i^* M_j + \langle D_i^* D_j \rangle M_i M_j^*}{2} \right|^2}{N_{\text{CEN}} \sum_{\substack{\text{element pairs,}\\ \text{sequences}}} |\langle D_i D_j^* \rangle|^2 \sum_{\substack{\text{element pairs,}\\ \text{sequences}}} |M_i^* M_j|^2}, \quad (1)$$

where N_{CEN} is the number of center frequencies f. The $M_i(f)$ are the modeled subaperture complex spectra (an asterisk denotes the complex conjugate), and the (i,j) element pair cross spectral data, formed from the subaperture beam data D_i and D_j^* are

$$\langle D_i D_j^* \rangle(f) = \frac{|D_{\text{RL}}|^2}{N_{\text{bins}}} \sum_{\text{bin}} D_i(f_{\text{bin}}) D_j^*(f_{\text{bin}}) / |D_{\text{REF}}(f_{\text{bin}})|^2, \quad (2)$$

where bin is the index for frequency f_{bin} in the average over the N_{bins} FFT bins about the center frequency f. The reference and RL normalizations are given by

$$|D_{\text{REF}}(f_{\text{bin}})|^2 = \sum_{\text{elts}} |D_i(f_{\text{bin}})|^2 / N_{\text{elts}}, \quad (3)$$

$$|D_{\text{RL}}|^2 = \sum_{\text{bin,elts}} |D_i(f_{\text{bin}})|^2 / (N_{\text{elts}} N_{\text{bins}}), \quad (4)$$

where N_{elts} is the number of elements in the summation with index elts. The final form of the cost function given in Eq. (1) satisfies $0 \le C \le 1$.

D. Cost envelopes

To examine the uncertainties in the inversion parameter estimates, cost distributions and the envelopes of their minima were constructed.[10,13,23] Broader parameter cost distribution envelopes translate into greater uncertainty in the parameter value solutions obtained from the inversion as compared to inversion solutions that produce narrower parameter cost distribution envelopes. A rigorous construction of probability distributions in a Bayesian approach[24,25] or a simple method of evaluating the widths of the cost distribution envelopes[18] can provide estimates of the inversion parameter value uncertainties but are not applied in this paper.

IV. SNR FOR INJECTED INTERFERERS

The effects of an interferer on source localizations are examined in Sec. V using a simulated interferer, and the effects of multiple injected data interferers on simultaneous source track and geoacoustic parameter inversions are discussed in Secs. VII A and VII B. For both analyses the interferers were injected into the data prior to the ABF subaperture data beam processing. The resulting SNR is estimated by

$$\text{SNR}(f) = 10 \log_{10} \left[I_{\text{SSOS}}(f) \Big/ \sum_{\text{int}} I_{\text{int}}(f) \right]. \quad (5)$$

The intensity $I_{\text{SSOS}}(f)$ from the source track beam is an average over the inversion interval sequences at frequency f. Similarly, $I_{\text{int}}(f)$ is the averaged intensity from the track beam following the int$^{\text{th}}$ interferer. The SSOS and interferer beam intensities in Eq. (5) are measured from separate ABF track beams to remove, as much as possible, interference from each other and from other sources. The SNR values discussed in Secs. V and VII are calculated for each inversion frequency and for each subaperture.

V. SOURCE LOCALIZATION FROM CBF AND ABF DATA WITH AN INTERFERER PRESENT

To compare the ability to localize sources using ABF and CBF subaperture data, the simulated field from a near-surface source interferer loud enough to produce a −20 dB SNR was injected into the subaperture element data. For the localization analysis the single interferer was simulated with a quasi-static mode model. The track of the simulated interferer is shown in Fig. 1(a). The simulated interferer source depth was 2 m. The effect of the interferer in the BTD is shown in Fig. 2(b).

A localization analysis was performed by evaluating the cost function in Eq. (1) with both ABF and CBF subaperture beam data produced from the element data with the injected

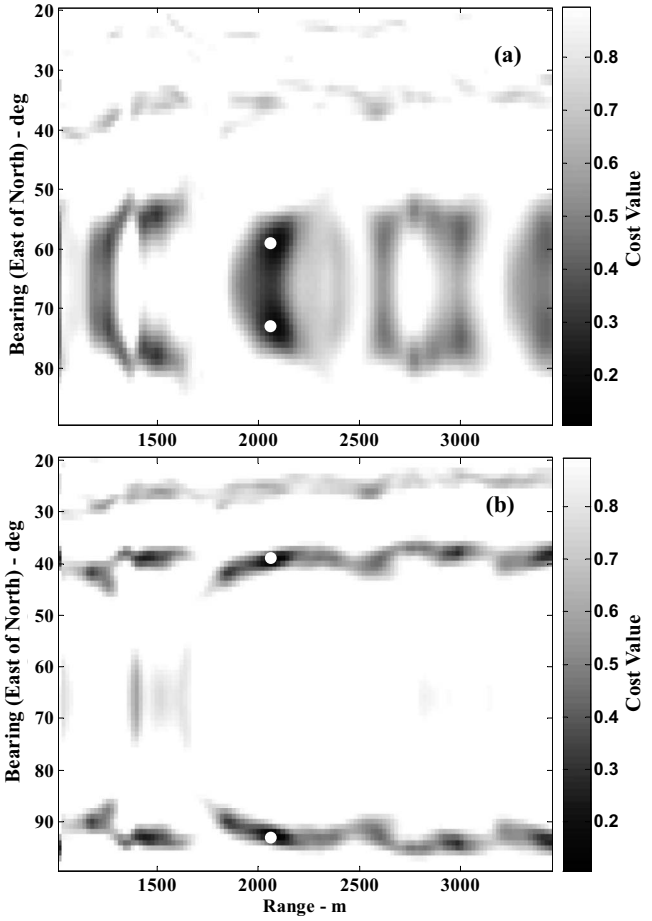


FIG. 4. Range-bearing ambiguity surface from (a) ABF and (b) CBF subaperture data at sequence 65. The location of the cost value minimum and the mirror image minimum are designated by white dots.

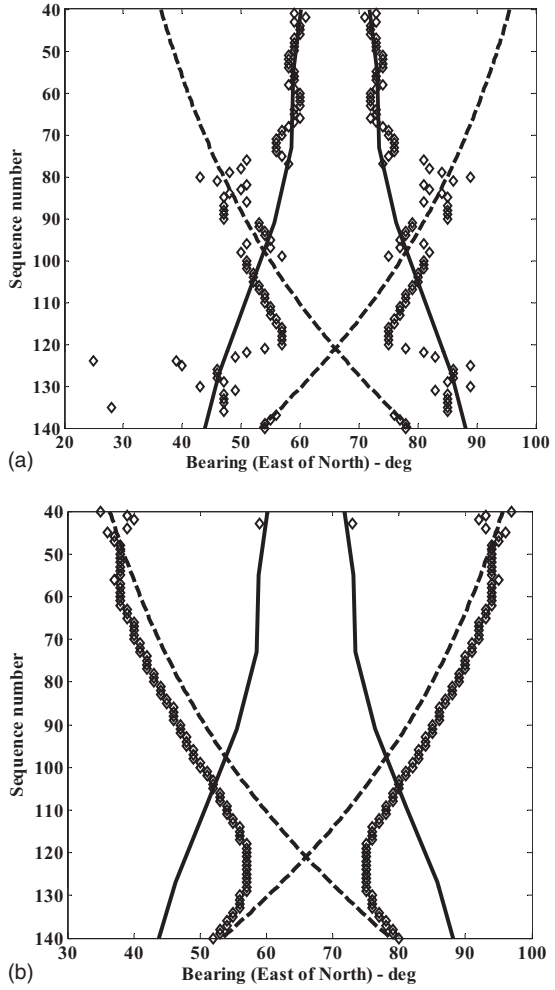


FIG. 5. Bearings (diamonds) of the cost value minima in the ambiguity surfaces for each sequence from 40 to 140, SSOS and mirror image SSOS track (solid), and interferer and mirror image interferer track (dashed) for (a) ABF data and (b) CBF data.

interferer. To construct the matched fields, the geoacoustic model was fixed at the values obtained from the inversion solution with the longer range data segment obtained in Ref. 10. Ambiguity surfaces for single time segments were produced by varying the source reference range and reference bearing and evaluating the cost function with data from ten equally spaced center frequencies starting from 120.9 Hz and ending at 188.4 Hz. The source locations from the reconstruction and the interferer location during these time segments are shown in Fig. 1(a). The ABF and CBF beams are conical about the LOB of a HLA, which produces in the ambiguity surface mirror image positions for the interferer and source about the LOB. The mirror of the bearing of the simulated interferer crosses the bearing of the SSOS.

The cost value minimum of 0.178 in the ambiguity surface for sequence 65, shown in Fig. 4(a), occurs at an approximate bearing of 73° east of north and at a range near 2100.0 m, which is near the reconstructed SSOS position for this time segment. The cost values in Fig. 4(a) are symmetric about the line of bearing. Near the mirror image of the SSOS at approximately 58° bearing east of north and at 2100.00 m range, the cost value is almost the same as that near the true source position. There is a sidelobe with a minimum cost value of 0.357 at a range of approximately 1270.0 m and at the same bearing as the cost value minimum. Another sidelobe with a cost minimum of 0.477 occurs at a range of approximately 2630.0 m. Thus, the cost value minimum near the actual SSOS position is easily distinguished from the cost value minima in the range sidelobes on the ambiguity surface.

In contrast the ambiguity surface shown in Fig. 4(b) from the CBF data for the same sequence has a cost value minimum of 0.0964 near the simulated interferer position at a bearing of 39° east of north and at a range of approximately 2100.0 m. The minimum cost at the mirror image source position with bearing 92° east of north is nearly the same. These cost value minima are much lower than for the true source in Fig. 4(a) because of the high level of the interferer. The cost value minimum in the sidelobe at 39° east of north and at an approximate range of 1450.0 m is 0.2473, and the minimum cost is 0.436 in the sidelobe at approximately 2610.0 m.

Next, ambiguity surfaces were constructed for sequences 40–141. The bearing at the cost value minimum on each surface, along with the source and interferer tracks and their mirror images, are shown in Fig. 5(a) for the ABF data and in Fig. 5(b) for the CBF data. The result for the ABF data is that the SSOS is accurately localized until the mirror image of the simulated interferer is within approximately 10° of the

SSOS bearing around sequence 80. The mirror image interferer track enters the subaperture beams that follow the SSOS bearing in the neighborhood of sequence 101. The cost value minima occur at the correct SSOS position again when the simulated interferer approaches the HLA endfire and the SSOS separation from the interferer exceeds 10° (around sequence 125). For the CBF data Fig. 5(b) shows that with the exception of one of the sequences, the cost value minimum for the ambiguity surface at each sequence occurs at the interferer track location for that sequence. Therefore, for −20 dB SNR the SSOS can be localized with the ABF subaperture data, but not with the CBF subaperture data, for SSOS-interferer bearing separations greater than 10°.

For a simulated interferer that produces −10 dB SNR, the SSOS localization from ABF data has the same performance as with a simulated interferer that produces a −20 dB SNR. The SSOS localizations from the CBF data with a simulated interferer that produces a −10 dB SNR succeeded sporadically and only for SSOS-interferer bearing separations of more than 30°.

VI. INVERSIONS WITH CBF AND ABF DATA NOT CONTAINING INJECTED INTERFERERS

Inversions of both CBF and ABF track beam data were performed to obtain both geoacoustic and source track parameter values. Simultaneous inversions for geoacoustic and source track parameter values are necessary because for soft-sediment environments source track and geoacoustic parameters are strongly coupled, e.g., the source positions that minimize the cost function depend sensitively on the geoacoustic parameter values.[11] The geoacoustic inversion solution obtained for the medium range data segment in Ref. 10 from CBF subaperture data, and the track reconstruction data provide parameter value estimates to initialize a simultaneous inversion for source track and geoacoustic parameter values from the CBF track beam data.

The solution from the inversion of CBF track beam data provides the starting values for a simultaneous inversion for source track and geoacoustic parameters with the ABF track beam data. This choice of the initial parameter values does not affect the purpose of this analysis which is to test whether the inversion solution obtained from the CBF data is preserved when the inversion is performed with subaperture ABF track beam data. Thus, the simulated annealing algorithm was initiated at a high temperature, and broad bounds were set for the inversion parameter values to permit the inversion search to pursue alternative solutions far from the initial parameter values. For the inversion from both CBF and ABF track beam data the bounds for each parameter value are listed in Table I.

The inversion frequencies were 91.2, 105.6, 113.4, 122.1, and 150.0 Hz. For the inversion data FFT sequences were chosen which did not include the seismic profilers. The solutions for the geoacoustic and source track parameter values from the inversions with the CBF and ABF data (after approximately 40 000 iterations) are shown in Table I. The solutions from the ABF and CBF data were nearly identical except for a negligible difference in the layer density and a slightly higher minimum cost value for the ABF inversion.

TABLE I. Source track and three-layer geoacoustic parameter inversion bounds and solutions from CBF and ABF data.

Parameter	Parameter value	Lower bound	Upper bound
Min. cost	0.349 CBF/0.373 ABF		
Range (m)	1856.7	1000	3000
Bearing (deg E of N)	72.21	55	95
Course (deg E of N)	225.32	190	260
Speed (m/s)	4.464	3.0	6.0
C_{t1}/C_w	0.984	0.95	1.1
H_1 (m)	39.004	10.0	60.0
C_{g1} (/s)	1.1904	0.5	2.5
ρ_1 (g/cm^3)	1.6464 CBF/1.6438 ABF	1.5	1.8
C_{t2}/C_{b1}	1.0202	0.95	1.1
α_{t1} (dB/m kHz)	0.0201	1.0×10^{-3}	0.2
α_{g1} (dB/m^2 kHz)	9.276×10^{-5}	0.	0.01
H_2 (m)	84.275	40.0	120.0
C_{g2} (/s)	1.4287	0.5	2.5
ρ_2 (g/cm^3)	1.7099	1.6	2.0
C_{sub}/C_{b2}	1.0321	1.0	1.05
α_{t2} (dB/m kHz)	0.0166	5.0×10^{-3}	0.2
α_{g2} (dB/m^2 kHz)	8.1507×10^{-5}	0.	0.01
α_{sub} (dB/m kHz)	0.0561	0.01	0.2
ρ_{sub} (g/cm^3)	2.3421	1.8	2.4

This exercise demonstrates that geoacoustic inversion and source localization can be performed simultaneously from subaperture ABF track beam data.

Cost distribution envelopes for two of the source track parameters from the CBF and ABF data inversions (with no interferers) are shown in Figs. 6(a) and 6(b). The cost envelopes from ABF and CBF data inversions exhibit approximately the same width. Figures 6(c) and 6(d) show cost envelopes for two of the geoacoustic parameters. For each parameter, the inversions with CBF and ABF data (with no interferers) produce comparable cost envelope widths.

To validate the inversion solutions obtained from CBF and ABF data, the parameter values obtained from the inversions were used to estimate transmission loss (TL) data measured from the SSOS. Measured RL exhibit frequency and range dependent structure that can be compared to modeled TL structure over the ranges in the inversion data. Measured TL are obtained by subtracting the measured RL from $10\log_{10}|S_f|$, where $|S_f|$ is the solution obtained from the inversions for the source level at frequency f.[18] Modeled TL are calculated for propagation from the SSOS to the subaperture phase centers. Comparisons of model and data TL at the ranges and frequencies of the inversion data are shown in Fig. 7 for the ABF data. The structures in the CBF data TL (not shown) and in the ABF data TL, i.e., in the data RL, are comparable except the ABF data contain larger differences between maxima and minima (as opposed to the CBF data) due to the reduction in sidelobe noise produced by the ABF. For example, the nulls at 1325 m at 105.6 Hz and at 1575 m for 150.0 Hz in Figs. 7(b) and 7(d) are approximately 5 dB deeper than for the CBF data. For this limited range data, the deeper nulls in the ABF data did not produce a better inversion solution than obtained from the CBF data.

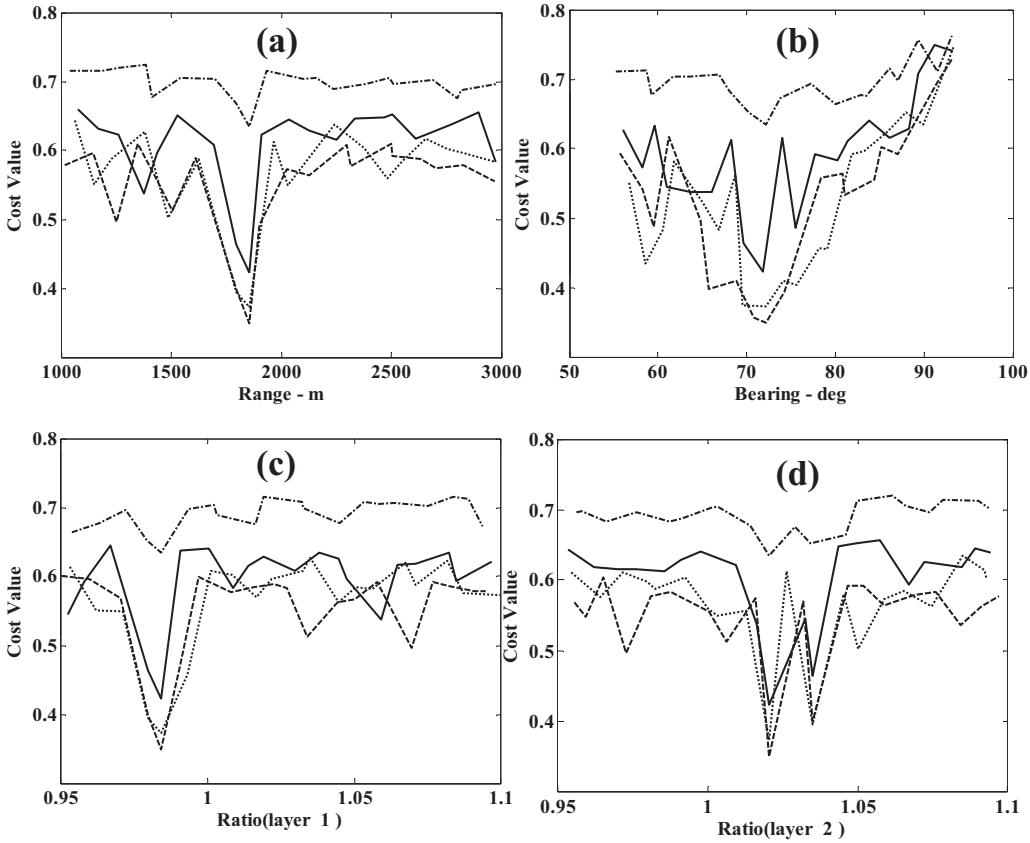


FIG. 6. Cost distribution minimum envelopes for initial SSOS track (a) range and (b) bearing and for the sound speed ratio for (c) layer 1 and (d) layer 2, obtained from inversions of both the CBF (dash) and ABF (dot) data with no interferers and from inversions of CBF (dash-dot) and ABF (solid) data with two interferers.

To further substantiate the inversion solution obtained with ABF data, the parameter values obtained from the inversions were used to estimate the TL from measured tow data recorded during another part of the experiment.[10,13] A comparison between the measured and modeled TL is shown in Fig. 8. The model TL values from the inversion solution produce accurate predictions of the measured TL out to long ranges (2–8 km), whereas the data for the inversions are from a source at ranges 1.0–1.7 km.

VII. INVERSIONS WITH INTERFERERS PRESENT

Having established that accurate inversions are possible with ABF data from a SSOS, the performance of inversions from ABF data, compared to those from CBF data, was investigated with strong interferers injected into the measured data and with data segments containing noise from seismic profilers. Interferers contaminate the inversion data through beam sidelobes, and the ability of ABF to reduce the sidelobe contamination without affecting the multipath contributions from the source will determine the ABF performance enhancement.

A. Two injected interferers

To study the performance of ABF against continuously emitting interferers, data snippets from time intervals that followed the inversion data segment [see Fig. 2(a)] were injected into the inversion data segments. Specifically, for each FFT frequency bin the complex spectra from the time snippets with interferers were added to the complex spectra from the inversion data. The injected snippets contain data from the SSOS at different ranges and bearing locations than in the inversion data, and the interferer levels were louder than the SSOS in the inversion data since in the injected data segments the SSOS was closer to the array. The loudest profiler emissions were excluded from the injected data segments. For example, as defined in Eq. (5), the SNRs were −3.7 dB at 91.2 Hz, −10.6 dB at 105.6 Hz, −9.3 dB at 113.4 Hz, −10.5 dB at 122.1 Hz, and −6.9 dB at 150 Hz for subaperture 1. Similar levels were obtained for the other subapertures shown in Fig. 1(b).

The BTD in Fig. 9 was constructed in the same way as the one in Fig. 2(a) and shows the effect of injecting data snippets representing two interferers centered around bearings of −45° and −10° [see Figs. 1(a) and Fig. 2(a)]. In comparison with the BTD in Fig. 2(a), the interferers degrade the SSOS multipath correlations in Fig. 9 over bearings 55°–75° in the inversion data segment, which may adversely affect the inversion results.

The inversions of the ABF and CBF subaperture beam data containing the injected interferers were initialized with the parameter values from Table I and performed with the broad bounds also in Table I. Starting with a high annealing temperature, a (possibly new) global minimum was located after sufficient inversion iterations were performed. Some parameter cost distribution envelopes from inversions of data with interferers are shown in Fig. 6. Cost envelopes from inversions of data with two interferers show minima at the

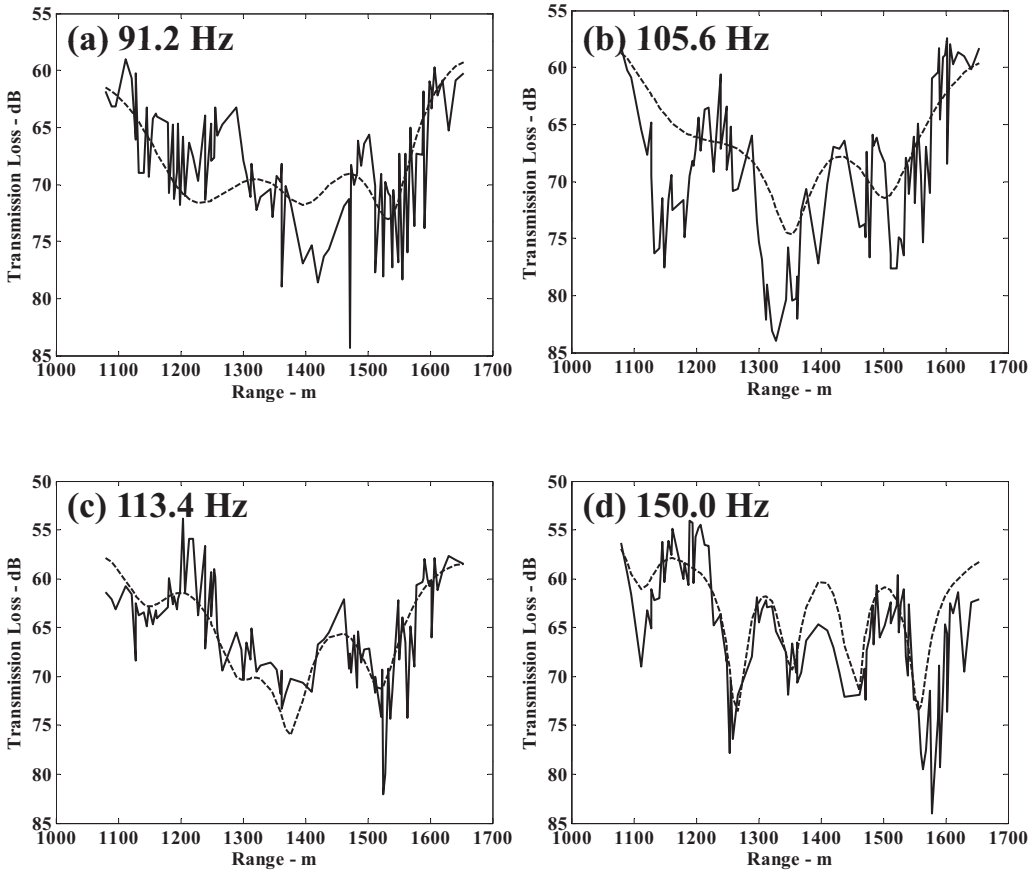


FIG. 7. Comparison of measured transmission loss (solid) with transmission loss calculated by a model (dash) with the parameter values from the solution of the inversion of the ABF data at (a) 91.2 Hz, (b) 105.6 Hz, (c) 113.4 Hz, and (d) 150.0 Hz.

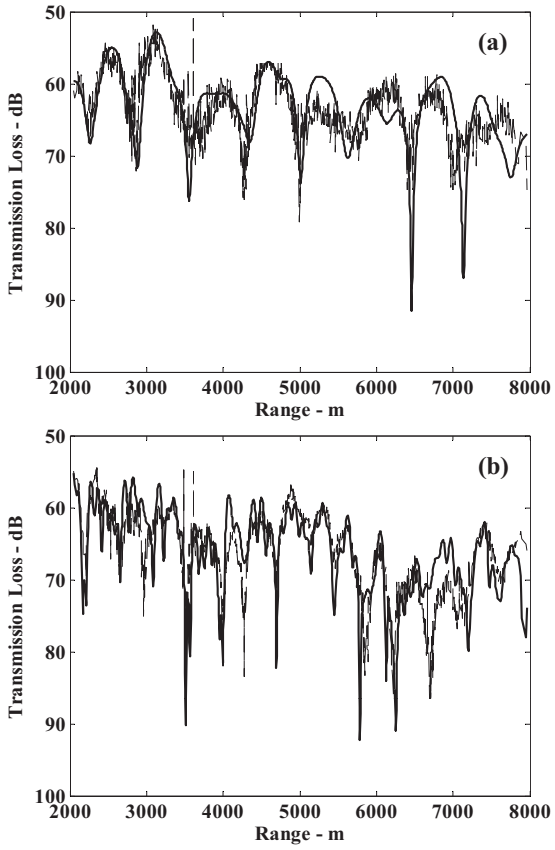


FIG. 8. Measured transmission loss from tow data (dash) compared to transmission loss calculated by a model (solid) with the parameter values from the solution of the inversion of ABF data at (a) 53 Hz and (b) 153 Hz.

same values as the distribution envelopes from the inversions of the data without interferers; however, the envelopes from the inversion of the CBF data with interferers are flat, which suggests that the parameter values in that solution have large uncertainty. In contrast, the envelopes obtained from the inversion of ABF data with interferers show well defined minima.

The measured RL structure from CBF and ABF track beam data containing two injected interferers are compared in Fig. 10 to the modeled RL ($10\log_{10}|S_f|-\mathrm{TL}$) calculated with the parameter values obtained from the solutions of the inversions of the data containing two interferers. Figure 10 shows that at the highest frequency the modeled RL struc-

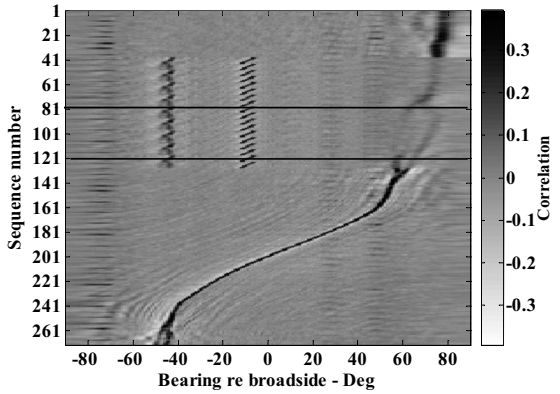


FIG. 9. Bearing time display with R/V Longhorn and two sources representing injected interferers replicated to span the inversion data sample.

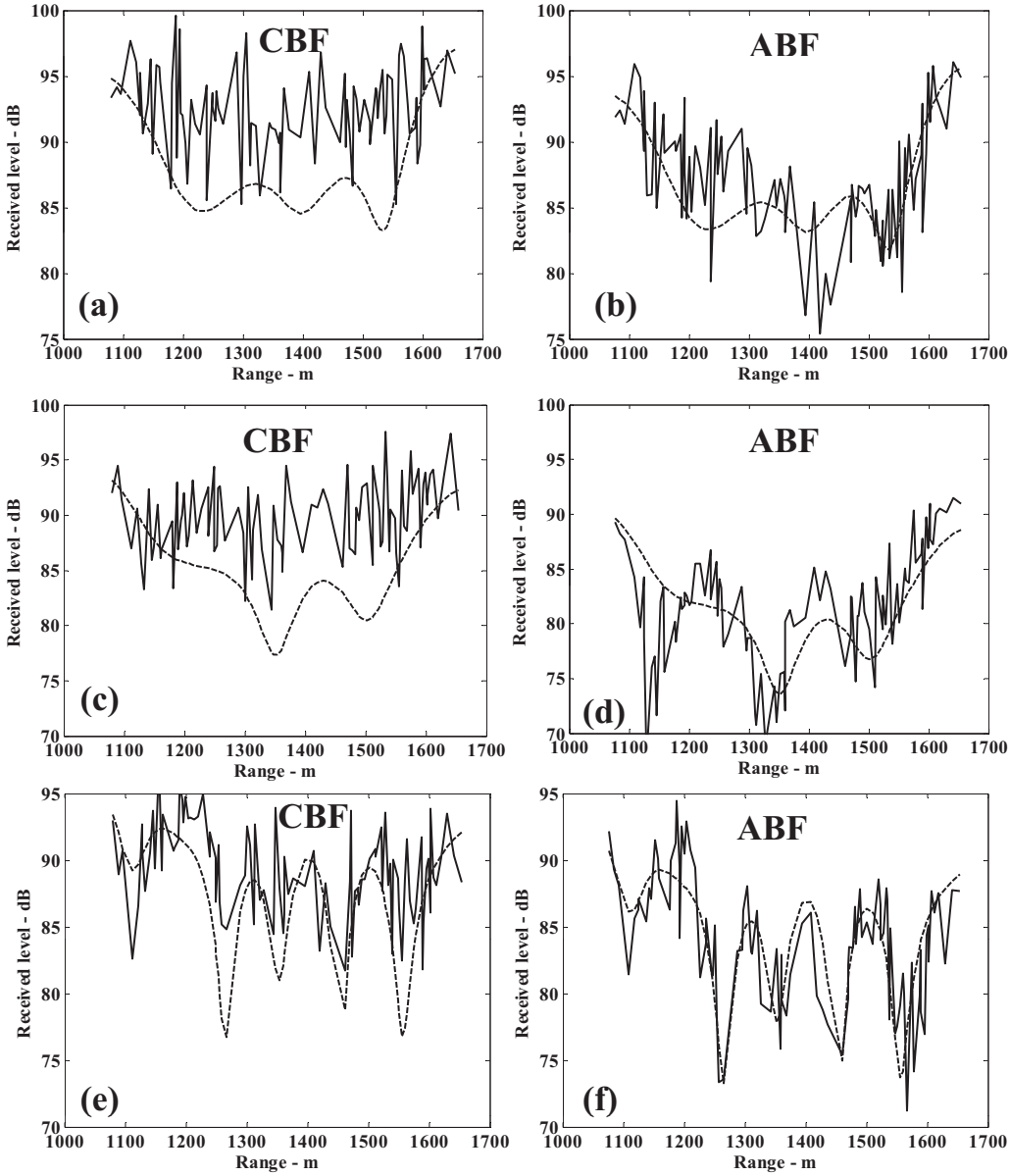


FIG. 10. Comparison of measured RL (solid) with RL calculated by a model (dash) with the parameter values from the solution of the inversion with two injected interferers for (a) CBF data at 91.2 Hz, (b) ABF data at 91.2 Hz, (c) CBF data at 105.6 Hz, (d) ABF data at 105.6 Hz, (e) CBF data at 150.0 Hz, and (f) ABF data at 150.0 Hz.

ture, produced from the parameters obtained with the CBF data solutions, exhibits multipath structure similar to that present in the modeled TL shown in Fig. 7 that was obtained with the parameter value solutions from inversions of CBF data without injected interferers. At the two lower frequencies the measured TL, i.e., the RL, from the CBF track beam data does not show the same multipath structure as the CBF and ABF data with no interferers. The CBF data contaminated with the two interferers lose much of the RL structure that is apparent with no injected interferers present, especially at the two lowest frequencies shown in Fig. 10. For the CBF data with two interferers, the broad cost envelopes shown in Fig. 6, along with the reduced RL structure, suggest that the values of the parameters obtained from the inversion solution have greater uncertainty than for the solution from ABF data.

The ABF track beam data containing two interferers (see Fig. 10) retains at each inversion frequency the structure shown in Fig. 7, which demonstrates the ability of the ABF to mitigate the effects of the interferers. The modeled RL produced from the solutions for inversions with the ABF track beam data containing two interferers (Fig. 10) retains approximately the same multipath structure at the two lower frequencies as the model TL produced from the solutions for inversions of the ABF track beam data without injected interferers. Both the CBF and ABF measured RL data in Fig. 10 for the highest frequency show most of the multipath structure that was exhibited in the measured TL in Fig. 7, although the CBF RL data structure in Fig. 10 contains more contamination from the interferers. Overall, the modeled RL produced from the solutions for the inversions of the CBF and ABF data containing two injected interferers reproduce the data RL, but the quantitative agreement for the ABF data is better due to the ability of the ABF to suppress the interferers.

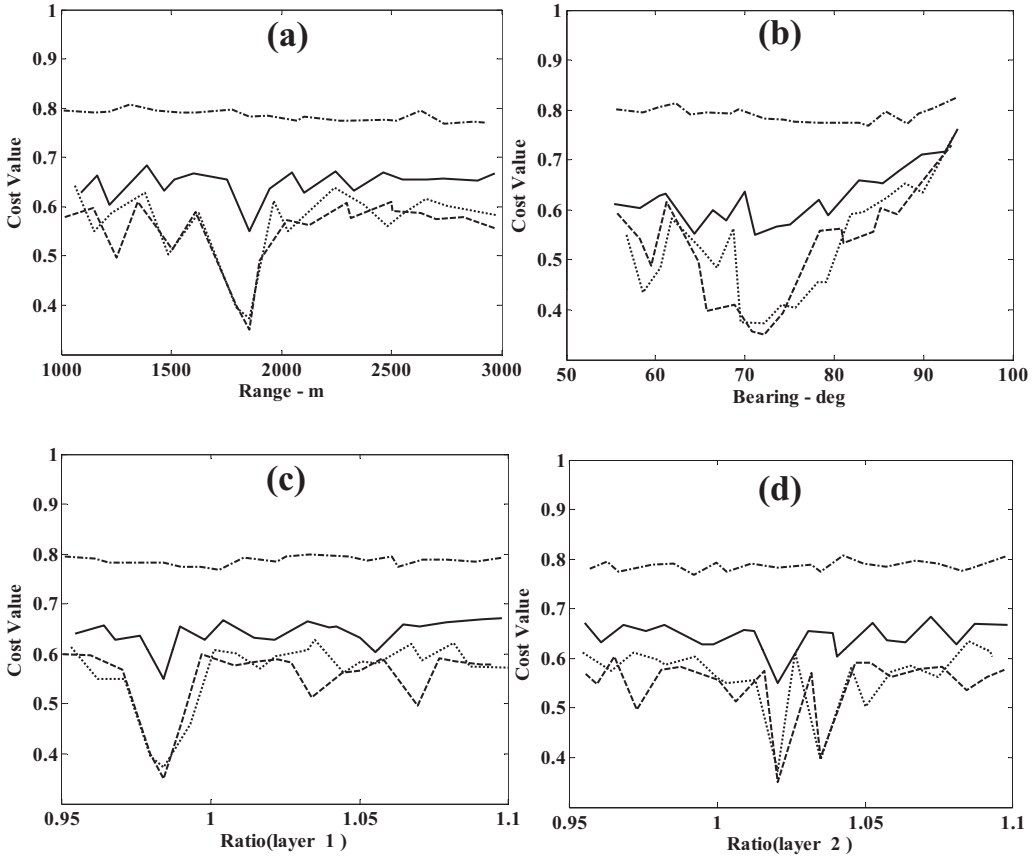


FIG. 11. Cost distribution minimum envelopes for initial SSOS (a) range and (b) bearing and for the sound speed ratio for (c) layer 1 and (d) layer 2, obtained from inversions of the CBF (dash-dot) and ABF (solid) data with three interferers present and from inversion of data without interferers for CBF (dash) and ABF (dot) data.

B. Three injected interferers

Next, the inversion process was repeated with interferer data injected from all three segments shown in Figs. 1(a) and 2(a). The SNRs defined in Eq. (5) for three interferers were −8.9 dB at 91.2 Hz, −11.5 dB at 105.6 Hz, −11.3 dB at 113.4 Hz, −12.8 dB at 122.1 Hz, and −12.0 dB at 150 Hz for subaperture 1. These SNR values are considerably lower than those in Sec. VII A for two interferers at most of the inversion frequencies.

The cost distribution envelopes from the solutions of inversions with CBF and ABF data with three interferers are shown in Fig. 11. The cost envelopes from the inversion of the CBF data have a high cost minimum at incorrect parameter values, and the cost distribution envelopes are mostly flat. The cost distribution envelopes from the ABF data inversions have the minimum at the correct solution; however, the envelopes are somewhat broader, and the cost minimum has increased compared to the cost minimum from the inversion of data with only two interferers.

The measured RL structure from the data containing three interferers is shown Fig. 12. For the CBF track beam data the measured RL at each frequency is obliterated in the CBF data by the effects of the interferers. In comparison, the measured RL from the ABF track beam data retains much of the multipath structure shown in Fig. 7, except at the higher frequencies, especially 150 Hz, which do show contamination from the interferers. These results suggest that there are limits on the capabilities of both CBF and ABF to suppress interferer noise for short subapertures.

The modeled $\mathrm{RL}=10\log_{10}|S_f|-\mathrm{TL}$ produced from the solutions of the inversions with CBF track beam data containing three injected interferers (Fig. 12) reflects the poor inversion solution obtained, i.e., the modeled RL is nearly flat. The modeled RL produced by the solutions from inversions with the ABF data containing three interferers retains most of the multipath structure apparent in the modeled TL shown in Fig. 7. Consequently, the modeled RL structure produced with the solutions from the ABF data containing three interferers is the same as in the TL structure in Fig. 8, but the modeled RL produced with the inversion solutions from the CBF data with three interferers cannot reproduce the RL structure of the tow line data, i.e., it produces RL with no structure for each frequency. This is because the SNR in the CBF data with three interferers was too low, and the inversion produced a substrate sound speed too small to support enough propagating modes in the waveguide to accurately reproduce measured RL structure.

C. Seismic profiler interferers

Inversions from data samples containing the seismic profiler emissions were performed to ascertain whether ABF processing would extend the usable data to lower frequencies than CBF processing during periods when the seismic profilers were active. If so, perhaps more geoacoustic information about the deeper sediment layers could be obtained even when profilers were active.

At several frequencies between 50 and 100 Hz inversion

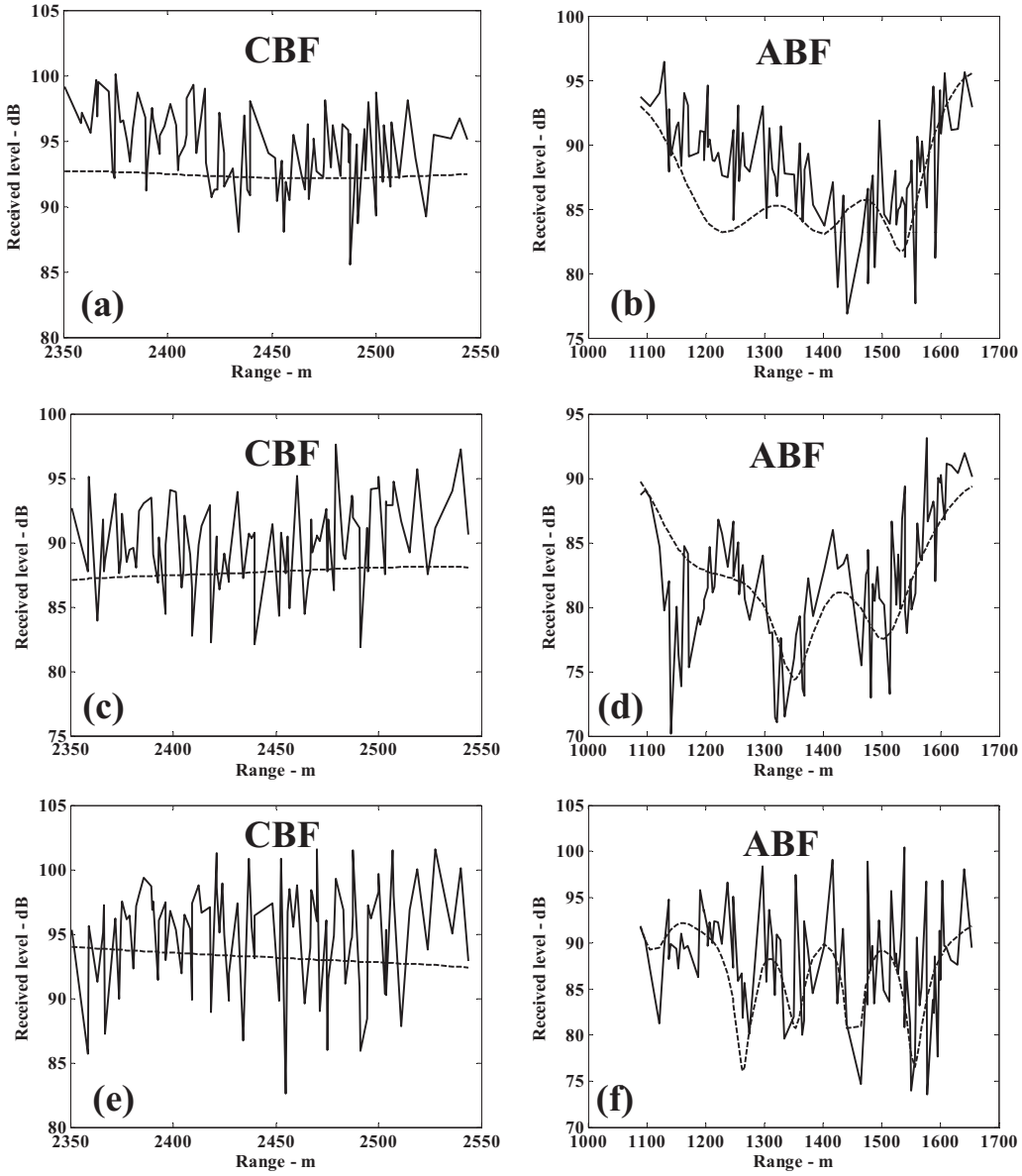


FIG. 12. Comparison of measured RL (solid) with RL calculated by a model (dash) with the parameter values from the solution of the inversion with three injected interferers for (a) CBF data at 91.2 Hz, (b) ABF data at 91.2 Hz, (c) CBF data at 105.6 Hz, (d) ABF data at 105.6 Hz, (e) CBF data at 150.0 Hz, and (f) ABF data at 150.0 Hz.

data without injected interferers were selected, specifically 53.4, 59.4, 68.4, and 88.7 Hz. An inversion was performed with only the data from these frequencies. The suppression of the seismic profiler interference in the ABF data was more notable at these frequencies than in the CBF data; however, some of the profiler interference still remained in the ABF data, although with a reduced level as compared to the CBF data. The ABF data inversion solution produced a global minimum cost value of 0.685, which was less than the cost minimum (0.81) from the CBF data inversion. The inversion of the ABF data from the low frequencies did produce the solution obtained with the higher frequency data (see Sec. VI), but with broader cost distribution envelopes. Despite the lack of direct information pertaining to the deeper layers, the modeled TL obtained with the parameter values from the inversion solution did match the data at the lower frequencies, which gives more confidence in the inversion solution obtained from only the higher frequency data. The inversion from CBF data with only the low frequencies gave a different (incorrect) solution with an inaccurate source localization and a water-sediment interface sound speed ratio >1.0, in contradiction to the known sediment properties in the area. This result suggests that inversion processing with ABF data from frequency bands that are inaccessible in CBF data is possible.

VIII. CONCLUSIONS

This paper has shown that accurate inversion solutions (for both geoacoustic and source track parameters) can be obtained from HLA subaperture ABF data. The inversion results from ABF data are similar to those from CBF subaperture data which contain a single SSOS and possibly low level interferers. The successful performance of inversions

with ABF data when multiple interferers at high levels are present was also established. The failure to obtain consistently accurate solutions from inversions of CBF data containing multiple interferers at high levels demonstrates the potential advantage of performing inversions with HLA subaperture ABF data. Finally, seismic profiler interference was suppressed by ABF for frequencies below 100 Hz, as compared to CBF, and demonstrates the possibility of performing inversions from ABF subaperture data at lower frequencies than from CBF data. Increasing the inversion data bandwidth to lower frequencies offers an advantage in that it provides additional data for inversion solution validation.

ACKNOWLEDGMENTS

This work was supported by the Independent Research and Development Program of Applied Research Laboratories, The University of Texas at Austin. The authors appreciate insightful discussions with Dr. D. P. Knobles.

[1]H. Cox, R. M. Zeskind, and M. M. Owen, "Robust adaptive beamforming," IEEE Trans. Acoust., Speech, Signal Process. **35**, 1365–1376 (1987).
[2]T. R. Messerschmitt and R. A. Gramann, "Evaluation of the dominant mode rejection beamformer using reduced integration times," IEEE J. Ocean. Eng. **22**, 385–392 (1997).
[3]T. M. Redheendran, "Initial evaluation of the dominant mode rejection beamformer," J. Acoust. Soc. Am. **104**, 1331–1344 (1998).
[4]N. O. Booth, A. T. Abawi, P. W. Schey, and W. S. Hodgkiss, "Detectability of low-level broad-band signals using adaptive matched-field processing with vertical aperture arrays," IEEE J. Ocean. Eng. **25**, 296–313 (2000).
[5]A. 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).
[6]H. Schmidt, A. B. Baggeroer, W. A. Kuperman, and E. K. Scheer, "Environmentally tolerant beamforming for high-resolution match field processing: deterministic mismatch," J. Acoust. Soc. Am. **88**, 1851–1862 (1990).
[7]C. Soares, S. M. Jesus, and E. Coelho, "Environmental inversion using high-resolution matched-field processing," J. Acoust. Soc. Am. **122**, 3391–3404 (1997).
[8]K. Yoo and T. C. Yang, "Adaptive matched beam processing for source localization," J. Acoust. Soc. Am. **106**, 2127 (1999).
[9]S. A. Stotts, "A robust spatial filtering technique for multisource localization and geoacoustic inversion," J. Acoust. Soc. Am. **118**, 139–162 (2005).
[10]R. A. Koch and D. P. Knobles, "Geoacoustic inversion with ships as sources," J. Acoust. Soc. Am. **117**, 626–637 (2005).
[11]R. B. Newman, "Simultaneous inversion for source track and sediment parameter values from acoustic signals generated by a ship of opportunity in a soft sediment shallow water waveguide," MS thesis, The University of Texas at Austin (2005).
[12]S. Stotts and B. Hawkins, "Application of adaptive beamforming to geoacoustic inversion and source localization," J. Acoust. Soc. Am. **118**, 1857 (2005).
[13]D. P. Knobles, R. A. Koch, L. A. Thompson, and P. E. Eisman, "Broadband sound propagation in shallow water and geoacoustic inversion," J. Acoust. Soc. Am. **113**, 205–222 (2003).
[14]E. A. Ritterman, "Geoacoustic inversion in multiple source environments," MS thesis, The University of Texas at Austin (2004).
[15]D. E. Grant, J. H. Gross, and M. Z. Lawrence, "Cross-spectral matrix estimation effects on adaptive beamforming," J. Acoust. Soc. Am. **98**, 517–524 (1995).
[16]S. Kirkpatrick, C. D. Gelatt, Jr., and M. P. Vecchi, "Optimization by simulated annealing," Science **220**, 671–680 (1983).
[17]W. L. Goffe, G. D. Ferrier, and J. Rogers, "Global optimization of statistical functions with simulated annealing," J. Econometr. **60**, 65–100 (1994).
[18]S. A. Stotts, R. A. Koch, S. M. Joshi, V. T. Nguyen, V. W. Ferreri, and D. P. Knobles, "Geoacoustic inversions of horizontal and vertical line array acoustic data from a surface ship source of opportunity," IEEE J. Ocean. Eng. In press (2010).
[19]R. A. Koch and J. E. LeMond, Software requirements specification for the NAUTILUS model (Version 1.0), ARL-TL-EV-01-18, April (2001).
[20]R. A. Koch and J. E. LeMond, Software test description for the NAUTILUS model (Version 1.0), ARL-TL-EV-01-19, April (2001).
[21]R. A. Koch and J. E. LeMond, Software design description for the NAUTILUS model (Version 1.0), ARL-TL-EV-01-17, April (2001).
[22]R. B. Newman and R. A. Koch, Software description for the NAUTILUS subroutine, ARL-TL-EV-03-09, April 18 (2003).
[23]S. A. Stotts, D. P. Knobles, R. A. Koch, D. E. Grant, K. C. Focke, and A. J. Cook, "Geoacoustic inversion in range-dependent ocean environments using a plane wave reflection coefficient approach," J. Acoust. Soc. Am. **115**, 1078–1102 (2004).
[24]S. E. Dosso, "Quantifying uncertainty in geoacoustic inversion. I. A fast Gibbs sampler approach," J. Acoust. Soc. Am. **111**, 129–142 (2002).
[25]S. E. Dosso and P. L. Nielsen, "Quantifying uncertainty in geoacoustic inversion. II. Application to broadband, shallow-water data," J. Acoust. Soc. Am. **111**, 143–159 (2002).

Observations and modeling of angular compression and vertical spatial coherence in sea surface forward scattering

Peter H. Dahl[a)]
Applied Physics Laboratory and Department of Mechanical Engineering, University of Washington, 1013 NE 40th Street, Seattle, Washington 98105

(Received 22 December 2008; revised 12 October 2009; accepted 24 October 2009)

Measurements and modeling of spatial coherence and related angular spreading associated with forward scattering from the sea surface are presented. The measurements were taken in waters 80 m deep off the New Jersey coast in August 2006. Acoustic signals from a source at depth 40 m were recorded on a vertical line array of length 1.4 m, centered at depth 25 m, and at range 200 m. Measurements in the 14–20-kHz frequency range are reported; the rms waveheight H was 0.16 m, setting kH as ~10 where k is acoustic wavenumber. A systematic study of measurements taken over four source–receiver bearing angles separated by 90° suggests a null influence of changing bearing angle or equivalently directional wave effects. Sound speed was characterized by a downward-refracting profile. Refraction modifies the vertical angular spread due to rough sea surface scattering, which can be understood from Snell's law. The Snell mapping is smooth, so an approximation based on the mean grazing angle provides a functional relation between the angular variance near the surface and that at the receiver. The latter is measurably reduced owing to refraction, the effect called angular compression, and a parameter that quantifies this effect is defined. © *2010 Acoustical Society of America.* [DOI: 10.1121/1.3268594]

PACS number(s): 43.30.Re, 43.30.Zk, 43.30.Hw [DRD] Pages: 96–103

I. INTRODUCTION

Forward scattering from the sea surface is important in the performance of many underwater acoustic applications, e.g., communication,[1,2] limits of broadband sonar performance,[3] higher frequency tomography,[4] and adaptive beam forming.[5] For sound frequencies of O(10) kHz and above, the term scattering rather than reflection generally applies because for typical sea surfaces $kH \gg 1$, where H is rms waveheight and k is acoustic wavenumber. Here the complex vertical spatial coherence of sound forward scattered from the sea surface is studied, with emphasis on the role of refraction caused by a thermocline in determining coherence magnitude and phase in addition to the expected dependence on rough surface properties. Modeling results of vertical spatial coherence are compared to field data that originate from the Shallow Water 06 (SW06) acoustic experiments conducted off the coast of New Jersey[6] in August 2006. The term spatial coherence as used here is a dimensionless quantity equal to the normalized second moment of the acoustic field computed between two receivers with separation d [7–9] and zero time lag. Implicit in this definition is that the received field be narrow band, which is typically satisfied in practice for experimental data characterized by a relative bandwidth $\ll 1$; thus there is an inverse relation between angular spreading of the received field and spatial coherence.[10]

Angular spreading is inherent to the process of forward scattering by a rough sea surface or boundary regardless of whether refraction occurs over the propagation path.[7] The measurements of vertical spatial coherence reported here correspond to propagation via a single sea surface bounce path or channel. Estimates of the magnitude of vertical spatial coherence $|\hat{\Gamma}|$ for such a channel decay monotonically with vertical receiver separation,[9] and a vertical coherence length d^* can be defined, e.g., where $|\hat{\Gamma}|$ reaches $e^{-1/2}$, such that $1/kd^*$ is proportional to the vertical angular spread. Note that in a shallow water, multipath channel such a monotonic decay in the coherence magnitude is typically manifested only at long range, i.e., large range-to-depth ratios.[11] The coherence length d^* in a given direction due to angular spreading induced by the ocean environment determines the effective aperture length for directional discrimination of a similarly oriented array.[10,12] In terms of specific vertical coherence, this measure also influences the communication data rate for both coherent[1] and time-reversal[13] processing schemes that utilize vertical receiving arrays.

Forward scattering is quantified in terms of the sea surface bistatic cross section per unit area σ, which is evaluated here using the small slope approximation[14–16] following the two-dimensional implementation by Dahl.[17] A model for vertical spatial coherence (or coherence along other directions) using a two-dimensional distribution of σ, combined with propagation to and from the sea surface, is derived using the van Cittert–Zernike theorem (vCZT).[9] In radio astronomy, the vCZT relates the spatial coherence function gathered from an array of telescopes to the Fourier transform of the brightness (intensity) distribution for the celestial object emitting the radiation.[18] In this application, the sea surface is taken to be an incoherently radiating source upon activation by an incident sound field, with source distribution constructed from σ. A second, related, modeling approach, motivated by the SW06 observations of a downward-refracting

a)Electronic mail: dahl@apl.washington.edu

ocean environment, is based on a probability density function (PDF) for vertical arrival angle that is constructed from the same bistatic scattering distribution. This approach clarifies the role of refraction and Snell's law in determining the coherence magnitude and phase.

Two key results are presented, the first being that measurements made over four source–receiver bearing angles separated by 90°, during which average directional wave conditions were stationary, are not statistically different. The directional wave properties have no observable influence on vertical spatial coherence at frequencies >14 kHz. The second concerns the extent to which the typical SW06 sound speed profile influences vertical spatial coherence. Downward refraction compresses the vertical angular spread that is caused by rough sea surface scattering and sensed by a receiver well below the sea surface. This angular compression produces a change in spatial coherence magnitude that is on the level of, or exceeds, that associated with a large change in sea surface roughness conditions.

Section II describes the experiment and environmental conditions that are relevant to the bistatic modeling. In Sec. III data analysis methods are described along with a detailed specification of bias and variance associated with the estimate of coherence magnitude. Modeling results and model-data comparisons are presented in Sec. IV. These are discussed in Sec. V, where an alternative modeling approach that highlights the role of refraction is described; conclusions are given in Sec. VI.

II. EXPERIMENT DESCRIPTION

A. Acoustic measurements

The acoustic measurements were made from aboard the research vessel *Knorr*. An omni-directional acoustic source was deployed from the stern of the R/V *Knorr* to depth 40 m; signals transmitted from the source were recorded on a moored receiving array (MORAY) system with remotely changeable receiving configurations. The MORAY was placed at position 39.0245°N, 73.0377°W, a location (depth 80 m) that defined the central site for SW06 experimental observations in the mid-frequency range. The acoustic source–receiver geometry was established by the position (station) of the R/V *Knorr*, referenced to its stern, with precise station-keeping achieved through its dynamic positioning system. Results presented here are from the 1.4-m vertical line array centered at depth 25 m and composed of four omni-directional hydrophones with separations of 0.2, 0.3, and 0.9 m. The MORAY was held vertical by a spherical float of diameter 1 m, submerged at depth 10 m, to produce a vertical and fairly stiff orientation for the line array. Instrumentation on the array recorded various metadata, the most relevant to this study being array tilt and heading (array inclination).

The measurements were made on 10 August 2006 between 0830 and 1500 UTC at four stations each at range 200 m from the MORAY and separated in direction between source and receiver by 90°. The four stations in source–receiver bearing angle were, in order of their sampling, 210°, 120°, 30°, and 300°. The transmitted signal was a 3-ms continuous wave (cw) pulse for which center frequencies of 14–20 kHz (every 2 kHz) were superimposed and transmitted simultaneously; a particular center frequency was recovered and analyzed in postprocessing (acoustic data sampled at 50-kHz sampling rate). The 3-ms pulse was transmitted 20 times; after each transmission, signal types designed for other experiments were transmitted resulting in a particular signal repetition period of 8 s. Upon completion of all transmissions, data were uploaded from the MORAY via telemetry. A second 20-transmission set was undertaken after which the R/V *Knorr* established position at the next source station to continue with the experiment. Though measurements in the frequency range 1–12 kHz were made simultaneously with the 14–20-kHz measurements, discussion here is limited to the higher frequencies.

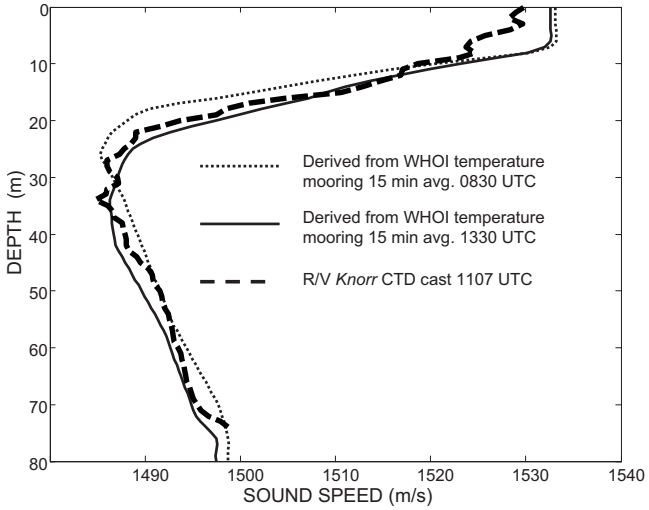


FIG. 1. Nominal sound speed conditions associated with the acoustic measurements made on 10 August 2006 between 0830 and 1500 UTC as represented by measuring a conductivity-depth-temperature (CTD) profile deployed from the R/V *Knorr* at 1107 UTC and derived from a nearby temperature mooring (Woods Hole Oceanographic Institution).

B. Environmental measurements

The sound speed conditions were characterized by a downward-refracting sound speed profile (Fig. 1), i.e., sound speed going from 1530 m/s at the surface to 1485 m/s at depth 25 m. Sea surface directional wave measurements were made using a 0.9-m-diameter TRIAXYS directional wave buoy deployed 800 m from the MORAY site and also by the Air-Sea Interaction Spar (ASIS) buoy[19] deployed by the University of Miami 1.5 km from the MORAY site. Figure 2(a) shows a comparison of non-directional spectra as determined from these two wave measurement systems, time averaged for the period of 0830–1500 UTC. The rms wave height H was 0.16 m with this value differing by no more than ±10% over the measurement period. Figure 2(b) plots the time-averaged directional wave spectrum estimated from the TRIAXYS buoy, showing a swell-wave system originating from the south and a wind-wave system originating from the southwest. The ASIS buoy also recorded wind speed at 6 m/s (±1 m/s) over the entire 6.5-h measurement period.

Detailed analysis of the 10 August conditions shows that no statistically significant bubble-related attenuation in sea

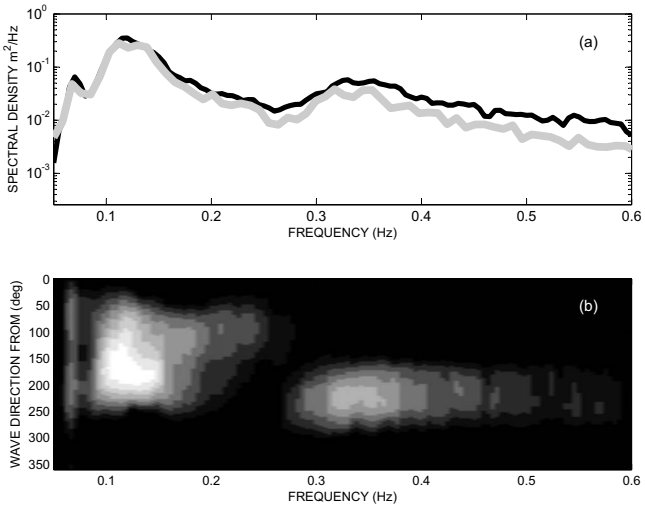


FIG. 2. (a) Time average (0830–1500 UTC) of the non-directional wave spectrum as measured 800 m from the MORAY system with wave buoy (black line) and 1500 m from MORAY with the ASIS buoy (gray line). (b) Equivalent time-averaged directional wave spectrum as measured 800 m from MORAY with wave buoy.

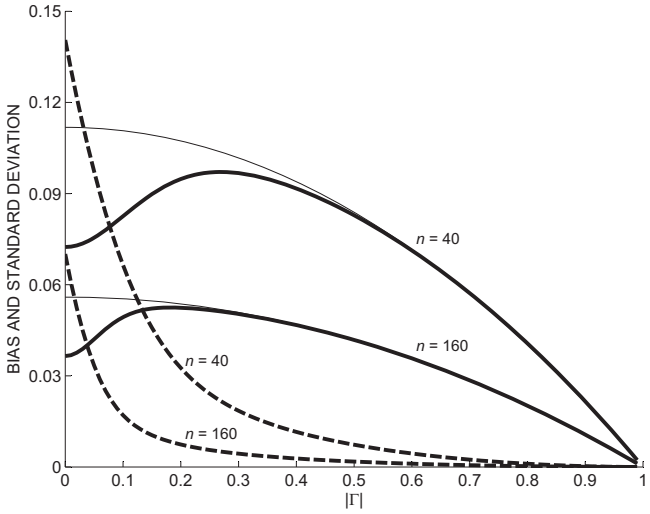


FIG. 4. Bias (thick, dashed line) and square root of variance (thick, solid lines) for estimates of magnitude coherence as a function of equivalent theoretical or population value $|\Gamma|$ for $n=40$ and 160 samples based on equations of Carter *et al.* (Ref. 22). The approximation for variance (also in square root form) corresponding to $n=40$ and 160 samples is shown by the thin lines.

surface forward scattering[20] was observed over the frequency range of 1–20 kHz. Bubbles have a minor role in the forward scattering process for this acquisition geometry, and the bubble void fraction was estimated to be less than 10^{-8}; such a void fraction, however, would be observable in high-frequency backscattering observations.

III. COHERENCE ESTIMATION

The 3-ms pulse length of the quasi-narrow band signals is sufficiently short to permit time resolution of individual arrivals (Fig. 3), yet sufficiently long such that the relative bandwidth remains ~0.02 or less, and the data are considered narrow band and represented by the center frequency of the pulse. Estimates of complex coherence $\hat{\Gamma}_{pq}$ between vertically separated receivers p and q are made according to an averaging method using the complex narrow band envelope.[9] A steady-state region of the surface bounce arrival is identified [near the triangle symbol in Fig. 3(b)]. Errors result if arrival content from either before or after the steady-state region is entered into the average, so the set of pulses is first time aligned to remove any jitter that can offset a given pulse arrival from the identified sample region. The alignment time shift is determined from one channel and applied in the same manner to the other three channels. Therefore this time shifting has no unwanted effect on the magnitude and phase of the estimates.

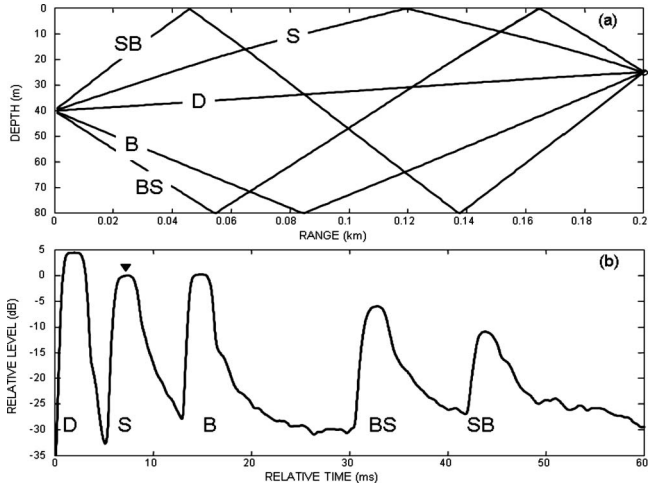


FIG. 3. (a) Ray diagram based on CTD cast made on 10 August at 1107 UTC (see Fig. 1), showing various arrivals for a source at depth 40 m, receiver at depth 25 m, and range 200 m. Arrivals are the direct path (D), surface path (S), bottom path (B), bottom-surface (BS) path, and surface-bottom (SB) path. (b) Time series of received level for 3-ms pulse of center frequency of 16 kHz based on the average of 20-ping transmissions made on 10 August at 10:00 UTC. The mark above the surface path (S) denotes region of steady-state return, and the time window within which coherence is estimated is centered here.

Bias and variance are important features of this estimate, particularly for low values (<0.3) of coherence magnitude that are presented and modeled in this study. First, an approximation[21] for the standard deviation of coherence magnitude is $(1-|\Gamma|^2)/\sqrt{2n}$, where Γ is the theoretical or population value for complex coherence and n is the number of ping receptions (independent measures) entering the average. (The $1/\sqrt{2}$ factor was mistakenly not included in Ref. 9, although with little impact as the resulting approximate expression was used only to define nominal uncertainty.) Second, Carter *et al.*[22] derive exact expressions for bias and variance of coherence magnitude for which the square root of variance follows the approximate behavior for large n and $|\Gamma|$ greater than ~0.3. For smaller magnitudes, however, the approximation's expected monotonic increase with decreasing $|\Gamma|$ no longer holds and variance begins to decrease (Fig. 4). This curious property is reproduced by a numerical evaluation of the equations for bias and variance of coherence magnitude for $n=40$ and 160 independent realizations (Fig. 4). In the estimates of coherence magnitude presented here, error bars correspond to ±1 standard deviation and the esti-

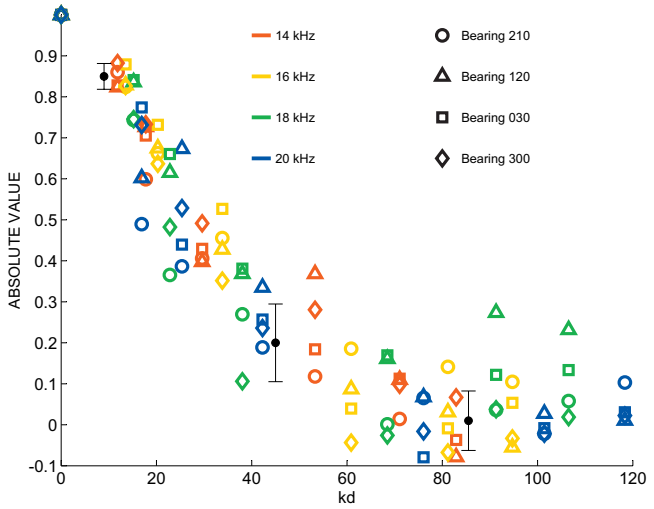


FIG. 5. Absolute value of vertical spatial coherence $|\Gamma|$ plotted as function of normalized receiver separation kd based on center frequencies of 14–20 kHz every 2 kHz. The different center frequencies are color-coded and symbols correspond to the four source–receiver bearing angles identified in the legend representing four source locations. Representative error bars are based on 40-ping average.

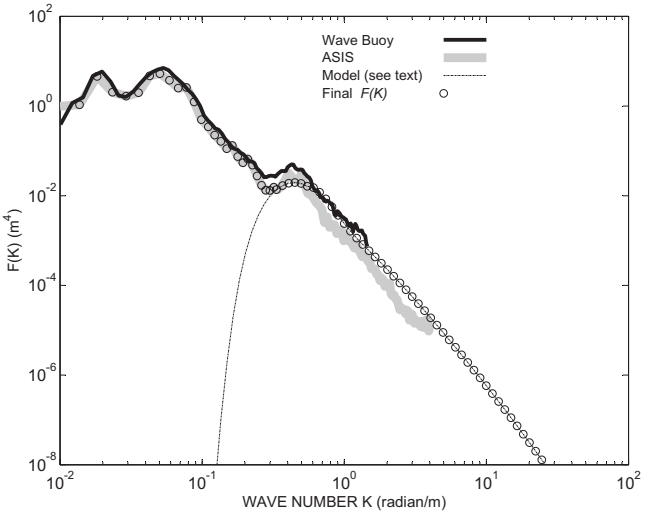


FIG. 6. The directionally averaged wavenumber spectrum $F(K)$ used to compute the bistatic cross section of the sea surface (circles). This is based on a combination of measured wave data (thick gray and black lines) and a model (Ref. 23) associated with wind waves, assuming a wind speed of 6 m/s and 30-km fetch (thin line).

mates have been corrected for bias; both quantities are based on the n-dependent exact expressions. These quantities also depend on a theoretical, or otherwise nominal, value of $|\Gamma|$ (Fig. 4), for which model estimates of $|\Gamma|$ as derived in Sec. IV B are used.

IV. RESULTS

A. The study over four source–receiver directions

Estimates of vertical spatial coherence from the four stations separated by 90° in source–receiver bearing (Fig. 5) correspond to the surface path geometry shown in Fig. 3. Data are plotted as a function of a normalized vertical receiver separation kd, where d is separation and k is acoustic wavenumber for frequencies of 14–20 kHz every 2 kHz. For each station the transmission set was sent twice, and thus there are 40 independent realizations of the sea surface sensed by the multi-frequency cw pulse.

Two conclusions emerge from data analysis. First, coherence magnitude estimates for frequencies 14–20 kHz tend to collapse onto one, monotonic decay curve that is a function of kd (Fig. 5). Second, estimates of spatial coherence are not influenced in any statistically significant way by the changing source–receiver bearing angle that spanned 360° during the measurement period on 10 August. These conclusions direct the subsequent analysis in three ways: (1) the measurements made over the four stations between 0830 and 1500 are averaged to reduce statistical uncertainty (estimates of coherence are re-computed based on $n=160$ transmissions rather than averaging station-curves based on $n=40$ transmissions), (2) results from 14 to 20 kHz are combined to improve the sampling in kd space, and (3) results are modeled based on a directionally averaged wave spectrum (Fig. 6).

B. Modeling and comparison with averaged measurements

The sea surface grazing angle θ_g, computed using the SW06 sound speed profile (Fig. 1), is ~15° for the specular ray associated with the surface bounce path [Fig. 3(a)]. This puts the roughness parameter $\chi=kH \sin\theta_g$ at ~5 for a frequency of 14 kHz. In view of this and the scaling of the vertical coherence data with parameter kd (Fig. 5), a model for spatial coherence based on incoherent bistatic scattering from the sea surface is used. The dimensionless bistatic scattering cross section per unit area, $\sigma(\xi,\eta)$, is computed using the small slope approximation[17] as a function of position on the sea surface as defined by cross-range variable ξ and down-range variable η at frequency 20 kHz. An estimate of the directionally averaged sea surface wavenumber spectrum $F(K)$ is necessary for scattering calculations and is generated (Fig. 6) using a combination of the measured wave data and a surface wave model by Plant.[23] The model contribution provides information for higher wavenumbers needed for $F(K)$ but not sensed by the ASIS buoy.

The strong thermocline and ensuing refraction is incorporated into modeling by using the SW06 profile of 1107 UTC (Fig. 1); modeling is also carried out using an isovelocity profile for comparison. Note that measurements of vertical spatial coherence over similar acquisition geometries were made in the East China Sea as part of ASIAEX,[9,24] where the primary influence of refraction was to produce a small change in the phase of vertical coherence, but little change in coherence magnitude. This is understood in part in the context of Fig. 1, e.g., the entire sound speed vs depth variation (applicable to the measurements) for the late spring environment in the East China Sea was 1525–1530 m/s. The more strongly downward-refracting sound speed profile characteristic of SW06, however, changes the incident and scattered angles that determine the level of the bistatic scattering cross section and compresses the set of vertical arrival angles

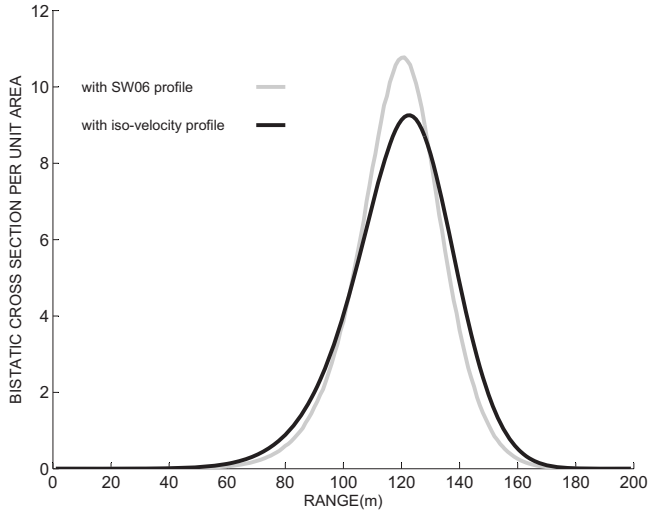


FIG. 7. The dimensionless bistatic cross section per unit area as computed along the line on the sea surface connecting source with receiver (bistatic angle of 0°), incorporating refraction effects (gray line) compared with iso-velocity conditions (black line). Calculations are based on the $F(K)$ in Fig. 6 and frequency of 20 kHz. The two lines differ near their maximum by ~0.8 dB. The source depth is 40 m, receiver depth is 25 m, and the range is 200 m.

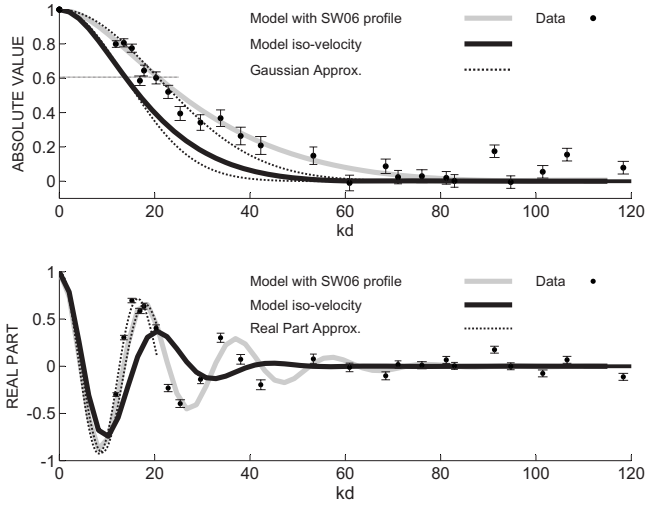


FIG. 8. (a) Estimates of the absolute value of vertical spatial coherence with error bars based on 160-ping average plotted as function of normalized receiver separation kd for center frequencies of 14, 16, 18, and 20 kHz. Solid lines represent a model for the absolute value based on the SW06 profile or iso-velocity profile. Two dotted lines represent Gaussian approximations for coherence magnitude as discussed in the text. The horizontal line corresponds to magnitude coherence equal to $e^{-1/2}$. (b) Estimates as in (a), but for the real part of vertical spatial coherence [error bars are multiplied by $\sqrt{2}$]. Two dotted lines represent approximations for coherence real part as discussed in the text.

at the receiver. Both changes in scattering and refraction ultimately influence vertical spatial coherence magnitude, although the latter has a much greater effect.

In terms of the bistatic forward scattering from the sea surface, the downward refraction tends to increase the maximum scattering level in the vicinity of the specular point compared to iso-velocity conditions. The effect is shown in Fig. 7 where the bistatic cross section, $\sigma(0,\eta)$, is plotted as a function of down-range variable η with the cross-range variable ξ set to 0 (corresponding to in-plane scattering); this is about 1.5 dB greater for the SW06 profile than the equivalent for the iso-velocity profile when using the $F(K)$ from Fig. 6 in the calculations. This difference is due in part to the reduction in χ for the refracting case, as θ_g is ~18° for the iso-velocity case compared with ~15° for the refracting case, resulting in a slightly larger value of χ and therefore lower scattering value.

The quantity $\sigma(\xi,\eta)/R_1(\xi,\eta)^2$, where $R_1(\xi,\eta)$ is the path corresponding to propagation between the source (depth 40 m) and a position on the sea surface, is used as a surrogate for the intensity distribution on the sea surface [for each value of $\sigma(\xi,\eta)$, an area of 1 m^2 is assumed to apply]. Similarly, the received intensity used in normalization is taken as this quantity divided by $R_2(\xi,\eta)^2$, where $R_2(\xi,\eta)$ corresponds to propagation from a position on the sea surface to a receiver. Both R_1 and R_2, along with the vertical arrival angle θ_V, are computed with ray theory to account for refraction (based on the 1107 UTC sound speed profile in Fig. 1). Using this intensity distribution on the sea surface and propagation from sea surface to receiver, the vCZT is applied to obtain a model for vertical spatial coherence[9] in terms of the parameter kd, where wavenumber k is taken to be the mean wavenumber[25] associated with the sound speed profile and modeling frequency (20 kHz). Note that R_1, R_2 (as used in normalization), θ_V, and σ are computed based on a single receiver at depth 25 m, as these vary little over the maximum change in receiver depth (1.4 m). The *difference* in path $R_2(\xi,\eta)$ between an intensity element on the sea surface and two receivers separated by depth d is important and is assumed to be equal to $d \sin \theta_V(\xi,\eta)$.

Two models for vertical coherence (Fig. 8), one based on the SW06 sound speed profile (Fig. 1) and the other based on iso-velocity conditions, are compared with the combined measurements of vertical spatial coherence from all four stations such that $\hat{\Gamma}$ is estimated based on $n=160$ observations, which further suppresses the bias and variance. The agreement between estimates $\hat{\Gamma}$ and model Γ based on the SW06 profile is evident for both coherence magnitude [Fig. 8(a)] and phase [Fig. 8(b)]. Taking kd^* as the point where each model absolute value reaches $e^{-1/2}$ gives 13.8 and 20.6 for the iso-velocity and SW06 (refraction) models, respectively. Gaussian approximations to the model curves, given by coherence magnitude equal to $\exp(-(kd/kd^*)^2/2)$, are shown by the two dashed lines [Fig. 8(a)] that coincide exactly with the models at kd^*. These along with the data illustrate that a Gaussian approximation for vertical coherence magnitude is a reasonable one for normalized separations less than kd^*; for higher kd, the magnitude assumes a different behavior.

Information concerning phase is displayed by the real part of vertical coherence [Fig. 8(b)]. For the model curves, the real part is approximated by $\mathrm{Re}\,\Gamma(kd) \approx \cos(kd \sin \bar{\theta}_V) \times (1-(\cos \bar{\theta}_V kd/kd^*)^2/2)$, again for normalized separations less than kd^*, where $\bar{\theta}_V$ is the mean vertical arrival angle. (Further discussion of $\bar{\theta}_V$ and the approximation for the real part is postponed to Sec. V.) For the SW06 case $\bar{\theta}_V$ is 20.9° and upon applying this angle to the approximation together

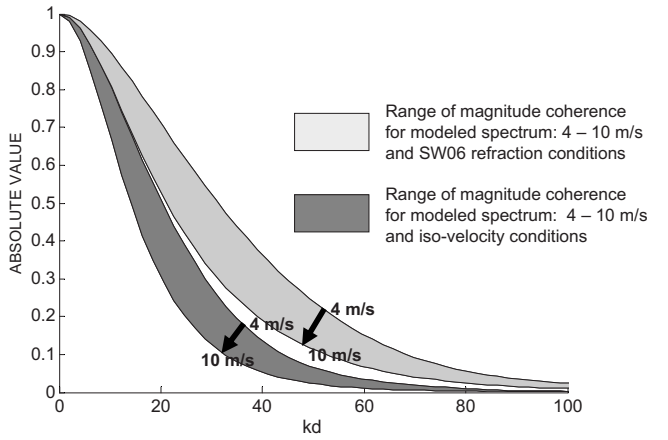


FIG. 9. Estimates of the absolute value of vertical spatial coherence $|\Gamma|$ plotted as function of normalized receiver separation kd for center frequency of 20 kHz. The lighter gray area represents a family of model coherence curves based on the SW06 sound speed profile. The upper and lower boundaries of the filled area represent specific model coherence curves based on model-generated sea surface wavenumber spectra that correspond to wind speeds of 4 and 10 m/s, respectively. The darker gray area has boundaries defined in the same manner, but curves are based on an iso-velocity profile.

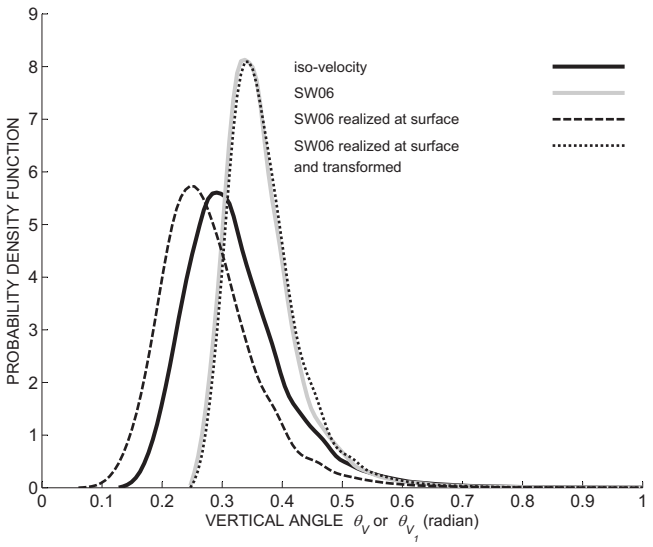


FIG. 10. PDF of received intensity as a function of particular vertical angle. Solid, gray curve incorporates SW06 refraction effects and solid, black curve assumes iso-velocity sound speed conditions; both are functions of vertical arrival angle θ_V. The black, dashed curve represents the PDF of received intensity as a function of vertical angle *as realized at the sea surface* θ_{V_1}; the black, dotted curve is the transformed version of this PDF based on Snell's law and is a function of vertical arrival angle θ_V.

with kd^* equal to 20.6, the initial oscillation at low kd observed in the real parts for the model and data is described well. In contrast, the model real part based on iso-velocity conditions is characterized by $\bar{\theta}_V \sim 18°$.

Significant fluctuations in vertical array tilt (e.g., of $\sim 1°$ or more) during the course of data collection would strongly degrade the magnitude estimates of vertical coherence. It can be demonstrated by simulation that the fluctuations in array tilt result in an increase in angular spread. This does not appear to be the case with the SW06 data, as the rms tilt was $<0.5°$ (this value is close to the resolution of the tilt sensor package). The array was subject to a mean tilt angle, however, which is close to $\sim 1°$ over the entire measurement period; this can affect estimates of the phase of vertical coherence but not the magnitude. Assuming the worst case, i.e., 1° tilt in a direction aligned with source-to-receiver propagation, and evaluating such tilting toward and away from the source, this would add and subtract, respectively, 1° from the observed $\bar{\theta}_V$. The effect is shown in Fig. 8(b) (two dashed lines) by plotting the aforementioned approximation for Re $\Gamma(kd)$ based on $\bar{\theta}_V \pm 1°$. Although the model based on SW06 observations establishes $\bar{\theta}_V$ as precisely 20.9°, small tilts in the array limit the ability to confirm this value to within about 1°. That said, the model real part based on the SW06 profile clearly provides a better fit to the observations.

V. DISCUSSION

An important effect is the increase in coherence magnitude over that expected for iso-velocity conditions, e.g., kd^* being 13.8 and 20.6 for the iso-velocity and SW06 (refraction) models, respectively [Fig. 8(a)]. For the SW06 conditions, this compression can produce a change in coherence magnitude that either is on the level of or exceeds that associated with a large change in the sea surface conditions (with mean square slope being the most relevant statistical moment of the waveheight spectrum[24]). This is illustrated in Fig. 9, where the coherence magnitude curve based on a model-generated[23] sea surface wavenumber spectra based on a wind speed of 4 m/s and using an iso-velocity profile nearly coincides with the magnitude curve based on a wind speed of 10 m/s but using the SW06 profile.

The role of refraction motivates an alternative modeling approach, and the model results (Figs. 8 and 9) can be similarly determined by constructing a PDF for received intensity vs vertical arrival angle $P_V(\theta_V)$ and taking the expected value of $e^{ikd \sin \theta_V}$ as follows:[24]

$$\Gamma(kd) = \int_{-\pi/2}^{\pi/2} e^{ikd \sin \theta_V} P_V(\theta_V) d\theta_V. \quad (1)$$

This is effectively a Fourier transform upon setting $\sin \theta_V \approx \theta_V$ and taking integration limits to infinity [viewed this way $\Gamma(kd)$ is the characteristic function for the random variable θ_V]. The received intensity from any given patch on the sea surface is proportional to $\sigma(\xi, \eta)/(R_1(\xi, \eta)R_2(\xi, \eta))^2$, and this quantity is readily sorted into discrete bins of θ_V to construct a numerical representation of the PDF $P_V(\theta_V)$ (see Fig. 10, solid, gray line), with expected value $\bar{\theta}_V$. In Fig. 8(a) Gaussian approximations for the two model curves for coherence magnitude are plotted corresponding to a Gaussian PDF for θ_V with mean $\bar{\theta}_V$ and standard deviation σ_{θ_V} equal to $1/kd^*$. Inserting a Gaussian PDF into Eq. (1) and expanding the exponential about $\bar{\theta}_V$ gives the aforementioned approximation for Re $\Gamma(kd)$ that highlights the behavior of the real parts [Fig. 8(b)]. Note, however, that an evaluation of $P_V(\theta_V)$ puts σ_{θ_V} equal to 0.065 rad, a value greater than $1/kd^*$ for the SW06 model curve (0.049 rad). The kind of skewed angular PDF, such as $P_V(\theta_V)$ (Fig. 10), has rms angular spread σ_{θ_V} such that $1/\sigma_{\theta_V}$ corresponds to $|\Gamma(kd)| \approx 0.75$ (Ref. 24) rather than $e^{-1/2}$ as would be the case where the PDF is truly Gauss-

ian. The definition of kd^* is somewhat arbitrary and here $|\Gamma(kd)|=e^{-1/2}$ is used with $1/kd^*$ interpreted as a measure proportional to corresponding angular spread but not necessarily equal to σ_{θ_V}.

Figure 10 displays four PDFs; the $P_V(\theta_V)$ corresponding to the SW06 profile (gray line) and the PDF corresponding to the iso-velocity profile (black line), which reproduce via Eq. (1) model curves for coherence identical to those in Fig. 8. The other two PDFs (dotted and dashed lines) are explained as follows: instead of vertical arrival angle, consider the distribution of received intensity as a function of vertical angle *as realized at the sea surface* or θ_{V_1}. For the iso-velocity case this gives the same PDF as the one associated with vertical arrival angle because there is no refraction; for the SW06 conditions, however, this gives the PDF shown by the dashed line in Fig. 10. Call this hypothetical PDF $P_{V_1}(\theta_{V_1})$ with standard deviation=0.083 rad, which is close to that of the iso-velocity case (standard deviation=0.088 rad). These cannot match exactly because the bistatic scattering differs slightly (Fig. 7). Snell's law relates θ_V (at receiver) to θ_{V_1} (at surface) via

$$\theta_V = \arccos(n \cos \theta_{V_1}), \tag{2}$$

where $n=1485/1530$ is the receiver depth to sea surface sound speed ratio. Using this, a transformed PDF, $P'_V(\theta_V)$, is generated from $P_{V_1}(\theta_{V_1})$ following the rule for determining the PDF for a function of a random variable.[26]

$$P'_V(\theta_V) = \frac{1}{|g|} P_{V_1}(\theta_{V_1}), \tag{3}$$

where

$$g = \frac{-n \sin \theta_{V_1}}{\sqrt{1-n^2 \cos^2 \theta_{V_1}}}, \tag{4}$$

and the inverse of Eq. (2) is used for θ_{V_1} in Eqs. (3) and (4). As anticipated, the transformed PDF $P'_V(\theta_V)$ nearly exactly recovers $P_V(\theta_V)$ (Fig. 10, dotted line). This illustrates how refraction in paths from the sea surface to receiver is responsible for compressing the angular spread that was initially approximately equal to that associated with iso-velocity conditions. Interestingly, spreading losses in such paths differ little (~0.1 dB) from their iso-velocity counterparts, and this difference would not be observable in field data whereas for coherence it is.

Snell's law relating sea surface and receiving vertical angles is smooth, so an expansion of Eq. (2) around the mean grazing angle also provides a useful functional relation between the angular variance near the sea surface var θ_{V_1} and that at the receiver var θ_V. This relation is var $\theta_V \approx \bar{g}^2$ var θ_{V_1}, where the constant $\bar{g}$, the compression factor, is given by the absolute value of Eq. (4) with θ_{V_1} replaced by the mean sea surface grazing angle for which θ_g serves as a good approximation. Given that $P_{V_1}(\theta_{V_1})$ is similar in its variance to the iso-velocity PDF, then an approximate relation between the vertical arrival angle standard deviation σ_{θ_V} and its iso-velocity counterpart $\sigma^{\text{iso}}_{\theta_V}$ is given by $\sigma_{\theta_V}/\sigma^{\text{iso}}_{\theta_V} \approx \bar{g}$. For example, recall that kd^* equals 13.8 and 20.6 for the iso-velocity and SW06 (refraction) models, respectively (Fig. 8), with the inverse of these values being proportional to angular spread. Using $\theta_g=15.3°$ gives $\bar{g}$ as ~0.7 and kd^* increases approximately by the factor of $1/\bar{g}$ owing to refraction, which is a significant effect.

VI. CONCLUSIONS

The primary conclusions that emerge from this study are as follows.

(1) The directional wave effects associated with the steady wind-wave conditions (rms waveheight of 0.16 m and wind speed of 6 m/s) did not have an observable influence on narrow band vertical spatial coherence (frequency range of 14–20 kHz). It is expected that this conclusion applies to the frequency range of O(10) kHz and above and for other wind-wave conditions.
(2) The thermocline conditions of SW06 produced strong downward refraction. In terms of sea surface forward scattering, this refraction compresses the distribution of received vertical arrival angles associated with rough surface scattering, and thereby increases vertical coherence.
(3) Models for the vertical spatial coherence that incorporates both sea surface roughness and the downward refraction conditions compare well with the observations in both magnitude and phase. The models show how refraction produces an increase in vertical coherence magnitude that is on the level of, or exceeds, that associated with a large change in sea surface wave conditions. For a given sea surface roughness, the opposite (decreased vertical coherence) is expected to occur for upward refracting conditions such as in arctic waters.
(4) A simplified model predicts the compression effect based on Snell's law, with the sound speed difference between sea surface and receiver, and surface grazing angle combined into a parameter referred to as the compression factor that predicts the change in coherence length, kd^*, due to refraction.

ACKNOWLEDGMENTS

This research was supported by the Office of Naval Research. The author thanks Neil Williams and Hans Graber from the University of Miami for making available the ASIS wind and wave data. Additional sound speed data were made available courtesy of the Woods Hole Oceanographic Institution, Ocean Acoustics Laboratory. The author wishes to thank the anonymous reviewers and Associate Editor for their helpful comments.

[1]J. Preisig, "Acoustic propagation considerations for underwater acoustic communications network development," Mob. Comput. Commun. Rev. **11**, 2–10 (2007).
[2]D. R. Dowling and D. R. Jackson, "Coherence of acoustic scattering from a dynamic rough surface," J. Acoust. Soc. Am. **93**, 3149–3157 (1993).
[3]R. L. Culver and D. L. Bradley, "On the relationship between signal bandwidth and frequency correlation for ocean surface forward scattered signals," J. Acoust. Soc. Am. **118**, 129–137 (2005).
[4]J. K. Lewis, "Determining tomographic arrival times based on matched filter processing: Considering the impact of ocean waves," J. Acoust. Soc. Am. **123**, 878–886 (2008).

[5]S. T. McDaniel, "Spatial covariance and adaptive beam forming of high-frequency acoustic signals forward scattered from the sea surface," IEEE J. Ocean. Eng. **16**, 415–419 (1991).
[6]D. Tang, J. F. Moum, J. F. Lynch, P. Abbot, R. Chapman, P. H. Dahl, T. F. Duda, G. Gawarkiewicz, S. Glenn, J. A. Goff, H. Graber, J. Kemp, A. Maffei, J. D. Nash, and A. Newhall, "Shallow Water '06: A joint acoustic propagation/nonlinear internal wave physics experiment," Oceanogr. **20**, 156–167 (2007).
[7]D. E. Weston, "Acoustic coherence loss due to ocean boundary interactions," in *Underwater Acoustics and Data Processing*, edited by Y. T. Chan (Kluwer Academic, Norwell, MA, 1989), pp. 55–67.
[8]D. E. Funk and K. L. Williams, "A physically motivated simulation technique for high-frequency forward scattering derived from specular point theory," J. Acoust. Soc. Am. **91**, 2606–2614 (1992).
[9]P. H. Dahl, "Forward scattering from the sea surface and the van Cittert–Zernike theorem," J. Acoust. Soc. Am. **115**, 589–599 (2004).
[10]P. Wille and R. Theile, "Transverse horizontal coherence of explosive signals in shallow water," J. Acoust. Soc. Am. **50**, 348–353 (1971).
[11]P. W. Smith, Jr., "Spatial coherence in multipath or multimodal channels," J. Acoust. Soc. Am. **60**, 305–310 (1976).
[12]W. M. Carey and W. B. Moseley, "Space-time processing, environmental-acoustic effects," IEEE J. Ocean. Eng. **16**, 285–301 (1991).
[13]D. Rouseff, "Intersymbol interference in underwater acoustic communications using time-reversal signal processing," J. Acoust. Soc. Am. **117**, 780–788 (2005).
[14]A. G. Voronovich, *Wave Scattering From Rough Surfaces* (Springer-Verlag, New York, 1993).
[15]S. L. Broschat and E. I. Thorsos, "An investigation of the small slope approximation for scattering from rough surface. Part II. Numerical studies," J. Acoust. Soc. Am. **101**, 2615–2625 (1997).
[16]S. L. Broschat and Y. Wang, "A practical cross-section for scattering from rough surfaces at very low grazing angles in both the forward and backward directions," Waves Random Complex Media **19**, 430–454 (2009).
[17]P. H. Dahl, "On bistatic sea surface scattering: Field measurements and modeling," J. Acoust. Soc. Am. **105**, 2155–2169 (1999).
[18]J. Pernia, *Coherence of Light*, 2nd ed. (Reidel, Boston, MA, 1985), pp. 33–36.
[19]H. C. Graber, E. A. Terray, M. A. Donelan, W. M. Drennan, J. C. Vanleer, and D. B. Peters, "ASIS—A new air-sea interaction spar buoy: Design and performance at sea," J. Atmos. Ocean. Technol. **17**, 708–720 (2000).
[20]P. H. Dahl, J. W. Choi, N. J. Williams, and H. C. Graber, "Field measurement and modeling of attenuation from near-surface bubbles for frequencies 1–20 kHz," J. Acoust. Soc. Am. **124**, EL163–EL169 (2008).
[21]G. M. Jenkins and D. G. Watts, *Spectral Analysis and Its Applications* (Holden-Day, San Francisco, CA, 1968).
[22]G. C. Carter, C. H. Knapp, and A. H. Nuttal, "Statistics of the estimate of the magnitude-coherence function," IEEE Trans. Audio Electroacoust. **21**, 388-389 (1973).
[23]W. J. Plant, "A stochastic, multiscale model of microwave backscatter from the ocean," J. Geophys. Res. **107**, 3120 (2002).
[24]P. H. Dahl, "Spatial coherence of signals forward scattered from the sea surface in the East China Sea," in *Impact of Littoral Environmental Variability on Acoustic Predictions and Sonar Performance*, edited by N. G. Pace and F. B. Jensen (Kluwer Academic, Dordrecht, The Netherlands, 2002), pp. 55–62.
[25]C. W. McCutchen, "Generalized source and the van Cittert–Zernike theorem: A study of spatial coherence required for interferometry," J. Opt. Soc. Am. **56**, 727–733 (1966).
[26]A. Papoulis, *Probability, Random Variables, and Stochastic Processes* (McGraw-Hill, New York, 1965).

Low-frequency target strength and abundance of shoaling Atlantic herring (*Clupea harengus*) in the Gulf of Maine during the Ocean Acoustic Waveguide Remote Sensing 2006 Experiment

Zheng Gong and Mark Andrews
Department of Electrical and Computer Engineering, Northeastern University, Boston, Massachusetts 02115

Srinivasan Jagannathan
Department of Mechanical Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

Ruben Patel
Institute of Marine Research, P.O. Box 1870, Nordnes, N-5817 Bergen, Norway

J. Michael Jech
Northeast Fisheries Science Center, 166 Water Street, Woods Hole, Massachusetts 02543

Nicholas C. Makris
Department of Mechanical Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

Purnima Ratilal
Department of Electrical and Computer Engineering, Northeastern University, Boston, Massachusetts 02115

(Received 30 January 2009; revised 21 October 2009; accepted 26 October 2009)

The low-frequency target strength of shoaling Atlantic herring (*Clupea harengus*) in the Gulf of Maine during Autumn 2006 spawning season is estimated from experimental data acquired simultaneously at multiple frequencies in the 300–1200 Hz range using (1) a low-frequency ocean acoustic waveguide remote sensing (OAWRS) system, (2) areal population density calibration with several conventional fish finding sonar (CFFS) systems, and (3) low-frequency transmission loss measurements. The OAWRS system's instantaneous imaging diameter of 100 km and regular updating enabled unaliased monitoring of fish populations over ecosystem scales including shoals of Atlantic herring containing hundreds of millions of individuals, as confirmed by concurrent trawl and CFFS sampling. High spatial-temporal coregistration was found between herring shoals imaged by OAWRS and concurrent CFFS line-transects, which also provided fish depth distributions. The mean scattering cross-section of an individual shoaling herring is found to consistently exhibit a strong, roughly 20 dB/octave roll-off with decreasing frequency in the range of the OAWRS survey over all days of the roughly 2-week experiment, consistent with the steep roll-offs expected for sub-resonance scattering from fish with air-filled swimbladders.
 [DOI: 10.1121/1.3268595]

PACS number(s): 43.30.Sf, 43.30.Pc, 43.30.Vh [RCG] Pages: 104–123

I. INTRODUCTION

To study the scattering characteristics, abundance and diurnal behavior of Atlantic herring (*Clupea harengus*), the most abundant fish species in and around Georges Bank during their Autumn spawning season,[1] an experiment using ocean acoustic waveguide remote sensing (OAWRS) was conducted in the Gulf of Maine from September 19 to October 6, 2006, concentrating on areas where herring shoals were most likely to form (Fig. 1). The experiment, known as OAWRS 2006, was conducted in conjunction with the U.S. National Marine Fisheries Service Annual Atlantic Herring Acoustic Survey of the Gulf of Maine and Georges Bank. Fish populations were instantaneously imaged over a 100 km diameter area by a mobile OAWRS system[2] with minute-to-minute updates to form wide-area movies of fish activity over many diurnal cycles, demonstrating the capacity of OAWRS to instantaneously image fish populations over wide areas in complex continental-shelf environments with highly variable bathymetry and oceanography. Shoals imaged by OAWRS typically comprised tens to hundreds of millions of individuals and stretched for many kilometers along the northern flank of Georges Bank. Concurrent conventional fish finding sonar (CFFS) surveys showed high spatio-temporal coregistration with fish shoals imaged by OAWRS and provided local areal population densities, as well as depth and length distributions of the fish populations. Concurrent trawl sampling, which showed Atlantic herring to be the overwhelmingly predominant species comprising the large shoals,[2,3] enabled onsite species identification and direct biological measurements of parameters such as fish length, swimbladder geometry, stomach content, and sexual development.

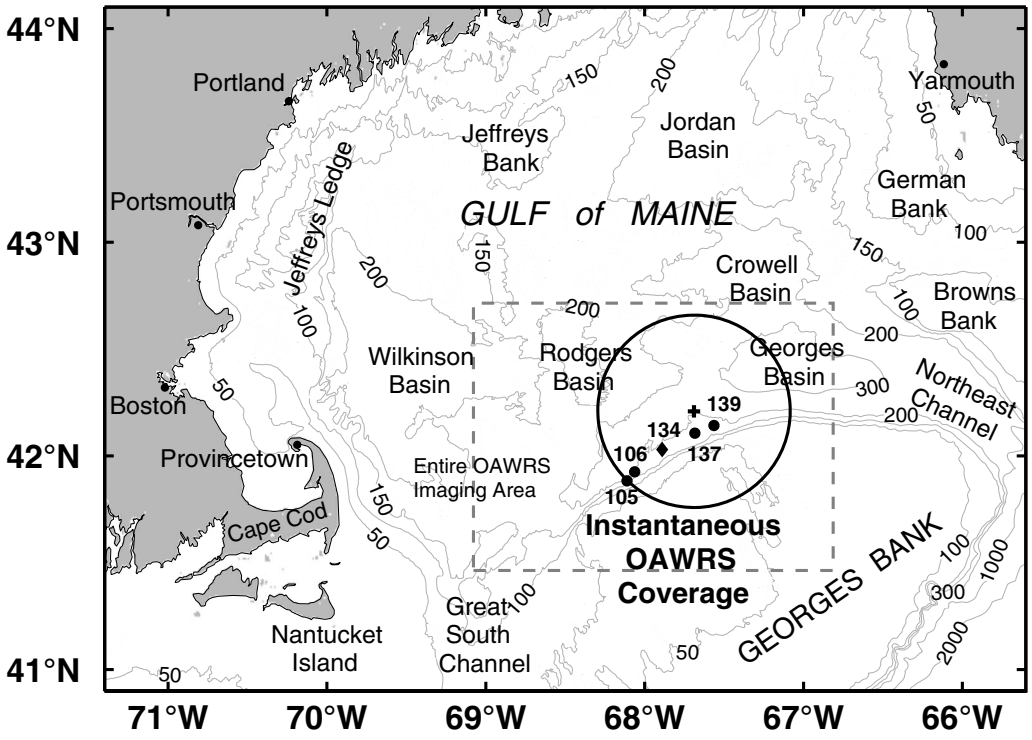


FIG. 1. Location of OAWRS 2006 experiment on the northern flank of Georges Bank in the Gulf of Maine. Plus indicates location of moored OAWRS source array deployed on Oct 1–3 at 42.2089N, 67.6892W, the coordinate origin for all OAWRS images in this paper. Circle shows typical area imaged by OAWRS, 100 km diameter and wider than Cape Cod, in 70 s. Geographic locations of trawls deployed by NOAA FRV *Delaware II* are overlain. Dots indicate trawls where herring were predominant species. In contrast, diamond indicates a trawl where silver hake and squids dominated. The gray dashed box bounds the area of OAWRS imaging during the OAWRS 2006 experiment.

Together with long-range transmission loss measurements, concurrent CFFS, and trawl data, the OAWRS imagery enabled (1) estimates of herring target strength to be made at low frequencies (300–1200 Hz) from which physical scattering mechanisms may be inferred, (2) herring spatial distributions and abundance to be estimated over ecosystem scales, and (3) regular diurnal patterns in herring behavior to be determined during the Autumn spawning season on Georges Bank.[3,4]

The mean scattering cross-section of an individual shoaling herring is found to consistently exhibit a strong, roughly 20 dB/octave roll-off with decreasing frequency in the range of the OAWRS survey over many measurement days, consistent with the steep roll-off expected for subresonance scattering from fish with air-filled swimbladders. These findings suggest that OAWRS can provide valuable evidence for remote species classification over wide areas since significant variations in the frequency dependence of target strength are expected across species due to differences in resonance. This is because the dominant source of acoustic scattering at low and mid-frequencies (less than 10 kHz) is the air-filled swimbladder for fish that have swimbladders.[5] Resonance frequencies depend on swimbladder volume, shape, ambient pressure, and the effect of surrounding tissues.[6,7] For many fish species of economic importance in the size ranging from 10 to 50 cm, resonances are expected to range from several hundreds of hertz to a few kilohertz.[6–9] Previous experimental investigation of resonance have been limited to small scale tank measurements with individual fish out of their natural environment or highly localized measurements made *in situ* with sources of rapid power roll-off below 1.7 kHz.[7,10,11]

The population that spawns on the northern flank of Georges Bank (Fig. 1) is the largest herring stock in the Gulf of Maine, and has both ecological and economic importance.[1] It has been surveyed annually by the U.S. National Marine Fisheries Service for roughly 1 decade during the Autumn spawning season.[12–14] Current estimates of the Georges Bank herring stock varies from 500 000 to 1×10^6 tons based on acoustic surveys and other assessment methods,[14] respectively. The National Marine Fisheries Service acoustic survey employs highly localized CFFS measurements along widely spaced line-transects, roughly 20–30 km apart,[14] trawl sampling at selected locations, and takes roughly 1 week to cover the northern flank of Georges Bank from east to west. As a result, annual stock estimates may be highly aliased in both time and space. One of the primary goals for OAWRS 2006 is to provide images of fish populations over the vast areas they inhabit that are unaliased in both space and time[2] so that more reliable abundance estimates may be obtained.

II. MULTI-SENSOR EXPERIMENT DESIGN AND RESOURCES

The OAWRS 2006 experiment was designed to coincide with the National Marine Fisheries Service annual herring survey of Georges Bank. It was conducted with four research vessels (RVs) that employed a suite of acoustic imaging sensors, several oceanographic monitoring systems, and trawls

TABLE I. OAWRS receiving array 3-dB angular resolution $\beta(\phi)$ at broadside $\phi=0$ and endfire $\phi=\pi/2$, and aperture length L as a function of imaging frequency f_c. A Hanning spatial window is applied in the beamforming.

f_c (Hz)	L (m)	$\beta(\phi=0)/\beta(\phi=\pi/2)$ (deg)
415	94.5	3.15/31.4
735	47.25	3.56/33.3
950	47.25	2.75/29.3
1125	23.625	4.65/38.1

for species identification. The OAWRS vertical source array and towed horizontal receiving array were separately deployed from two medium sized UNOL vessels, RV *Endeavor* and RV *Oceanus* respectively, for bistatic measurement of echo returns. The instantaneous areal coverage of the OAWRS system in a single transmission is shown in Fig. 1. The vertical source array transmitted a suite of individual Tukey-windowed linear frequency modulated (LFM) pulses of 1 s duration and 50 Hz bandwidth centered at a suite of frequencies with a repetition interval of 150 s for each center frequency.[15] Broadband LFM pulses centered at 415 and 735 Hz were transmitted seconds apart, then after 75 s those centered at 950 and 1125 Hz were transmitted seconds apart, and the process was repeated. Transmissions radiated with azimuthal symmetry about the OAWRS source array, for which more information is available in Ref. 16, with source level continuously monitored with two desensitized hydrophones deployed from RV *Endeavor*. Scattered returns were acquired with a horizontal receiving array, the ONR five-octave research array, towed by RV *Oceanus* along designated tracks. The multiple nested sub-apertures of the array span 50–3750 Hz frequency range. Returns measured within each linear section of the array are processed by beamforming and matched filtering with angular resolution shown in Table I. The receiving array also contained one desensitized hydrophone which was used to measure transmitted signals from the source array for transmission loss and source level calibrations. Two calibrated acoustic targets made of air-filled rubber hose,[17] approximately 30 m long and 7 cm in diameter with known scattering properties,[18] were vertically deployed at selected locations to enable accurate charting of scattered returns in both range and azimuth. One of the targets was moored with lower end 5 m off the seafloor and the other was centered at 140 m in waters 200 m deep.

Over the course of the OAWRS 2006 experiment, more than 3000 wide-area images of the ocean environment were acquired by the OAWRS system for each of the four LFM center frequencies leading to more than 12 000 images in total. Similarly, more than 12 000 transmission loss measurements were made over the survey area to calibrate our transmission loss model. The length of each RV *Oceanus* towed array track was typically 15 km. With a nominal tow speed of 2 m s^{-1} for the receiver ship, a total of roughly 75 images of the ocean environment were generated per frequency along each track. Minute-to-minute updates of the OAWRS imagery made it possible to closely monitor herring activity over wide areas and observe patterns of spatial distribution evolve over the course of each day. The inherent left-right

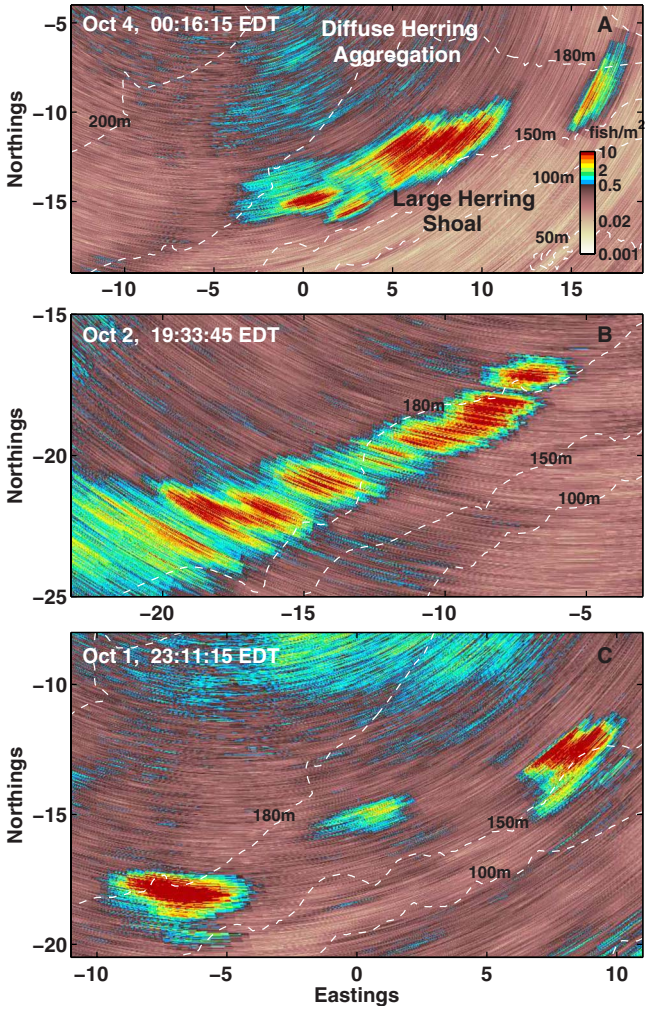


FIG. 2. (Color online) [(A)–(C)] OAWRS images of areal fish density zoomed-in around massive herring shoals, with densities exceeding 10 fish m^{-2} in population centers. Measured during evening to midnight hours of October 4, 2, and 1, respectively. (A) The total population of herring in the large dense shoal is roughly 170×10^6, and that in the diffuse cloud outside the large shoal is roughly 70×10^6. Imaged shoal populations of herring are approximately 86×10^6 and 70×10^6 respectively for (B) and (C). Uncertainty in the abundance estimate is 17–20%. Note that the figures are plotted on different scales, and the coordinate origin is the source location shown in Fig. 1.

ambiguity about the horizontal line-array's axis in the OAWRS images were resolved mainly by varying receiver ship heading, sometimes only slightly for several transmissions by what we call a "Crazy Ivan" for immediate results, as well as ship position. These approaches for ambiguity resolution are described in Refs. 19–21.

Examples of the massive fish shoals instantaneously imaged by OAWRS near Georges Bank are shown in Fig. 2. The massive shoal imaged during midnight hours of October 4 within the 150–180 m bathymetric contour in Fig. 2(A), for example, extends 15×5 km^2 and comprises roughly 170×10^6 fish distributed about several population centers. The area occupied by this shoal is approximately equal to that of Manhattan Island in New York. The fish population in the diffuse cloud region to the north is comprised of over 70×10^6 individuals.

Concurrent localized imaging of fish aggregations at OAWRS-directed locations was conducted by two other re-

TABLE II. Conventional fish finding sonars, SIMRAD EK60 and EK500 specifications. The angular 3-dB beamwidth is denoted by β, the pulse duration by PD, and repetition rate by RR. The resolution diameter, Res, is calculated for 200 m water depth.

Sensor	f (kHz)	β (deg)	Res (m)	PD (ms)	RR (s^{-1})
EK60	38				
	120	7	24	1	1
	200				
EK500	18	11	39	2	
	38	12	42	1	0.5
	120	7	24	1	

search vessels, the RV *Hugh Sharp* and the NOAA FRV *Delaware II*, using two downward-directed CFFS systems, the SIMRAD EK60 and EK500 echosounders, respectively. Both the EK60 and EK500 echosounders insonify the water column directly beneath the survey vessel simultaneously at three frequencies to provide the local depth dependence of dominant fish layers within their instantaneous resolution footprints, of between 24–50 m diameter, and estimates of volumetric and areal fish population densities. Specifications of these two echosounders appear in Table II. A Reson 7125 Seabat multi-beam sonar (400 kHz) system was also deployed from RV *Hugh Sharp* with an angular swath of 128°. It was useful in providing detailed three-dimensional morphology of smaller fish groups located in the mid-water column.[22] A high-speed rope trawl[23] deployed by NOAA FRV *Delaware II* enabled species identification[14] at OAWRS-directed locations.

Physical oceanography was monitored by sampling water-column temperature and salinity with expendable bathythermographs (XBTs) and conductivity-temperature-depth (CTD) sensors at regular hourly intervals from all four research vessels. The water-column sound speed profile was found to be relatively constant in space and time over the 2006 OAWRS survey, as shown by the compilation of over roughly 200 samples taken during the experiment in Fig. 3. No mesoscale oceanographic features such as eddies were found or expected. The small fluctuations about the mean profile are due to mild internal wave activity that causes well-understood short-term Gaussian field fluctuations in acoustic transmission that have an intensity standard deviation that can be reduced to a small fraction of the mean by stationary averaging.[24–26] An instrumented tow cable was also deployed from the RV *Hugh Sharp* to provide continuous measurement of temperature. This oceanographic information was used to carefully update horizontal locations and depths of the OAWRS source (typically centered at 60–70 m) and receiving arrays (centered at 105 m) (Ref. 15) to optimize OAWRS imaging of fish groups. Decisions were often based on the outputs of the range-dependent acoustic model (RAM), based on the parabolic equation, for multimodal waveguide transmission loss in the range-dependent Georges Bank environment with the hourly sound speed profile updates and known bathymetry.[27]

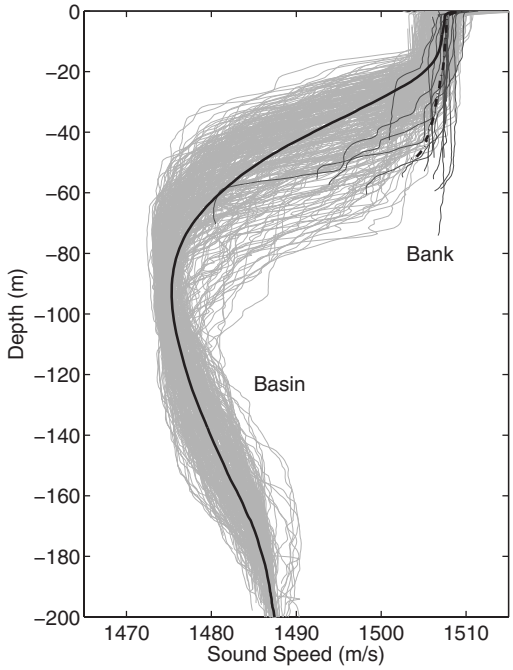


FIG. 3. Profiles of water-column sound speed from XBT and CTD measurements made from all four research vessels on the Northern Flank of Georges Bank and Georges Basin during OAWRS 2006.

III. DATA PROCESSING AND ANALYSIS

A. Generating instantaneous wide-area OAWRS images of the ocean environment

Wide-area images of instantaneous scattered intensity spanning 100 km in diameter were generated in near real-time for every broadband transmission centered at each of the four frequencies, f_c=415, 735, 950, and 1125 Hz. For each transmission, the pressure data on the receiving array were first beamformed to determine the azimuth of the arrivals, then matched filtered with the source signal, and charted in range using two-way travel time.[19–21] Each image was then mapped onto geographic space using the GPS latitude and longitude information of the source and receiving array. A nominal sound speed of 1475 m/s that minimizes charting errors was used to convert the travel-time of the signal to range.[21,28] The range resolution $\Delta\rho$ of the OAWRS system is approximately 15 m after matched-filtering, and the azimuthal resolution $\beta(\phi, f_c)$ associated with each frequency band both at broadside and endfire is tabulated in Table I. A Hanning spatial window was applied in the beamforming to reduce sidelobe levels by more than 30 dB from the main lobe. A detailed explanation of the image formation process is provided in Refs. 2, 20, and 21.

A standard procedure of averaging three consecutive instantaneous OAWRS images and two adjacent range cells is used for all OAWRS images presented here. This leads to an experimentally determined standard deviation in log-intensity of roughly 1.5 dB,[3] consistent both with theory and previous experiments.[2,3,16,20,21] This standard deviation is negligible compared to the dynamic ranges of features in the OAWRS images and the variations in herring target strength measured across frequency in the OAWRS range.

B. Estimating areal fish population density from instantaneous OAWRS imagery

Here we describe how areal fish population density over wide areas may be estimated from OAWRS intensity images of the ocean environment. At typical OAWRS operating frequencies from hundreds of hertz to a few kilohertz, most fish are acoustically compact scatterers, with swimbladder sizes that are much smaller than the wavelength. The sonar equation approach is then valid for analyzing scattering from fish since the scattered field from each individual is omnidirectional, making propagation and scattering factorable even in a waveguide.[29] The expected scattered intensity from fish aggregations after matched-filtering is dominated by the incoherent intensity or variance of the scattered field and multiple scattering effects are negligible for the densities found here, as shown in Refs. 30 and 31. As a result, given a source at $\boldsymbol{r}_0$ transmitting a broadband signal with bandwidth centered at f_c and a receiver at $\boldsymbol{r}$, the expected scattered intensity, $\langle I_s(\boldsymbol{\rho}_m,f_c)\rangle$, within the OAWRS resolution footprint of area $A(\boldsymbol{\rho}_m|\Delta\rho,f_c)$ centered at horizontal location $\boldsymbol{\rho}_m$ can be expressed as

$$\langle I_s(\boldsymbol{\rho}_m,f_c)\rangle = \sum_{i=1}^{M(\boldsymbol{\rho}_m)} \langle|Q(f_c)|^2\rangle \times(4\pi)^4\langle|G(\boldsymbol{r}_i|\boldsymbol{r}_0,f_c)G(\boldsymbol{r}|\boldsymbol{r}_i,f_c)|^2\rangle \times\left\langle\frac{|S(\boldsymbol{r}_i,f_c)|^2}{k^2}\right\rangle, \quad (1)$$

where $M(\boldsymbol{\rho}_m)$ is the number of fish within the resolution cell, $\boldsymbol{r}_i$ is the location of the ith fish, $|Q(f_c)|^2$ is the source intensity, $G(\boldsymbol{r}_i|\boldsymbol{r}_0,f_c)$ and $G(\boldsymbol{r}|\boldsymbol{r}_i,f_c)$ are the waveguide Green's functions from the source to each scatterer and from each scatterer to the receiver, respectively, $S(\boldsymbol{r}_i,f_c)$ is the fish scatter function, and k is the wavenumber.

The expected intensity in a fluctuating waveguide from uniformly distributed targets within the resolution footprint can be approximated as

$$\langle I_s(\boldsymbol{\rho}_m,f_c)\rangle \approx \langle|Q(f_c)|^2\rangle\gamma(\boldsymbol{\rho}_m,f_c)\sum_{i=1}^{M(\boldsymbol{\rho}_m)}\left\langle\frac{|S(\boldsymbol{r}_i,f_c)|^2}{k^2}\right\rangle, \quad (2)$$

where $\gamma(\boldsymbol{\rho}_m,f_c)=\langle|(4\pi)^2G(\boldsymbol{r}_m|\boldsymbol{r}_0,f_c)G(\boldsymbol{r}|\boldsymbol{r}_i,f_c)|^2\rangle$ for sufficiently narrow depth layers H and areal footprints over which $\gamma(\boldsymbol{\rho}_m,f_c)$ becomes effectively constant, as shown for the OAWRS 2006 fish shoal imaging in Ref. 4. The last factor of Eq. (2) can be written as

$$\sum_{i=1}^{M(\boldsymbol{\rho}_m)}\left\langle\frac{|S(\boldsymbol{r}_i,f_c)|^2}{k^2}\right\rangle = \sum_{i=1}^{M(\boldsymbol{\rho}_m)}\int\int\int\frac{|S(\boldsymbol{r}_i,f_c)|^2}{k^2}P(\boldsymbol{r}_i)d\boldsymbol{r}_i^3 = M(\boldsymbol{\rho}_m)\bar{\sigma}(\boldsymbol{\rho}_m,f_c), \quad (3)$$

where $P(\boldsymbol{r}_i)$ is the probability density of finding the ith fish at location $\boldsymbol{r}_i$, and $P(\boldsymbol{r}_i)=1/A(\boldsymbol{\rho}_m|\Delta\rho,f_c)H$ for uniformly distributed fish shoals, $\bar{\sigma}(\boldsymbol{\rho}_m,f_c)$ is the average scattering cross-section of an individual fish over the OAWRS resolution footprint and the depth layer, $n_{A,\mathrm{OAWRS}}(\boldsymbol{\rho}_m)=M(\boldsymbol{\rho}_m)/A(\boldsymbol{\rho}_m|\Delta\rho,f_c)$ is the mean areal fish population density within the resolution footprint, and $A(\boldsymbol{\rho}_m|\Delta\rho,f_c)\approx\rho_m\Delta\rho\beta(\phi,f_c)$ is the range and azimuth-dependent spatial resolution of the OAWRS imaging system.[21]

Inserting Eq. (3) into Eq. (2) and taking $10\log_{10}$ of both sides, we obtain the scattered intensity level in decibels;

$$L(\boldsymbol{\rho}_m,f_c)\approx SL(f_c)+TTL(\boldsymbol{\rho}_m,f_c)+\mathrm{SS}_{\mathrm{OAWRS}}(\boldsymbol{\rho}_m,f_c) + 10\log_{10}(A(\boldsymbol{\rho}_m|\Delta\rho,f_c)), \quad (4)$$

where $L(\boldsymbol{\rho}_m,f_c)=10\log_{10}\langle I_s(\boldsymbol{\rho}_m,f_c)\rangle$, $TTL(\boldsymbol{\rho}_m,f_c)=10\log_{10}\gamma(\boldsymbol{\rho}_m,f_c)$ describes the expected second moment of depth averaged propagation to and from the fish layer averaged over the resolution footprint of the OAWRS system, $SL(f_c)=10\log_{10}\langle|Q(f_c)|^2\rangle$ is the spectral source level, and $\mathrm{SS}_{\mathrm{OAWRS}}(\boldsymbol{\rho}_m,f_c)$ is the scattering strength.

From Eqs. (2)–(4), OAWRS scattering strength can be expressed as

$$\mathrm{SS}_{\mathrm{OAWRS}}(\boldsymbol{\rho}_m,f_c)=\mathrm{TS}_{\mathrm{OAWRS}}(f_c) + 10\log_{10}\langle n_{A,\mathrm{OAWRS}}(\boldsymbol{\rho}_m)\rangle, \quad (5)$$

where $\mathrm{TS}_{\mathrm{OAWRS}}(f_c)=10\log_{10}\bar{\sigma}(\boldsymbol{\rho}_m,f_c)$ in units of dB re 1 m^2 is the target strength corresponding to the average scattering cross-section of an individual fish over the OAWRS resolution footprint and depth layer within the bandwidth centered at f_c.

The terms in Eq. (4) are evaluated separately for each of the four OAWRS LFM waveforms with different center frequencies f_c. A calibrated stochastic transmission loss model based on the parabolic equation[27] for a range-dependent fluctuating ocean waveguide is used to estimate the random Green's functions and determine $TTL(\boldsymbol{\rho}_m,f_c)$ following the approach described in the Appendix E and Ref. 16. Expected source level is estimated from one-way propagated signals received by a desensitized hydrophone on the moving receiver array using the approach of Ref. 16. The two monitoring hydrophones on the source ship were used to verify the source level estimates. Our analysis indicates the source transmitted a stable output over the course of each day.

The application of Eq. (4) to estimate scattering strength from OAWRS imagery is illustrated in Figs. 2 and 3 of Ref. 32 and in Ref. 33. Scattering strength is a useful parameter for characterizing submerged objects, both distant and nearby, because it is independent of the spatially varying transmission loss and areal resolution footprint of the imaging system. Once the target strength expected of an individual fish is known, an areal fish population density image can be obtained from a scattering strength image[33] using Eq. (5). The target strength corresponding to the average scattering cross-section of an individual fish at OAWRS operating frequencies is estimated by matching between OAWRS and CFFS areal fish population density measurements where simultaneous sampling through stationary fish populations is available.

C. Estimates of areal fish population density from CFFS

The CFFS measurements at 38 kHz are used to provide local estimates of areal fish population density.[14,34,35] The 7° 3-dB beamwidth yields an instantaneous circular survey area

TABLE III. Physical parameters of modeled fish species and their measured target strength at 38 kHz with a CFFS.

Species	Atlantic herring	Acadian redfish	Silver hake
L_{FL} (cm)	19–30	15–39	2–35
L_{TL} (cm)	20–34	16–41	2–35
Depth (m)	120–190	120–190	10–75
$\langle TS_{CFFS} \rangle$ (dB)	−39.7[a]	−38.9[b]	N/A
σ_{CFFS} (dB)	1.3[c]	2.4	N/A
ρ_f[d] (kg m^{-3})	1071	1080	1050
ζ[e] (Pa s)	50	50	20
ϵ[f]	5–10	6	8
κ_{nb}[g]	0.05	0.05	0.03
κ_z[h]	x[i]	0.05	0.03

[a]Mean target strength of herring calculated using Eq. (7).
[b]Mean target strength of redfish calculated using equation in Ref. 71.
[c]Standard deviation of derived herring target strength at 38 kHz incorporating fish length and depth distribution from CFFS and trawl surveys.
[d]Fish flesh density.
[e]Viscosity of fish flesh.
[f]major-to-minor-axis ratio of fish swimbladder.
[g]Fish swimbladder volume to fish body volume ratio at neutral buoyancy depth.
[h]Fish swimbladder volume to fish body volume ratio at depth.
[i]x is a linear function of ambient pressure at depth given by $x = \kappa_{nb}(P_{nb}/P_x)$, where P_{nb} is the ambient pressure at neutral buoyancy depth z_{nb}, and P_z is the ambient pressure at any depth z.

of 24 m diameter directly under the survey vessel at 200 m water depth. Volumetric scattering from all targets within the conical beam were measured. The localized areal fish population density in fish m^{-2}, denoted by $n_{A,CFFS}$, can be estimated using

$$n_{A,CFFS} = \frac{4\pi}{\bar{\sigma}_{bs}} \int_{z_1}^{z_2} s_v dz, \tag{6}$$

where s_v is the volume backscattering coefficient[36] in m^{-1}, z_1 and z_2 delimit the depth bounds for fish aggregations, and $\bar{\sigma}_{bs} = 4\pi 10^{\langle TS_{CFFS} \rangle/10}$ is the mean backscattering cross-section of an individual at 38 kHz in units of m^2, where $\langle TS_{CFFS} \rangle$ is the corresponding mean target strength at ultrasonic frequency.

The expected target strength for an individual fish at 38 kHz varies with species, depth, and total fish length. Here, the expected TS_{CFFS} in dB of an individual herring of total length L_{TL} in centimeters at depth z in m is obtained from Eq. (5) of Ref. 37,

$$TS_{CFFS} = 20 \log_{10} L_{TL} - 2.3 \log(1 + z/10) - 65.4, \tag{7}$$

and then converted to σ_{bs}. The mean backscattering cross-section $\bar{\sigma}_{bs}$ is obtained as the weighted average over the total length and depth distribution of the fish aggregations. From trawl surveys of the imaged fish populations in OAWRS, herring was the overwhelmingly dominant species comprising the large shoals, which had small fractions of redfish and silver hake. Estimates of the mean TS_{CFFS} for individual herring and redfish based on our trawl measurements (Appendix A) of the length distribution are provided in Table III. The expected target strength of herring and redfish over similar depth extent at 38 kHz are close, varying at most by 1 dB, albeit their different length distributions. In contrast, their low-frequency target strength near resonance varies significantly, as discussed in Sec. IV C. These modeled TS_{CFFS} values are in good agreement with those obtained by experimentally analyzing the CFFS backscattered field from individual fish distinguishable in the periphery of various aggregations consistent with 0.1 dB mean squared errors reported in Ref. 37.

Figures 4(D) and 4(E) illustrate the application of Eq. (6) to estimate areal fish density for herring aggregations in the 120–180 m water depth range.

D. Estimating low-frequency target strength by matching OAWRS and CFFS population densities

Here we describe our procedure for estimating the low-frequency target strength corresponding to the average scattering cross-section of an individual shoaling herring over the resolution footprint of the OAWRS system by correlating OAWRS data with simultaneous measurements made along CFFS transects. The target strength of herring at 950 and 1125 Hz is found to be significantly higher than at 415 and 735 Hz, making much lower herring densities observable at these higher frequencies. At the lower frequencies of 415 and 735 Hz, the herring target strength is weaker causing the scattered returns to be background saturated at moderate fish densities. Due to the receiving array's sub-aperture design, OAWRS images at 950 Hz have the best cross-range resolution, making this an optimal frequency for wide-area sensing. An alternative approach for target strength estimation, based on differencing pairs of OAWRS wide-area scattering strength images at two distinct frequencies, is applied in Sec. III E to determine target strength at the lower frequencies. The target strength estimates are summarized in Table IV.

Calibrated acoustic targets were deployed on October 2–3 enabling independent and precise geographic charting of OAWRS images. By making small adjustments to the charting speed and array orientation, scattered returns from calibrated targets were accurately charted to the correct range-azimuth resolution cell relative to the source and receiver. This ensures that scattered returns from all other targets, including the fish aggregations, have been accurately charted as well. In this section we focus on data acquired on October 2–3 when calibrated target data were available and present target strength estimates for other days in the Appendix B.

Close to midnight on October 2, both OAWRS and CFFS systems simultaneously co-registered a massive herring shoal between the 150 and 180 m isobaths on the northern flank of Georges Bank, as shown in Figs. 4(A)–4(C). The observations were made continuously over a 90-min period between 23:30 Eastern Daylight Time (EDT) on October 2 and 01:00 EDT on October 3. Measurements from the two systems are highly correlated during the course of the observations because of the statistical stationarity of the fish populations even though their resolution footprints are significantly different. The OAWRS system monitored and sampled the temporal and spatial evolution of the shoal's horizontal morphology at intervals of 75 s without aliasing. Concurrently, the CFFS system crossed the same shoal twice along a U-shaped transect with two parallel transects 1.5 km apart.

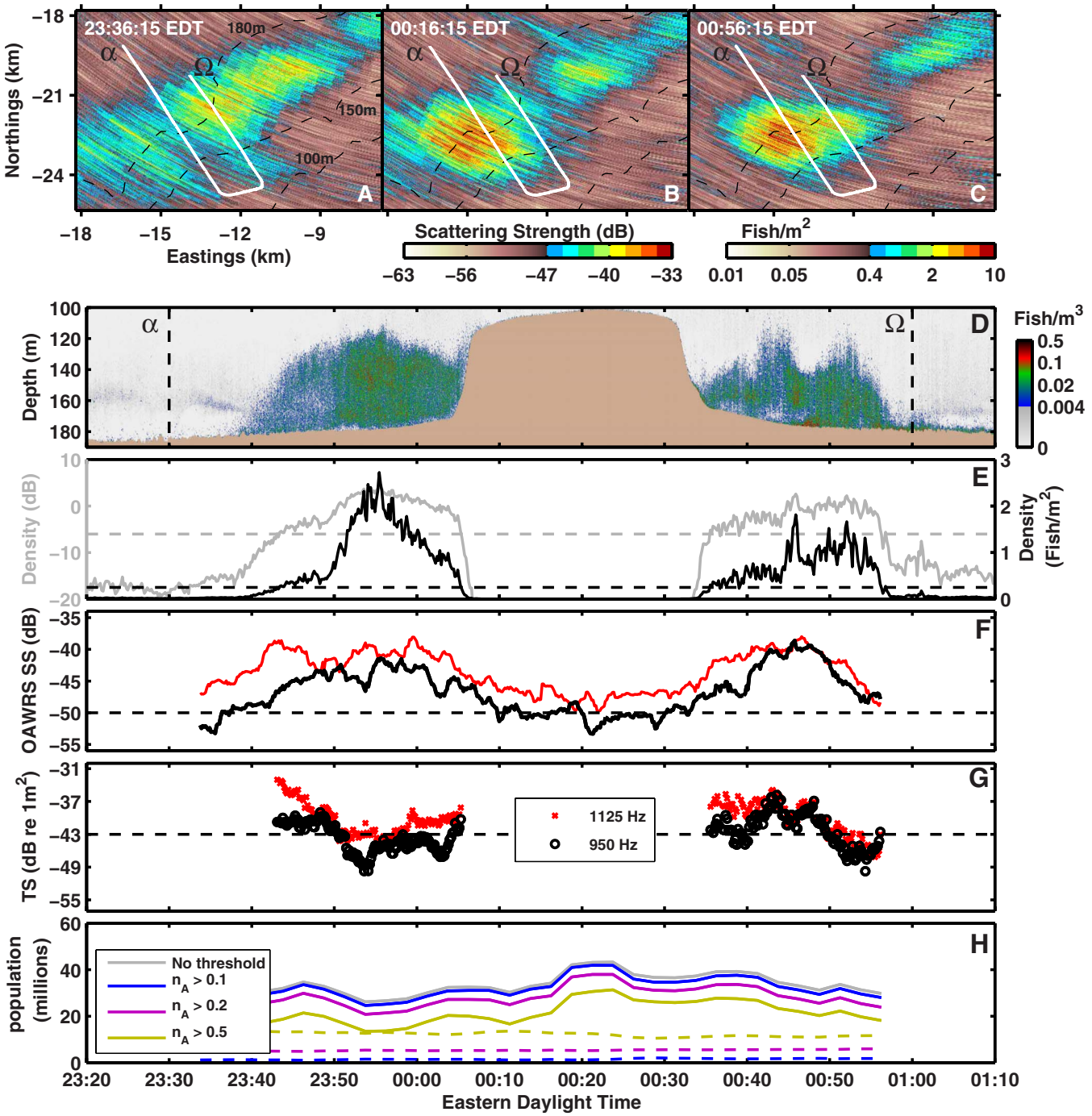


FIG. 4. (Color online) Herring target strength at 950 and 1125 Hz estimated by matching areal fish density in OAWRS and CFFS data acquired during midnight hours of October 2. [(A)–(C)] A sequence of instantaneous OAWRS scattering strength images zoomed into the region containing a massive herring shoal with overlain CFFS line-transect (solid line) made at nominal tow-speed of 2.5 m s^{-1}. (D) CFFS time-depth echogram provides local depth distributions of fish aggregations. Dashed lines at 23:30 EDT and 01:00 EDT correspond to transect start and end points α and Ω, respectively. (E) The areal fish population densities inferred from CFFS measurements following Eq. (6) are plotted as a function of time in black, and the corresponding areal fish population densities in dB, $10\log_{10}(n_{A,\mathrm{CFFS}})$, are plotted in gray. (F) The OAWRS scattering strength measurements and (G) instantaneous target strength estimates along CFFS line-transects at 950 and 1125 Hz. Target strength estimates near the edge of shoals are not accurate because of nonstationarity. (H) Population of herring within the area shown in (A)–(C) determined with various OAWRS fish density n_A thresholds. Solid line gives population above the threshold and dotted line gives population below the threshold.

The depth distribution of the fish population, within roughly 40–60 m of the seafloor, is relatively consistent across the two CFFS transects, as shown in Fig. 4(D).

To accurately estimate low-frequency target strength, we confine our present analysis to contiguous space-time segments that consistently register significant, stationary scattering from fish aggregations in both OAWRS and CFFS systems. We derive threshold values for CFFS population density and OAWRS scattering strength. The segmented data above these thresholds are used for target strength estimation. The CFFS threshold is set at 0.2 fish m^{-2}, as shown in Fig. 4(E). For OAWRS, two square areas of dimension 6.2×3 and 1.57×5.58 km^2 that continuously register significant fish scattering and diffuse background reverberation,

TABLE IV. Mean low-frequency target strength estimates. The $\hat{TS}_{cor}$ estimates are obtained by correlating OAWRS with CFFS data along CFFS transect. This approach is only applied to OAWRS data at 950 and 1125 Hz. For the other frequencies, the $\hat{TS}_{sc}$ estimates are obtained by the approach of differencing OAWRS images. The Diff is the expected target strength difference between the given frequency and 950 Hz.

Date	f_c (Hz)	$\hat{TS}_{cor}$ (dB *re* 1 m^2)	Diff (dB)	$\hat{TS}_{sc}$ (dB *re* 1 m^2)	$\hat{\sigma}_{TS}$ (dB)
Oct 2	415	N/A	≈−17.5	≈−60.5	2
	735	N/A	<−8	<−51.0	>3
	950	−43.0	N/A	−43.0[a]	0.7
	1125	−40.3	≈7	≈−36.0	1.7
Oct 3 transect1	415	N/A	≈−17	≈−57.9	2
	735	N/A	≈−8.5	≈−49.4	1.5
	950	−40.9	N/A	−40.9	0.8
	1125	−35.6	≈8	≈−32.9	1.5
Oct 3 transect2	415	N/A	<−17	<−58.7	>3
	735	N/A	≈−7.5	≈−49.2	1.5
	950	−41.7	N/A	−41.7	0.5
	1125	−37.0	≈7.5	≈−34.2	1
Sep 27	415	N/A	<−15	<−57	>3
	735	N/A	≈−12.5	≈−54.5	1.7
	950	−42	N/A	−42	0.7
	1125	N/A	≈5	≈−37	1.5
Sep 29	415	N/A	<−10	<−54.4	>3
	735	N/A	<−7	<−51.4	>3
	950	−44.4	N/A	−44.4	0.8
	1125	N/A	≈10	≈−34.4	2

[a]Base target strength at 950 Hz used in the differencing approach.

respectively, throughout the course of observation are first examined. The histogram of scattering strength values within these areas, averaged over multiple OAWRS images, are plotted in Fig. 5(D). The histograms are approximately Gaussian. The OAWRS threshold is then set at −50 dB for 950 Hz to distinguish fish scattering from the background. This threshold is roughly 2 standard deviations below the mean for the fish histogram and roughly 2 standard deviations above the background mean.

Employing Eq. (5), and assuming local stationarity of fish population, we set the areal fish density within the OAWRS resolution footprint to that simultaneously sampled by CFFS transect through the OAWRS footprint, $n_{A,\mathrm{OAWRS}} \approx n_{A,\mathrm{CFFS}}$. The resulting target strength estimates for fish in these contiguous shoaling regions at 950 Hz are shown in Fig. 4(G). The differences in target strength estimates along the transect are due to the fact that the OAWRS and CFFS systems have different resolution footprint sizes, and so the true mean fish areal densities within the OAWRS resolution cell may be overestimated or underestimated by the CFFS system given nonstationary spatial distributions, as occurred at shoal boundaries. The combination of measurements from many space-time locations from both systems should yield mean target strength estimates with small variance by virtue of the law of large numbers as discussed in Appendix C.

Similar statistical analyses have been conducted for OAWRS data at 1125 Hz, with estimated target strength appearing in Fig. 4(G). This approach is also applied to infer herring target strength at 950 and 1125 Hz using OAWRS and CFFS data on October 3, where two contiguous shoal segments are imaged. Estimated target strengths for these two segments are provided in Table IV. The target strength estimates at 950 Hz for the three data sets are consistent, with a standard deviation of roughly 1 dB.

The approach of this section is not used to estimate target strength at 415 and 735 Hz because the herring are much weaker scatterers at these frequencies, as seen in Fig. 5(A) where only the densest shoal population centers stand above background scattering levels. An alternative approach to estimate herring target strength at these lower frequencies is developed and applied in Sec. III E.

E. Frequency dependence of target strength estimated by differencing OAWRS scattering strength images over wide areas

Here, we develop an alternative approach to estimate target strength expected of an individual shoaling herring by differencing pairs of OAWRS scattering strength images acquired at two distinct frequencies over the entire area of the shoal. We apply this to data at 415 and 735 Hz. The approach is illustrated by the conceptual diagram shown in Fig. 6. From Eq. (5), we observe that scattering strength in areas containing fish increases logarithmically with areal density n_A. Here, f_1 represents a low frequency, such as 415 Hz, where the target strength for fish is lower than at another frequency f_2, such as 950 Hz. The background scattering strength from sources other than fish is expected to be statistically stationary and can be identified by its mean level which is frequency dependent. The total scattering strength measured at any given pixel is a sum of the contribution from fish and other background effects. The difference between the total scattering strength across various pixels at the two frequencies then follows the trend illustrated in Fig. 6(B), where at very low fish densities, the scattering strength is dominated by the background reverberation, and at very high fish densities by fish scattering. The difference in the scattering strength at low fish densities therefore provides a measure of the difference in background reverberation. The difference in the scattering strength at high fish densities is equal to the target strength difference for fish at these frequencies. If the target strength at one of the frequencies is known accurately, then the target strength at the other frequency can be obtained.

This approach is implemented for pairs of OAWRS images using 950 Hz as the base frequency. The difference in scattering strength is calculated and plotted for OAWRS data acquired between 22:00 and 22:45 EDT on October 3 in Figs. 7(A)–7(C) for various frequency pairs. We observe the scattering strength difference in the background is roughly 1 dB between 1125 and 950 Hz, but the fish target strength difference is larger, roughly 7.5 dB. Between 950 and 415 Hz, the background scattering strength difference is roughly 1.5 dB, but the fish target strength difference is more than 17 dB. Between 950 and 735 Hz, no conclusion can be

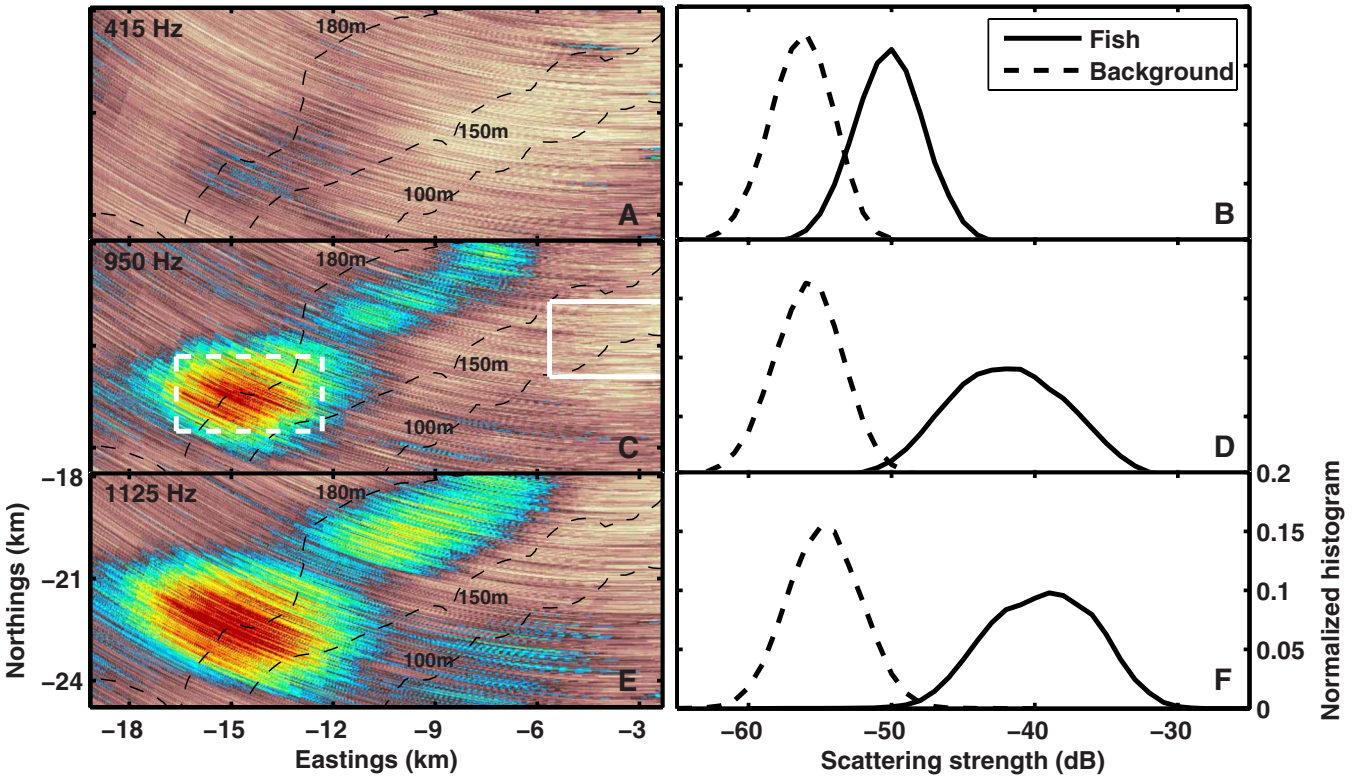


FIG. 5. (Color online) The intensity of scattered returns from shoals is highly frequency-dependent. The histograms illustrate that it is easier to detect shoals over background regions at higher frequencies. Simultaneous trawls show shoals are overwhelmingly comprised of herring while background regions yield negligible herring (Table IV, Fig. 12). [(A), (C), and (E)] OAWRS images of herring shoal acquired simultaneously at three distinct frequency bands centered at 415, 950, and 1125 Hz at 00:41:15 EDT on October 2. The colorscale used in (A), (C), and (E) is the same as in Figs. 4(A)–4(C). [(B), (D), and (F)] Histograms of scattering strength values at locations within the shoal (areas inside the dashed box) and in a background region (areas inside the solid box) plotted for comparison. The 735 Hz data are ambient noise limited in background areas due to weak source level and is not shown.

drawn about background scattering because the 735 Hz data were dominated by ambient and nearby shipping noise since the source level for this frequency was lower. The fish target strength difference between 950 and 735 Hz is roughly 7.5 dB. These results are tabulated in Table IV.

As the OAWRS imaging frequency increases from 415 to 1125 Hz, the target strengths of both fish populations and background levels also increase. From the histograms of Fig. 5, the scattering strength increase with frequency is greater for the fish shoals than background levels, making it easier to detect fish aggregations at the higher frequency. At 415 Hz, only densely populated fish regions are distinguishable from the background. Fish densities at shoal peripheries are typically too low to be detectable.

The background levels in Figs. 7(A)–7(C) can be used to derive the minimum detectable fish densities in the OAWRS system at various frequencies. From Fig. 5(D), if we require that fish returns stand at least 1 standard deviation above the background mean to be detectable, then scattering strength levels that are above roughly −52 dB at 950 Hz would be detectable. This corresponds to a minimum detectable areal density of roughly 0.1 fish m^{-2}. The minimum detectable densities at the other frequencies are tabulated in Table V. These are based on scaling the fish densities up or down depending on the background mean scattering strength level at the other frequencies (Fig. 7) and also accounting for the target strength differences. These results are consistent with those obtained from analyzing the histograms in Fig. 5.

IV. RESULTS AND DISCUSSION

A. Measured abundance

Herring areal population densities and abundances are estimated by subtracting the estimated target strength expected of an individual herring from OAWRS scattering strength images as explained in Sec. III B. Areal fish density is first calculated by applying Eq. (5) at each pixel in an OAWRS image. Integrating over the area of a shoal then provides an estimate of its abundance. We illustrate abundance estimation with OAWRS images generated at 950 Hz since these have the best spatial resolution.

Figures 4(A)–4(C) show the areal fish density for a sequence of instantaneous OAWRS images close to the midnight hours of October 2 where average density within the shoal often exceeds 10 fish m^{-2}. Figure 4(H) shows the population over time within the area shown in Figs. 4(A)–4(C) for consecutive OAWRS images from 23:30 EDT on October 2 to 01:00 EDT on October 3. When no threshold is applied, we simply integrate the densities throughout the area. The total population of roughly 40×10^6 fish includes fish in the shoaling region as well as diffuse fish clouds outside the shoal. Contributions from outside the fish shoal, which may include background from the seafloor, are estimated to account for less than 10% of the total population in the area shown.

To exclude background reverberation, a density threshold is selected to segment shoaling regions where fish scat-

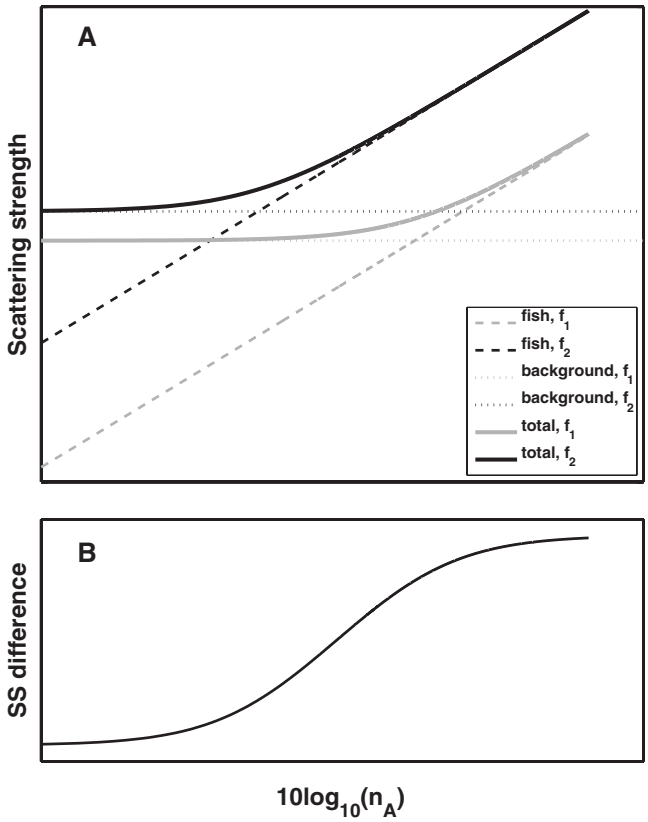


FIG. 6. (A) Schematic of scattering strength levels at two distinct frequencies f_1 and f_2 as a function of local fish density and (B) their difference at a given OAWRS pixel. The scattering strength difference equals the mean background level difference for low fish densities, while at high fish densities, the scattering strength differences equals the fish target strength difference.

tering is dominant. Figure 4(H) shows the total population of fish above and below various density thresholds, 0.1, 0.2, and 0.5 fish m^{-2}. The optimal density threshold for segmenting the shoaling region is roughly 0.2 fish m^{-2} since below this threshold populations stay fairly constant as expected for background levels. This threshold also corresponds to a scattering strength of roughly −50 dB which is roughly the value where the scattering strength histograms for background and

TABLE V. Minimum detectable fish density (M.D.D) in OAWRS imagery.

f_c (Hz)	415	950	1125
M.D.D (fish m^{-2})	4	0.1	0.04

shoaling regions intersect in Fig. 5. The population of fish in the shoaling region varies over time between 30×10^6 and 40×10^6.

The massive shoal shown in Fig. 2(A) is comprised of over 240×10^6 fish, with roughly 170×10^6 in the large consolidated shoal and 70×10^6 in the diffuse fish aggregation region. The population of fish in the shoals of Figs. 2(B) and 2(C) are roughly 86×10^6 and 70×10^6, respectively. As discussed in the Appendix D, the uncertainty in the abundance estimate is roughly 17–20%.

B. Measured low-frequency target strength

Following the approach of Secs. III D and III E, the target strength corresponding to the mean scattering cross-section of an individual shoaling herring is estimated as a function of frequency in the 300–1200 Hz range of the OAWRS system from measured scattered returns and measured and modeled transmission loss as shown in Fig. 10 and Table IV. The target strength data show a consistent dependence in both level and roll-off with decreasing frequency of roughly 20 dB/octave for all measurements, which spanned five shoals on 4 days. The invariance of the results from shoal to shoal and day to day is consistent with the low measured standard deviations obtained for each shoal by stationary averaging. The very strong roll-off in frequency is consistent with that found just below the resonance peak of a system undergoing damped harmonic oscillation.[38]

C. Using measured low-frequency target strength to infer swimbladder properties

Air-filled swimbladders typically comprise roughly 5% of fish body volume at neutral buoyancy depth.[39] To remain neutrally buoyant as hydrostatic pressure changes with depth, fish need to regulate the amount of gas in their swimbladders to maintain the 5% ratio.[40] Given this ratio and total

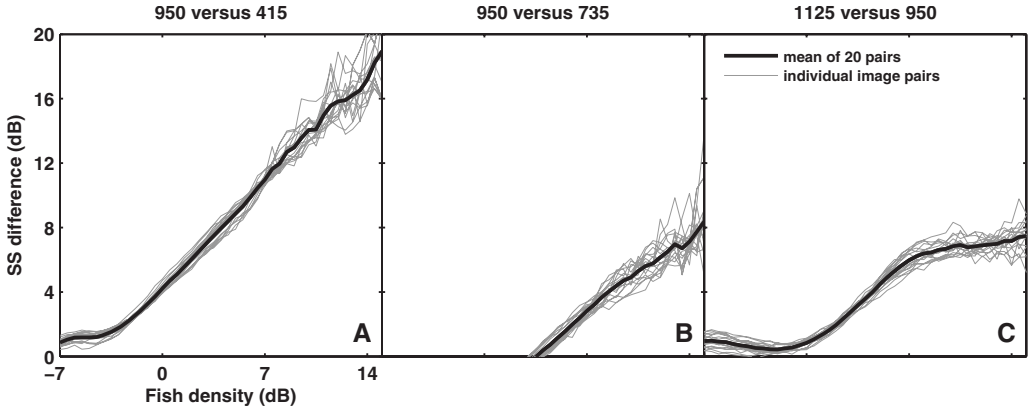


FIG. 7. OAWRS scattering strength level differences for the indicated frequency pairs as a function of areal fish density in dB *re* 1 m^2 for data acquired between 22:00 and 22:45 EDT on October 3. The scattering strength difference at high fish densities equals the target strength difference for the frequency pair shown.

fish volume, swimbladder volume can be determined at any depth as can neutral buoyancy depth if the relationship between pressure and volume is known for the swimbladder. One highly plausible relationship is Boyle's law,[41,42] where the product of pressure and volume remains constant, which has been demonstrated in the laboratory with a single dead herring,[43] but has not been directly confirmed in the wild where it is difficult to make *in situ* measurements of the physiology of free-ranging fish at depth.

Herring are physostomes, fish with open swimbladders connected to the gut and colon.[44] There are three hypotheses by which herring inflate their swimbladders: (1) by gulping air at the surface,[44–47] (2) by bacterial fermentation in the gut,[44,48] and (3) by secretion of gas from the blood stream into the swimbladder.[49–52] Nøttestad[52] found hypotheses (1) and (2) implausible for the Norwegian spring spawing herring in his study that remain in deep layers, are not observed near the surface, and are not feeding. This led him to suggest hypothesis (3). Nero *et al.*[7] conducted on site experiments on the northern flank of Georges Bank with spawning herring by adding weights to captured herring until they sank. They concluded "that these herring contained up to at least a three times greater volume of gas than a neutrally buoyant fish at the sea surface," and arrived at neutral buoyancy depths as great as 60 m from these and low-frequency acoustic target strength measurements. They found their results to be consistent with measurements of Pacific herring in Puget sound,[48] and suggested hypotheses (2) and (3) as plausible explanations for their observations. Similarly, Fänge[51] supported hypothesis (3) by noting that "The herring lacks obvious gas depository structures (*rete mirabile*, gas gland), but has relatively high O_2 values in the swimbladder (up to 21.5%),[53] and observations of release of gas bubbles from vertically migrating herring[54] indicate that some gas secretion may occur."

The conditions of the present experiment were not only similar to those of Nero *et al.*,[7] but also to those of Nøttestad[52] in that the spawning herring were only observed in deep layers and not near the surface where it is unsafe due to predator attack as both Nøttestad[52] and Makris *et al.*[3] noted, and a vast majority of the herring captured in trawl samples were observed to have no large prey (copepods) in their stomachs. The latter point, however, needs to be tempered because bacterial content was indeterminate in the samples, food resources are plentiful near the seafloor where the herring shoals of the present study were found, and gas production by bacteria in herring stomachs can last more than 90 h (Ref. 55) after ingestion. It is typically associated with a delay due to phase lag in bacterial growth.[55] Since the diffusion rate of gas out of the swimbladders of caged herring at fixed depth is also found to be small,[40] corresponding to less than a 0.3 dB decrease in target strength per day, herring may maintain bacterial gas for long periods with minimal feeding on large prey. Such feeding is known to increase with gonadal development.[56] These facts suggest that in addition to Nøttestad's[52] hypothesis (3), hypothesis (2) may also remain highly plausible for our experiment as suggested by Nero *et al.*[7] for a similar location and season.

Here we compare the estimated low-frequency target strength obtained from experimental data with that derived from Love's widely used model for resonant scattering from a fish swimbladder.[6,7,57] This comparison enables estimates of swimbladder volume to be inferred for the shoaling herring observed in this study. Love models the fish swimbladder as an elongated-spheroidal, viscous, heat-conducting shell which encloses an air cavity with surface tension at the inner surface.[6,57] This leads to well-understood damped resonance behavior. While the material in and around the swimbladder has more complex elastic composition and structural constraints than that in the Love model that could lead to more complex scattering, the Love model has been successfully tested in experimental settings where strong resonances have been observed,[6,7,57–61] and probably provides an accurate description of the first order physics near resonance. Following Refs. 7 and 61–64, the swimbladder is modeled as a resonant, air-filled prolate-spheroid with a major-to-minor-axis ratio of 5–10 and major-axis to fish length ratio of 0.13–0.17 based on our trawl samples and CFFS measurements for herring. Only changes in minor-axis are assumed to contribute to swimbladder volume change due to physiological constraints.[7,45,65,66] We use the herring length distribution [Fig. 8(A)] measured from trawl samples and depth distributions determined by CFFS line-transects. The fish weight (W) to length (L) relationship is approximated by a normal distribution with a mean given by an empirically determined length-weight regression (gray line) from length-weight measurement of 1219 herring samples and a standard deviation of ±20% of the mean calculated from the length-weight data (dots), as shown in Fig. 9. Properties of modeled fish, such as the flesh density, viscosity, and swimbladder volume at depth are tabulated in Table III. Given these constraints, swimbladder volume or equivalently swimbladder semi-minor-axis is the only unknown variable in the Love model

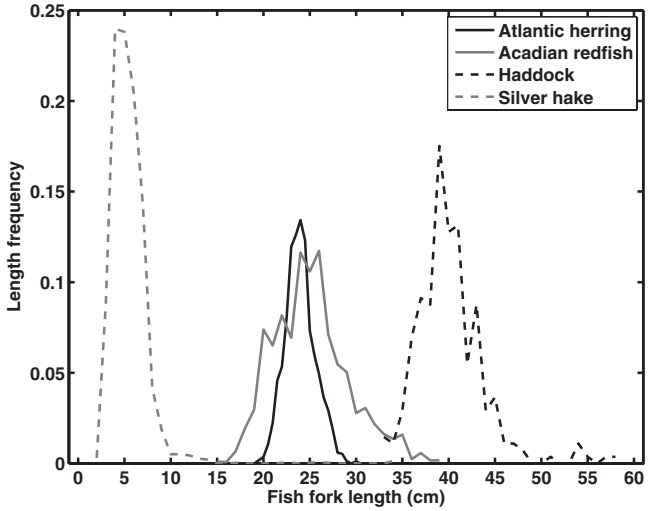


FIG. 8. Fork length distributions of most frequently caught species, Atlantic herring, Acadian redfish, Haddock, and Silver hake, from trawls deployed on Georges Bank (Fig. 1). The mean fork length of herring is 24.2 cm with a standard deviation 6.8% of the 24.2 cm mean. The equation $L_{TL}=1.103L_{FL}+0.01$ (Ref. 7) is used to convert herring's fork length to the total length, where L_{TL} and L_{FL} are in cm. The mean fork length of redfish is 26.2 cm with a standard deviation 15% of the 26.2 cm mean. The equation $L_{TL}=1.033L_{FL}-0.038$ (Ref. 72) is used to convert redfish's fork length to total length. Silver hake's fork length is the total length.

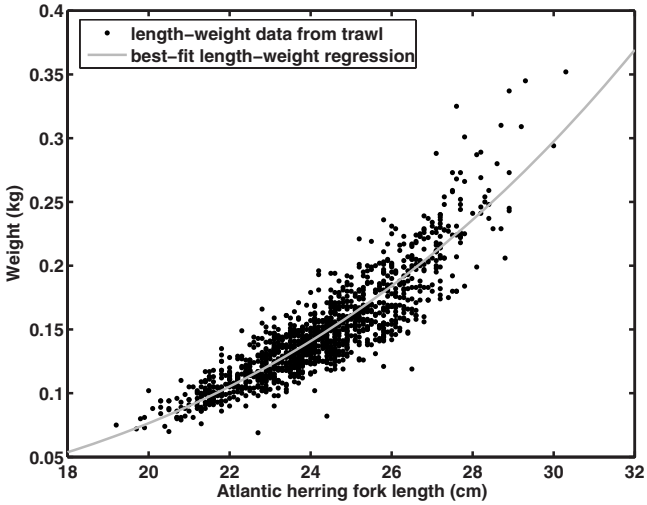


FIG. 9. Atlantic herring length-weight regression calibration. The dots are the length-weight data obtained from the trawl-survey conducted by U.S. National Marine Fisheries Service in conjunction with OAWRS 2006 experiment, and the gray solid line indicates the derived best-fit length-weight regression, which can be expressed as $W=aL^b$, where W is the weight of herring in kg, L is the fork length of herring in cm, and a and b are empirical regression parameters. For this trawl dataset, $a=3.35\times10^{-6}$ and $b=3.35$.

that can lead to a significant change in fish target strength at and below the resonance peak. Swimbladder volume is assumed to vary with pressure according to Boyle's law,[41,42] from which neutral buoyancy depth can be uniquely determined. Neutral buoyancy depth is then assumed to be a Gaussian random variable with mean and standard deviation determined by least-squares fit between measured and modeled target strengths.

The best least square fits between our measured target strength data and the Love model appear in Figs. 10(A)–10(E) for five fish shoals imaged by OAWRS on four different days, and consistently show a broad resonance peak with maximum at roughly 1.5–1.7 kHz and swimbladder semi-minor-axis of 3–5 mm. The model to data match is excellent, with negligible mean-squared error, which is significant because it would not be possible to obtain a good match between the measured frequency dependence and the Love model if the overall level of the measured target strength data had a significant scale factor error that was much larger than measured errors of roughly 1 dB. This consistency gives added confidence to the veracity of both the data and model. As expected from the roughly 20 dB/octave frequency roll-off of the data, the best fit of the model is for a resonance just above the highest frequency data point available in the current set of measurements. The model fits of Fig. 10 can be interpreted with the aid of Table VI which shows the volume and corresponding swimbladder minor axes radii given the measured herring length distributions, as well as possible neutral buoyancy depths based on Boyle's law. Neutral buoyancy depths were found to correspond to roughly half the mean shoal depth given the measured herring length distribution, spanning 20–34 cm with a mean of 26.7 cm, and depth distribution in a layer between 120 and 190 m from trawl and CFFS sampling.

The localized measurements of Nero *et al.*[7] for herring target strength data in the 1.5–5.0 kHz range show target strength levels consistent with those found in our best-fit curve for frequencies above roughly 2.2 kHz. This can be seen by noting their best-fit neutral buoyancy depth curve (50 m) falls within 1–2 dB of ours for all examples above roughly 2.2 kHz for the measured fish depth, length, and density distributions of this study. The Love-model based neutral buoyancy depths and resonances found here are also consistent with those measured by OAWRS in 2003 south of Long Island, NY, in scattering from shoals that evidence suggest were also Atlantic herring.[2,67] Arbitrarily constraining the neutral buoyancy depth to be near the surface, say, at 4 m and using Boyle's law for fish at 120–190 m depth yields a corresponding 3.7 kHz resonance and leads the Love model to a herring target strength 20–30 dB lower (Fig. 11) than that measured by the OAWRS system. Such low target strengths are not only inconsistent with the target strength measurements of Fig. 10 and Nero *et al.*,[7] they also violate conservation of energy since the corresponding transmission losses required to match our measured sound pressure levels returned from the fish shoals would have to be less than those found in even a perfectly reflecting waveguide without any medium attenuation, i.e., the waveguide would have to somehow add energy to the source signal by two to three orders of magnitude. This can be seen in Fig. 14 which shows the measured transmission loss to be within 2–3 dB of that found in a perfectly reflecting waveguide without attenuation.

A wide distribution of swimbladder volumes and corresponding neutral buoyancy depths within any shoal is likely and could potentially unify the various existing data sets by superposition, with larger swimbladder volumes dominating at the lower end of the possible resonance spectrums and smaller volumes dominating the higher end. None of the systems used in the field so far, however, could test this since it would require acquisition of simultaneous data both well below and well above all contributing resonant frequencies. Since low-frequency target strength measurements near resonance are far more sensitive to small changes in swimbladder minor-axis or volume than at much higher CFFS frequencies, such lower frequency measurements have the potential to resolve the *in situ* swimbladder volumes of wild herring at depth with much greater accuracy. The measurements of target strength for herring as a function of depth reported by Ona[37] at 38 kHz, for example, show standard deviations of 8–10 dB and minimal 3–5 dB reduction from the surface to 300 m,[66] which falls well within these standard deviations. These standard deviations significantly exceed the 5.6 dB expected purely from stationary Gaussian field fluctuations by the Central limit theorem,[26] shown to apply to scattering from fish of random orientation by Dahl and Mathisen.[68] This may suggest a bias in fish orientations or a wide spread in swimbladder volumes in the Ona[37] data, which could easily include the span of volumes measured here and in Ref. 7. Fässler *et al.*[43] suggested that projected dorsal area variations exhibit greater depth variation than the mean target strength measurements of Ona.[37] We find the Ona[37] mean trend in depth to be consistent with that expected from Boyle's law since at their frequencies ka, the product of acoustic wavenumber to the semi-minor-axis of the swimbladder, is typi-

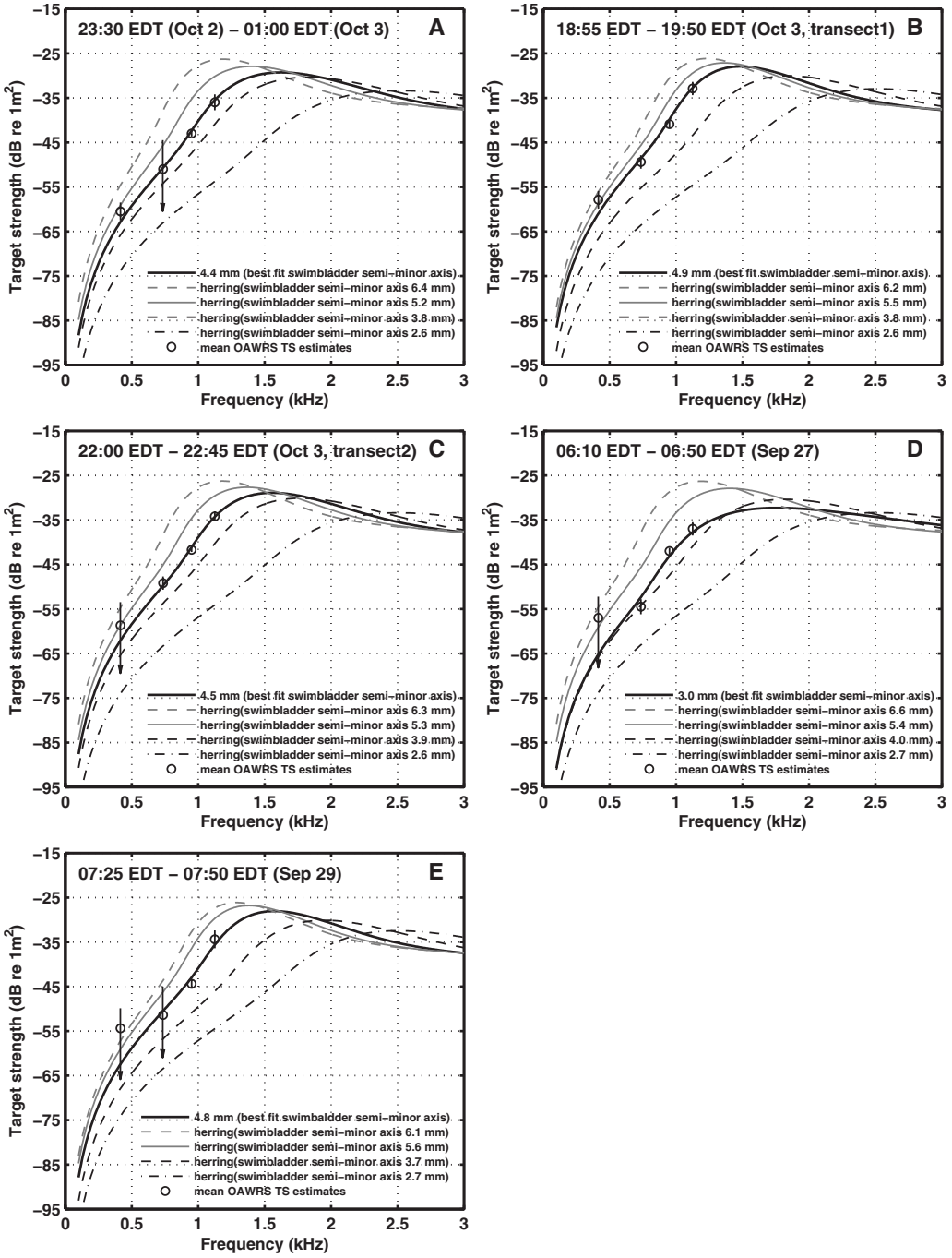


FIG. 10. Experimentally determined low-frequency target strength corresponding to the average scattering cross-sections of shoaling herring observed from OAWRS imagery acquired from five shoals on 4 days (Table VI) at 415, 735, 950, and 1125 (circles) with standard deviations (error bars). Comparison with Love-model mean target strength for shoaling herring, with physical parameters tabulated in Table III, of different swimbladder semi-minor axes over the shoals' depth distributions (lines). The best least-squares fits shown are obtained only using target strength estimates of each shoal (Table IV) with standard deviations less than 3 dB. Arrows indicate the target strength uncertainties due to potential masking from background scattering (Sec. III E) for given frequencies. The best-fit means and standard deviations of inferred swimbladder volume, swimbladder semi-minor axes, and corresponding neutral buoyancy depths of each shoal were tabulated in Table VI. (A) Shoaling herring, distributed between 120 and 185 m (Fig. 4), imaged with the OAWRS system from 23:30 EDT October 2 to 01:00 EDT on October 3. (B) Shoaling herring, distributed between 135 and 175 m, imaged with the OAWRS system from 18:55 to 19:50 EDT on October 3. (C) Shoaling herring, distributed between 120 and 175 m, imaged with the OAWRS system from 22:00 to 22:45 EDT on October 3. (D) Shoaling herring, distributed between 120 and 185 m, imaged with the OAWRS system from 06:10 to 06:50 EDT on September 27 (Fig. 13). (E) Shoaling herring, distributed between 150 and 180 m, imaged with the OAWRS system from 07:25 to 07:50 EDT on September 29.

cally less than unity and target strength is no longer proportional to projected area as in the large *ka* limit.

Evidence from the extensive OAWRS, CFFS, and trawl surveys conducted during OAWRS 2006 offers no plausible alternative to herring as the primary constituent and source of scattering in the shoals imaged by OAWRS. Consideration of the effect of two other species, however, also present in many trawls but in far fewer numbers (Table VII), still provides useful perspective. These are silver hake and Acadian redfish. While silver hake were found at shallower depths than herring, their much shorter lengths (Fig. 8) lead to individual target strengths more than 20 dB lower than those

TABLE VI. Experimentally inferred means and standard deviations of swimbladder volume, $\hat{v_{sb}}$ and $\sigma_{v_{sb}}$, semi-minor-axis, $\hat{a_z}$ and σ_{a_z}, over the depth distributions of the shoals, and corresponding means and standard deviations of neutral buoyancy depth, $\hat{d_{nb}}$ and $\sigma_{d_{nb}}$, where neutral buoyancy depth is restricted to water-column depths of 0–200 m in the least squares fit. All three parameters are assumed to be Gaussian random variables completely characterized by their respective means and standard deviations.

Date	Time[a]	Layer depth (m)	$\hat{v_{sb}}$ (ml)	$\sigma_{v_{sb}}$ (ml)	$\hat{a_z}$ (mm)	σ_{a_z} (mm)	$\hat{d_{nb}}$ (m)	$\sigma_{d_{nb}}$ (m)
Oct 2	23:30–01:00(Oct 3)	120–185	3.41	0.41	4.4	0.26	82	11
Oct 3	18:55–19:50	135–175	4.27	0.31	4.9	0.18	108	9
Oct 3	22:00–22:45	120–175	3.62	0.38	4.5	0.24	85	10
Sep 27	06:10–06:50	120–185	1.61	1.49	3.0	0.96	31	48
Sep 29	07:25–07:50	150–180	3.98	0.21	4.8	0.12	107	6

[a]Time periods during which both OAWRS and CFFS systems simultaneously co-registered contiguous shoal segments. The time is in EDT.

measured at the location of the herring shoals at OAWRS frequencies (Fig. 11). Given that (1) both CFFS and trawl samples found silver hake in considerably lower areal densities than herring, (2) their individual target strengths are roughly 20 dB lower, and (3) OAWRS transmission loss was much greater in the shallower layers where silver hake resided as part of the experimental design, contributions from silver hake can be ruled out as a plausible explanation for shoals imaged by OAWRS. As shown in Fig. 11(A), the presence of redfish in the maximum percentages determined by trawl has a negligible effect on the best-fit target strength compared to that of modeling herring alone [Figs. 10(A)–10(E)]. This should be expected because the low-frequency dependence of background returns when no shoals are present is effectively negligible compared to the dependence when shoals are present (Fig. 5) and simultaneous trawls showed the shoals to be overwhelmingly comprised of herring but the background to yield negligible amounts of herring (Table IV, Fig. 12). So, even though redfish are physoclists with trawl sample lengths (Fig. 8) typically greater than those of the herring, the measured areal densities of redfish are too low to have a significant impact on the average scattering cross-section measured in the observed shoals. Including greater percentages of redfish in the modeling or including measured percentages and shallower neutral buoyancy depths for herring leads to far worse fits and high mean-square error when matched with the data, as shown in Fig. 11(B).

D. Space, time, and frequency dependencies

From another perspective, since the spatio-temporal population distributions of the large shoals versus background levels are consistent among OAWRS, CFFS, and trawls, as shown here and in Ref. 3, the spatio-temporal distributions of silver hake, redfish, or any other contaminant

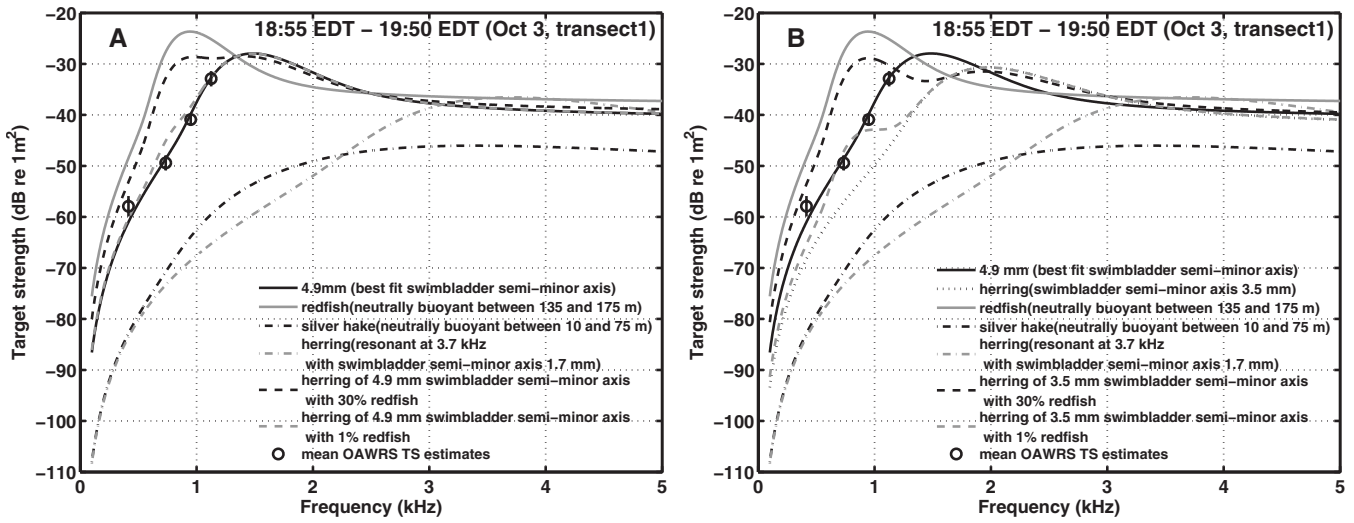


FIG. 11. (A) An example of Love-model target strength corresponding to the average scattering cross-section of an individual for mixed species content and swimbladder semi-minor axes shown. Comparison of Love-model target strength with experimentally determined mean target strength estimates of shoaling herring, distributed between 135 and 175 m, imaged with the OAWRS system from 18:55 to 19:50 EDT on October 3. Presence of redfish in trawl-determined percentage (dashed gray line) has negligible effect on best-fit target strength compared to herring alone (solid black line), while including an unrealistic percentage of redfish yields far worse fits (dashed black line). The silver hake, found at shallower water depth (<75 m) in trawls, have resonance peak above 3 kHz (dash-dot curve) making their contribution negligible. Herring target strength with a resonance frequency at 3.7 kHz (dashed gray line) based on Love's model using length and depth distribution obtained from CFFS and trawl measurements is found to be neutrally buoyant at 4 m, and is 20–30 dB lower than those measured by the OAWRS system. (B) Same as (A) but for shallower herring neutral buoyancy depths. Arrows indicate potential target strength uncertainties for given frequencies.

TABLE VII. Concurrent high-speed rope trawl deployed by NOAA FRV *Delaware II* within or in the vicinity of large herring shoals imaged by OAWRS during the OAWRS 2006 experiment in the Gulf of Maine and Georges Bank at shoal depth. The number of most frequently caught species in each trawl deployment, including Atlantic herring, Acadian redfish, Silver hake, and Haddock are tabulated. Trawls 134, 137, and 139 were made with simultaneous OAWRS imagery, and trawls 137 and 139 were made directly through shoals as shown in Figs. 12(B) and 12(C). In contrast, trawl 134 was made in a region with no shoal formed (Fig. 12(A)), but one shoal would form in the vicinity 4 h later. Trawl 105 was made through shoals imaged by OAWRS 95 min before, and trawl 106 was made through shoals imaged by OAWRS 20 min later.

Deployment	Atlantic herring	Acadian redfish	Silver hake	Haddock	Total catch
105	8030 (99.98%)	0	0	0	8032
106	634 (96.79%)	0	14 (2.14%)	2 (0.31%)	655
134	3 (11.54%)	0	14 (53.85%)	0	26
137	333 (76.03)	0	94 (21.46%)	0	438
139	796 (74.74%)	9 (0.85%)	208 (19.53%)	23 (2.16%)	1065
Total catch of experiment	9796 (96.54%)	9 (0.07%)	330 (2.67%)	25 (0.2%)	10216

species would have to consistently follow those of herring if they were a major contributor to OAWRS returns, which is both implausible and contrary to trawl and CFFS data as well as the frequency dependence of OAWRS returns in shoals versus background (Fig. 5). Also, neither our trawl nor CFFS data show any evidence of shallow fish layers that could account for the prominent OAWRS returns that co-registered in space and time with the deep shoals measured by simultaneous trawl and CFFS sampling. Indeed, considering experimental causality, it was consistently necessary for OAWRS to *first find* and *then direct* trawl and CFFS vessels to the locations of these deep shoals because they are so difficult to find with conventional methods given the fact they occupy areas many orders of magnitude smaller than the wide areas over which they may roam.

V. CONCLUSION

The low-frequency target strength of Atlantic herring (*Clupea harengus*) is estimated from experimental data acquired from shoaling herring in the Gulf of Maine during the Autumn 2006 spawning season in the 300–1200 Hz range using simultaneous ocean acoustic waveguide remote sensing, conventional fish finding sonar, and trawl surveys. The target strength expected of an individual is found to have a strong nonlinear dependence on frequency consistent with resonant scattering from an air-filled swimbladder given measured fish length, depth distributions, and experimentally inferred swimbladder volumes based on Love's model, which indicate the herring remain negatively buoyant in layers near the seafloor for extended periods. The OAWRS system used in this study employed an instantaneous imaging diameter of 100 km with regular minute-to-minute updates enabling unaliased monitoring of fish populations over ecosystem scales. This included detection and imaging of shoals of Atlantic herring containing hundreds of millions of individuals, as confirmed by concurrent trawl and CFFS surveys that were directed to the shoals' locations by OAWRS. High spatial-temporal coregistration was found between herring shoals imaged by OAWRS and concurrent CFFS line-transects.

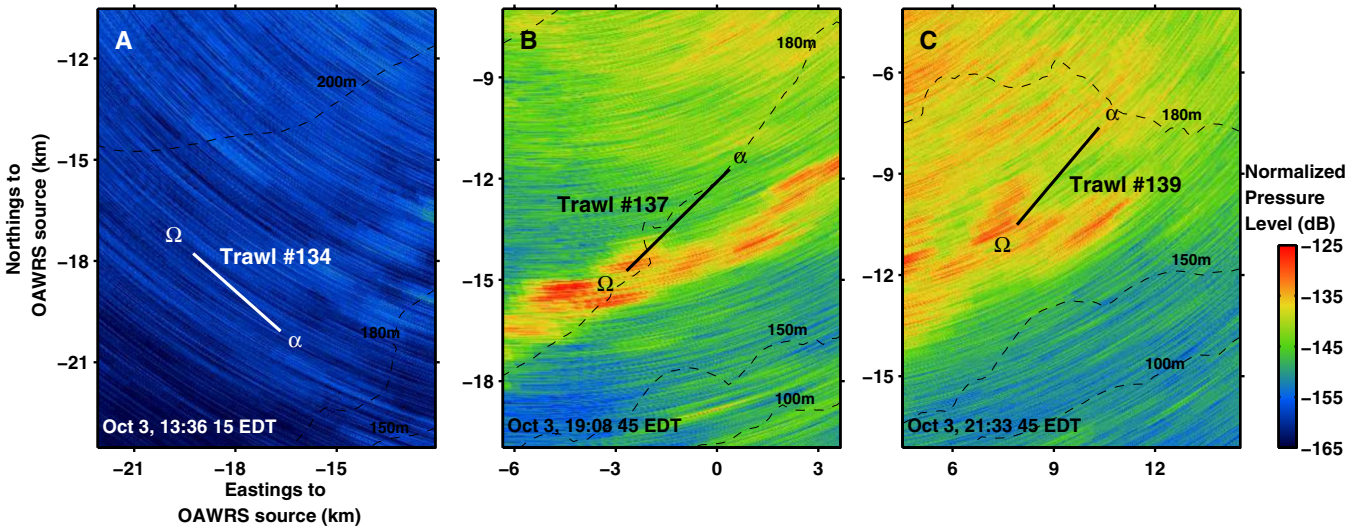


FIG. 12. (Color online) [(A)–(C)] Locations of trawls over simultaneous OAWRS images. Trawls 137 and 139 were made directly through shoals as shown in (B) and (C). In contrast, trawl 134 was made in a region with no shoal (A), but one shoal would form in the vicinity 4 h later. The OAWRS source locations are the coordinate origin in all OAWRS images. On October 3, the OAWRS source ship was moored at 42.2089N, 67.6892W. The trawls in (A)–(C) were deployed and towed along the solid lines starting at α and ending at Ω. The dashed lines indicate the contours of 100, 150, 180, and 200 m water depth.

TABLE VIII. Concurrent high-speed rope trawl deployed by NOAA ship FRV *Delaware II* within or in the vicinity of large herring shoals imaged by OAWRS system during the OAWRS 2006 experiment in the Gulf of Maine and Georges Bank at shoal depth. The dates, times (Eastern Daylight Time), deploy depths, and geographic locations of the trawls are tabulated.

Deployment	Date	Time	Deploy depth (m)	Begin Lat	Begin Lon	End Lat	End Lon
105	09/26	14:21:40-15:16:32	130–160	41° 53.00N	68° 06.62W	41° 51.95N	68° 10.45W
106	09/26	19:23:24-19:57:20	130–160	41° 55.49N	68° 03.80W	41° 55.05N	68° 06.33W
134	10/03	13:21:12-13:54:01	140–170	42° 01.80N	67° 53.39W	42° 03.06N	67° 55.24W
137	10/03	18:47:09-19:27:04	140–170	42° 06.36N	67° 41.02W	42° 04.76N	67° 43.27W
139	10/03	21:16:35-21:52:46	140–170	42° 08.54N	67° 33.83W	42° 07.00N	67° 35.63W

ACKNOWLEDGMENTS

This research was supported by the National Oceanographic Partnership Program, the Office of Naval Research, the Alfred P. Sloan Foundation, and the National Oceanic and Atmospheric Administration and is a contribution to the Census of Marine Life. We would like to thank the science parties, officers, and crew of the research vessels *Oceanus, Endeavor, Delaware II*, and *Hugh Sharp*. We would also like to thank Olav Rune Godø and Redwood W. Nero for the many interesting discussions on the topic, and the research assistants who helped in the data analysis.

APPENDIX A: CONCURRENT NATIONAL MARINE FISHERIES SERVICE ANNUAL ATLANTIC HERRING ACOUSTIC SURVEY IN THE GULF OF MAINE

In conjunction with OAWRS 2006 experiment, U.S. National Marine Fisheries Service conducted annual Atlantic herring survey of the Gulf of Maine and Georges Bank. Five trawls, 105, 106, 134, 137, and 139, targeted at 130–170 m depth (Table VIII) were deployed within or in the vicinity of the large fish shoals imaged by OAWRS system, which enabled onsite species identification and biological measurements. The locations of these trawls are shown in Fig. 1. The trawls 134, 137, and 139 were made with simultaneous OAWRS imagery, and trawls 137 and 139 were made directly through shoals, as shown in Figs. 12(B) and 12(C). As tabulated in Table VII, Atlantic herring is found consistently to be the overwhelmingly predominant species in trawls 105, 106, 137, and 139. Trawl 134, shown in Fig. 12(A), was deployed in a region with no shoal formed, but one would form in the vicinity 4 h later, and like both OAWRS and CFFS does not show high concentrations of herring before the shoal formed. Trawl 105 was made through shoals imaged by OAWRS 95 min before, and trawl 106 was made through shoals imaged by OAWRS 20 min later. In addition, a small percentage of juvenile silver hake, Acadian redfish, and haddock are also present in the trawls. Most of the juvenile silver hake were caught at shallower water depth (<75 m) as the deeper trawls were deployed. Histograms of the measured length distributions of the most frequently caught species are plotted in Fig. 8. The conversion of herring length to weight by regression analysis is given by $W=0.003\,35L^{3.35}$, based on length-weight measurements of 1219 herring caught in trawls throughout the course of the survey.

APPENDIX B: ESTIMATING LOW-FREQUENCY TARGET STRENGTH OF SHOALING HERRING WITHOUT CALIBRATED TARGET DEPLOYMENT

Here we present the estimated low-frequency target strength of a shoaling herring based on simultaneous OAWRS and CFFS sampling line-transects of large shoals on September 27 and 29, 2006 when no calibrated targets were available to aid in charting. Only minor adjustments of less than 1.5° from the array heading sensors were needed to ensure optimal coregistration between OAWRS and CFFS locations of shoals, which were typically the only landmark available in the absence of the calibrated targets. The target strength estimates shown here are consistent with those presented for the other days when calibrated targets were available in the main text.

During the early morning hours of September 27, between 06:10 and 06:50 EDT, both OAWRS and CFFS coregistered a massive herring shoal 10–15 km southeast of the OAWRS source, spanning more than 2 km in range and 5 km in azimuth, as shown in Figs. 13(A)–13(C). A dense layer of herring, shown in Fig. 13(D), was found consistently spanning 120–170 m in the water column by simultaneous CFFS line-transects. The resulting target strength estimates at 950 Hz obtained by means of the approach described in Sec. III D are plotted in Fig. 13(G). The target strength estimates at 415, 735, and 1125 Hz are tabulated in Table IV using the approach described in Sec. III E, where target strength at 950 Hz is used as the base frequency. No conclusion can be made for the target strength difference between 950 and 415 Hz, because fish density is not high enough to make scattering from herring at 415 Hz distinguishable from background levels. The same approaches are also applied to estimate the target strength expected for an individual shoaling herring imaged on September 29 between 07:25 and 07:50 EDT. The results are provided in Table IV.

APPENDIX C: STATISTICAL ANALYSIS FOR LOW-FREQUENCY TARGET STRENGTH OF HERRING

Here we describe the statistical approach used to estimate the mean and standard deviation of the low-frequency target strength of fish populations imaged concurrently by OAWRS and CFFS. We confine our analysis to regions where both systems image continuous segments of statisti-

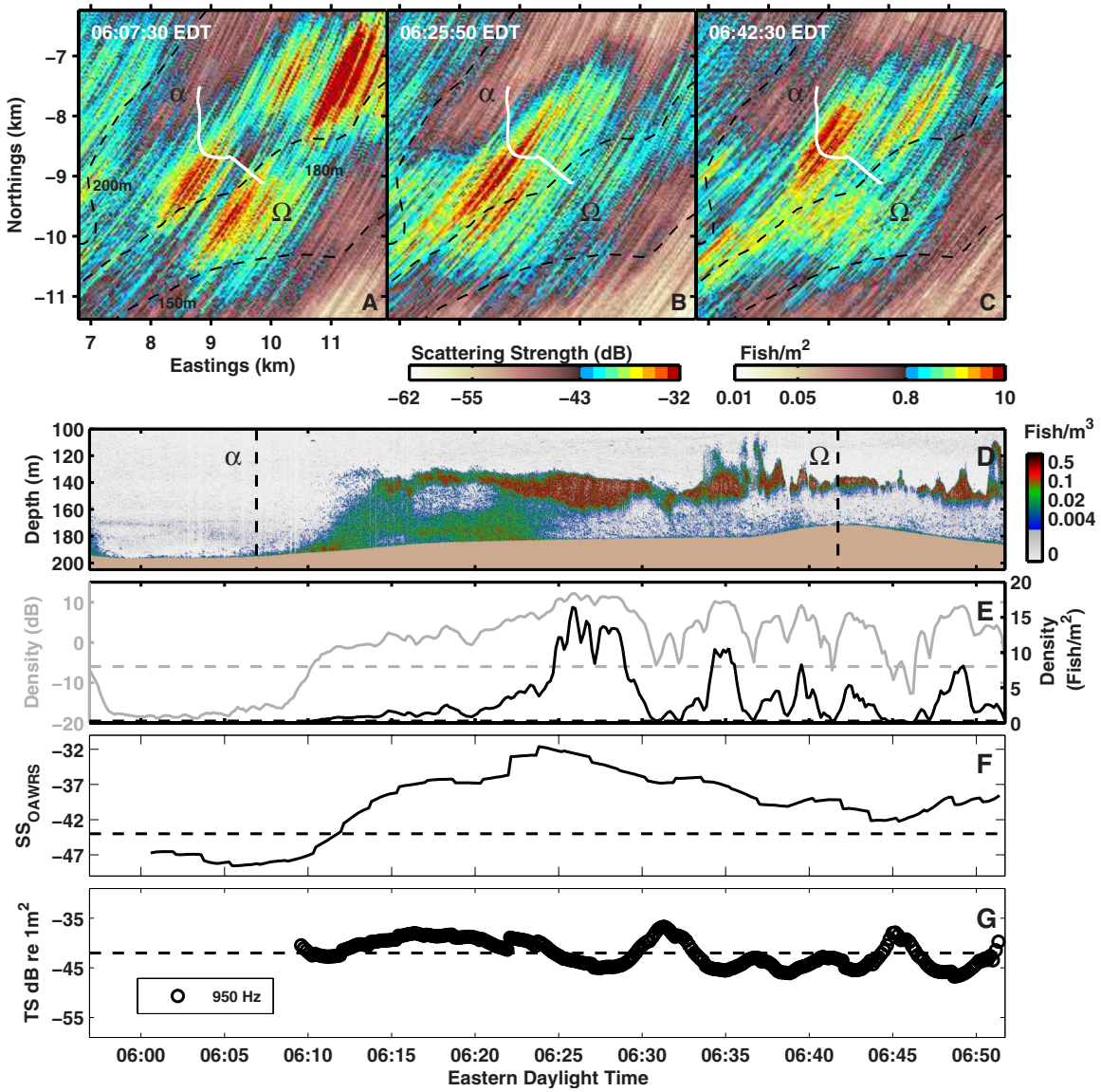


FIG. 13. (Color online) Herring target strength at 950 Hz estimated by matching the areal density of OAWRS and CFFS data acquired during the early morning hours of September 27. Similar to Fig. 4 but for a contiguous shoal 40-min segment starting at 06:10 EDT on September 27.

cally stationary fish populations that span many resolution cells in both the CFFS and OAWRS systems.

Employing Eq. (5), let TS_i be independent estimates of low-frequency target strength obtained for each resolution cell after sufficient spatial and temporal averaging in anti-log to eliminate the bias when converted to log units and to insure that the TS_i are Gaussian random variables.[26] Assuming stationarity, let their means be $\overline{TS}$ and standard deviations be σ_{TS}. Then, given N independent measurements of the target strength along the CFFS transect, the resulting linear estimator for mean target strength is[69]

$$\widehat{\overline{TS}} = \frac{1}{N}\sum_{i=1}^{N} TS_i \tag{C1}$$

and the corresponding estimator for the standard deviation of the target strength distribution is obtained by[69]

$$\hat{\sigma}_{TS} = \sqrt{\frac{1}{(N-1)}\sum_{i=1}^{N}(TS_i - \widehat{\overline{TS}})^2}. \tag{C2}$$

The variance of the mean target strength estimate is

$$\mathrm{Var}(\widehat{\overline{TS}}) = \frac{\hat{\sigma}_{TS}^2}{N}. \tag{C3}$$

Applying Eq. (C3), the standard deviation in the mean target strength estimate for the data in Fig. 4 is roughly on the order of 0.7–0.8 dB.

APPENDIX D: QUANTIFYING ERROR IN ABUNDANCE ESTIMATES

Here we quantify the error in the abundance estimates from OAWRS data presented in this paper. From Eqs. (4) and (5), error in $n_{A,\mathrm{OAWRS}}$ at any pixel is caused by a fixed OAWRS target strength estimate, which is the same for all

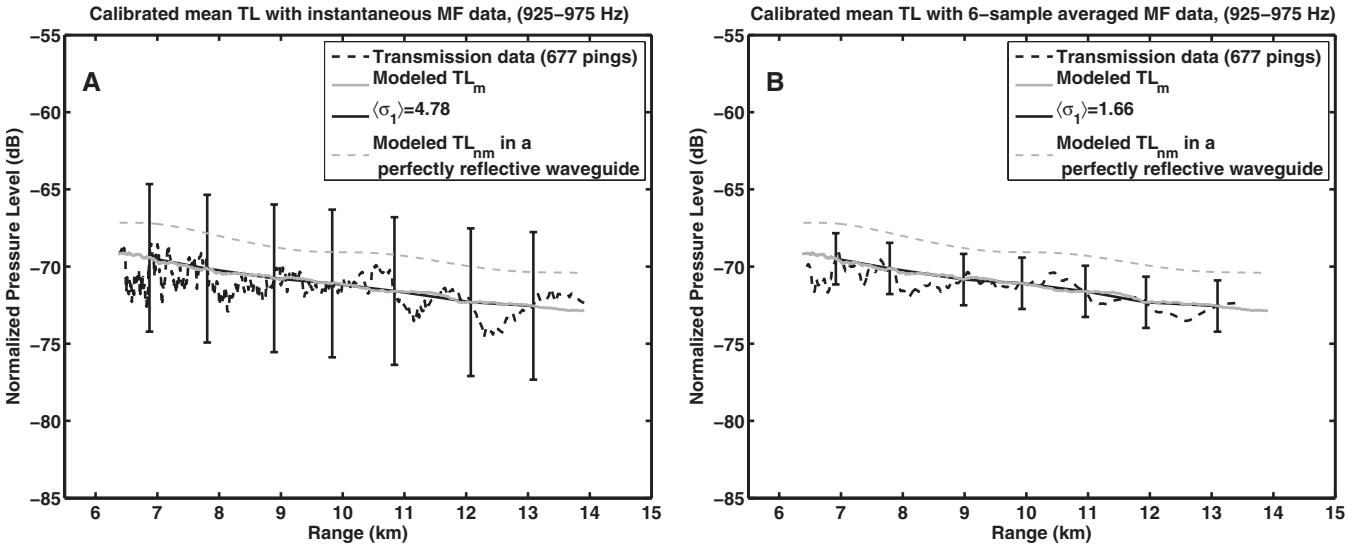


FIG. 14. (A) Experimentally determined mean and standard deviation (4.78 dB) of 677 measured instantaneous broadband one-way TL data after matched filter with 950 Hz center frequency. Plotted as a function of range, with modeled one-way TL overlain. Transmission data acquired by a single desensitized hydrophone on OAWRS receiver array on October 1–3 2006. Modeled TL computed by Monte-Carlo simulation with parabolic equation inputing measured oceanography and bathymetry of Georges Bank environment. Modeled TL (dashed gray line) in a perfectly reflective waveguide with no water-column attenuation by normal-mode model. (B) After six-sample averaging, the rms error is reduced to 1.7 dB, consistent with theory for stationary averaging of intensities of circular complex Gaussian random fields (Refs. 16 and 26).

pixels, and the OAWRS scattering strength, which varies from pixel to pixel. Summing large numbers of pixels reduces the percentage error from scattering strength fluctuations in the population density estimate to negligible values by the law of large numbers, leaving the error to be dominated by that in the OAWRS target strength estimate, which for the 950 Hz imagery is consistently between 0.7 and 0.8 dB throughout the experiment, corresponding to an error of 17–20% in total population.

APPENDIX E: CALIBRATION OF BROADBAND MEAN TL WITH ONE-WAY PROPAGATED EXPERIMENTAL DATA IN THE NORTHERN FLANK OF GEORGES BANK

In wide-area sonar applications, two-way TL must be efficiently and accurately estimated in order to invert for scattering strength or target strength of scatterers.[2,32] Here, RAM is used to estimate two-way TL expected over wide areas with measured bathymetry and oceanography of the northern flank of Georges Bank environment. For each OAWRS image, TL was computed along radials separated by roughly 1.5°, half the receiver array broadside resolution, for forward transmission from the source, and again for return transmission to the receiver array, by multiple Monte-Carlo realizations per radial. Each Monte-Carlo realization employed a different measured sound speed profile every 500 m to incorporate the effects of the fluctuating ocean waveguide.

In the OAWRS 2006 experiment, a very complete set of more than 12 000 transmission loss measurements spanning ranges from hundreds of meters to tens of kilometers, and more than roughly 200 sound speed profile measurements were made over the survey area. These were used to calibrate the parabolic equation model used to estimate wide-area transmission loss, population density, and target strength. Measured sound speed was found to follow a roughly Gaussian distribution about the mean at each depth and to be relatively uniform in time and space over the entire OAWRS imaging area (Fig. 3) with no evidence of horizontal features besides the expected short term fluctuations from internal wave activity.

The best fit between measured and modeled transmission loss was obtained for expected sandy bottom conditions of sound speed of 1.7 km s^{-1}, density of 1.9 g cm^{-3} and attenuation of 0.8 dB λ^{-1}, and in-water-column attenuation of 6×10^{-5} dB λ^{-1}, as shown in Fig. 14 using the maximum likelihood method.[16] Figure 14 shows the standard deviation of one-way TL for instantaneous broadband measurements is roughly 4.8 dB, and after six-sample averaging is on the order of 1.7 dB, following previous experimental results and theory.[16,26] We find the mean measured TL to be within 2–3 dB of that of a perfectly reflecting lossless waveguide of the same depth (dashed gray line in Fig. 14). The *scattered field from large fish shoals* after *two-way propagation in a waveguide* also obeys circular complex Gaussian field statistics by the central limit theorem and has experimentally measured standard deviation of roughly 1.5 dB, after our standard six-sample averaging, for the current experiment[4] which also follows that expected from theory and numerous past experiments where internal waves, ocean turbulence, eddies, and variations in scatterer orientation have been shown to cause short term intensity fluctuations with a standard deviation that can be reduced to a small fraction of the mean after stationary averaging.[20,21,24,26,67,70]

[1]http://www.gma.org/herring/ (Last viewed January 29, 2009).
[2]N. C. Makris, P. Ratilal, D. T. Symonds, S. Jagannathan, S. Lee, and R. W. Nero, "Fish population and behavior revealed by instantaneous continental shelf-scale imaging," Science **311**, 660–663 (2006).
[3]N. C. Makris, P. Ratilal, S. Jagannathan, Z. Gong, M. Andrews, I. Bertsa-

tos, O. R. Gødo, R. W. Nero, and J. M. Jech, "Critical population density triggers rapid formation of vast oceanic fish shoals," Science **323**, 1734–1737 (2009).
[4]N. C. Makris, P. Ratilal, S. Jagannathan, Z. Gong, M. Andrews, I. Bertsatos, O. R. Godø, R. W. Nero, and J. M. Jech, "Critical population density triggers rapid formation of vast oceanic fish shoals," materials and methods are available as supporting material on *Science* online (2009).
[5]B. S. McCartney and A. R. Stubbs, "Measurement of the acoustic target strength of fish in dorsal aspect, including swimbladder resonance," J. Sound Vib. **15**, 397–420 (1971).
[6]R. H. Love, "A comparison of volume scattering strength data with model calculations based on quasisynoptically collected fishery data," J. Acoust. Soc. Am. **94**, 2255–2268 (1993).
[7]R. W. Nero, C. H. Thompson, and J. M. Jech, "*In situ* acoustic estimates of the swimbladder volume of Atlantic herring (*Clupea harengus*)," ICES J. Mar. Sci. **61**, 323–337 (2004).
[8]K. G. Foote, "Importance of the swimbladder in acoustic scattering by fish: A comparison of gadoid and mackerel target strengths," J. Acoust. Soc. Am. **67**, 2084–2089 (1980).
[9]D. V. Holliday, "Resonance structure in echoes from schooled pelagic fish," J. Acoust. Soc. Am. **51**, 1322–1332 (1972).
[10]R. W. Nero, C. H. Thompson, C. Feuillade, and R. H. Love, "A highly reflective low cost backscattering target," IEEE J. Ocean. Eng. **26**, 259–265 (2001).
[11]T. K. Stanton, D. Chu, J. M. Jech, and J. D. Irish, "A broadband echosounder for resonance classification of swimbladder-bearing fish," in Proceedings of the 2007 IEEE Oceans Conference, Aberdeen, Scotland, (2007).
[12]W. J. Overholtz, "The Gulf of Maine-Georges Bank Atlantic herring (*Clupea harengus*): Spatial pattern analysis of the collapse and recovery of a large marine fish complex," Fish. Res. **57**, 237–254 (2002).
[13]W. J. Overholtz and K. D. Friedland, "Recovery of the Gulf of Maine herring (*Clupea harengus*) complex: perspectives based on bottom trawl survey data," Fish. Bull. **100**, 593–608 (2002).
[14]W. J. Overholtz, J. M. Jech, W. L. Michaels, L. D. Jacobson, and P. J. Sullivan, "Empirical comparisons of survey design in acoustic surveys of Gulf of Maine-Georges Bank Atlantic herring," J. Northw. Atl. Fish. Sci. **36**, 127–144 (2006).
[15]N. C. Makris and P. Ratilal, "OAWRS Gulf of Maine 2006 Experiment Cruise Report," sponsored by the Office of Naval Research and Alfred P. Sloan Foundation, 19 September–6 October, 2006.
[16]M. Andrews, T. Chen, and P. Ratilal, "Empirical dependence of acoustic transmission scintillation statistics on bandwidth, frequency, and range on New Jersey continental shelf," J. Acoust. Soc. Am. **125**, 111–124 (2009).
[17]C. I. Malme, "Development of a high target strength passive acoustic reflector low-frequency sonar applications," IEEE J. Ocean. Eng. **19**, 438–448 (1994).
[18]E. T. Küsel and P. Ratilal, "Effects of incident field refraction on scattered field from vertically extended cylindrical targets in range-dependent ocean waveguides," J. Acoust. Soc. Am. **125**, 1930–1936 (2009).
[19]N. C. Makris and J. M. Berkson, "Long-range backscatter from the Mid-Atlantic Ridge," J. Acoust. Soc. Am. **95**, 1865-1881 (1994).
[20]N. C. Makris, L. Z. Avelino, and R. Menis, "Deterministic reverberation from ocean ridges," J. Acoust. Soc. Am. **97**, 3547–3574 (1995).
[21]P. Ratilal, Y. Lai, D. Symonds, L. A. Ruhlmann, J. R. Preston, E. K. Scheer, M. T. Garr, C. W. Holland, J. A. Goff, and N. C. Makris, "Long range acoustic imaging of the continental shelf environment: The Acoustic Clutter Reconnaissance Experiment 2001," J. Acoust. Soc. Am. **117**, 1977–1998 (2005).
[22]T. C. Weber, H. Pena, and J. M. Jech, "Multibeam and single-beam sonar observations of Atlantic herring in the Gulf of Maine," J. Acoust. Soc. Am. **122**, 3003–3004 (2007).
[23]R. C. Dotson and D. A. Griffith, "A high-speed rope trawl for collecting coastal pelagic fishes," California Cooperative Oceanic Fisheries Investigation, Progress Reports **37**, 134–139 (1996).
[24]P. G. Bergmann, "Intensity fluctuations," in *The Physics of Sound in the Sea, Part I: Transmission* (National Defense Research Committee, Washington, DC, 1946).
[25]I. Dyer, "Statistics of sound propagation in the ocean," J. Acoust. Soc. Am. **48**, 337–345 (1970).
[26]N. C. Makris, "The effect of saturated transmission scintillation on ocean acoustic intensity measurements," J. Acoust. Soc. Am. **100**, 769–783 (1996).
[27]M. D. Collins, "A split-step Pade solution for the parabolic equation method," J. Acoust. Soc. Am. **93**, 1736–1742 (1993).
[28]G. Bar-Yehoshua, "Quantifying the effect of dispersion in continental shelf sound propagation," MS thesis, Massachusetts Institute of Technology, Cambridge, MA (2002).
[29]P. Ratilal, Y. Lai, and N. Makris, "Validity of the sonar equation and Babinet's principle for scattering in a stratified medium," J. Acoust. Soc. Am. **112**, 1797–1816 (2002).
[30]M. Andrews, Z. Gong, and P. Ratilal, "High resolution population density imaging of random scatterers with the matched filtered scattered field variance," J. Acoust. Soc. Am. **126**, 1057–1068 (2009).
[31]T. R. Hahn, "Low frequency sound scattering from spherical assemblages of bubbles using effective medium theory," J. Acoust. Soc. Am. **122**, 3252–3267 (2007).
[32]A. Galinde, N. Donabed, M. Andrews, S. Lee, N. C. Makris, and P. Ratilal, "Range-dependent waveguide scattering model calibrated for bottom reverberation in a continental shelf environment," J. Acoust. Soc. Am. **123**, 1270–1281 (2008).
[33]D. Symonds, "Instantaneously imaging and continuously monitoring fish populations over continental-shelf scales using ocean acoustic waveguide remote sensing," Ph.D. thesis, Massachusetts Institute of Technology, Cambridge, MA (2008).
[34]K. G. Foote, "Fish target strengths for use in echo integrator survey," J. Acoust. Soc. Am. **82**, 981–987 (1987).
[35]K. G. Foote, A. Aglen, and O. Nakken, "Measurement of fish target strength with a split-beam echo sounder," J. Acoust. Soc. Am. **80**, 612-621 (1986).
[36]D. N. MacLennan, P. G. Fernandes, and J. Dalen, "A consistent approach to definitions and symbols in fisheries acoustics," ICES J. Mar. Sci. **59**, 365–369 (2002).
[37]E. Ona, "An expanded target-strength relationship for herring," ICES J. Mar. Sci. **60**, 493–499 (2003).
[38]A. P. French, *Vibrations and Waves*, MIT Introductory Physics Series (CRC, Boca Raton, FL, 1971).
[39]F. R. Harden Jones and N. B. Marshall, "The structure and functions of the teleostean swimbladder," Biol. Rev. **28**, 16–83 (1953).
[40]D. N. MacLennan and E. J. Simmonds, *Fisheries Acoustics*, 2nd ed. (Chapman and Hall, London, 1992), pp. 177–182.
[41]F. R. Harden Jones and P. Scholes, "Gas secretion and resorption in the swimbladder of the cod *Gadus morhua*," J. Comp. Physiol. **155**, 319–331 (1985).
[42]O. R. Godø and K. Michalsen, "The use of data storage tags to study cod natural behaviour and availability to abundance surveys in the Barents Sea," International Council for the Exploration of the Sea (1997).
[43]S. M. M. Fässler, P. G. Fernandes, S. I. K. Semple, and A. S. Brierley, "Depth-dependent swimbladder compression in herring *Clupea harengus* observed using magnetic resonance imaging," J. Fish Biol. **74**, 296–303 (2009).
[44]V. M. Brawn, "Physical properties and hydrostatic function of the swimbladder of herring (*Clupea harengus L*)," J. Fish. Res. Board Can. **19**, 635–656 (1962).
[45]J. H. S. Blaxter and R. S. Batty, "The herring swimbladder as a gas reservoir for the acousticolateralis system," J. Mar. Biol. Assoc. U.K. **59**, 1–10 (1979).
[46]J. H. S. Blaxter and J. R. Hunter, "The biology of the Clupeoid fishes," Adv. Mar. Biol. **20**, 1–223 (1982).
[47]J. H. S. Blaxter and R. S. Batty, "The herring swimbladder: Loss and gain of gas," J. Mar. Biol. Assoc. U.K. **64**, 441–459 (1984).
[48]R. E. Thorne and G. L. Thomas, "Acoustic observations of gas-bubble release by Pacific herring *Clupea harengus pallasi*," Can. J. Fish. Aquat. Sci. **47**, 1920–1928 (1990).
[49]G. Sundnes, T. Enns, and P. F. Scholander, "Gas secretion in fishes lacking *rete mirabile*," J. Exp. Biol. **35**, 671–676 (1958).
[50]G. Fahlén, "Morphological aspects on the hydrostatic function of the gas bladder of *Clupea harengus* L.," Acta. Univ. Lund. Sect. II **1**, 1–49 (1967).
[51]R. Fänge, "Gas exchange in fish swim bladder," *Reviews of Physiology Biochemistry and Pharmacology* (Springer-Verlag, Berlin, 1983), Vol. **97**, pp. 111–158.
[52]L. Nøttestad, "Extensive gas bubble release in Norwegian spring-spawning herring (*Clupea harengus*) during predator avoidance," ICES J. Mar. Sci. **55**, 1133–1140 (1998).
[53]G. Fahlén, "Morphology of the gas bladder of *Coregonus laveratus* L.," Acta. Univ. Lund. Sect. II **28**, 1–37 (1967).
[54]G. Sundnes and P. Bratland, "Notes on the gas content and neutral buoyancy in physostome fish," Report on Norwegian Fishery and Marine In-

vestigations **16**, 89–97 (1972).
[55]J. A. Cranston, "Lxi. studies in gas production by bacteria, II. Denitrification and bacterial growth phases," Biochem. J. **24**, 529–548 (1930).
[56]J. S. Link and J. Burnett, "The relationship between stomach contents and maturity state for northwest Atlantic fishes: New paradigms?," J. Fish Biol. **59**, 783–794 (2001).
[57]R. H. Love, "Resonant acoustic scattering by swimbladder-bearing fish," J. Acoust. Soc. Am. **64**, 571–580 (1978).
[58]C. H. Thompson and R. H. Love, "Determination of fish size distributions and aerial densities using broadband, low-frequency measurements," ICES J. Mar. Sci. **53**, 197–201 (1996).
[59]R. W. Nero and M. E. Huster, "Low-frequency acoustic imaging of Pacific salmon on the high seas," Can. J. Fish. Aquat. Sci. **53**, 2513–2523 (1996).
[60]R. W. Nero, C. H. Thompson, and R. H. Love, "Abyssopelagic grenadiers: the probable cause of low-frequency sound scattering at great depths off the Oregon and California coasts," Deep-Sea Res., Part I **44**, 627–645 (1997).
[61]R. W. Nero, C. H. Thompson, and R. H. Love, "Low-frequency acoustic measurements of Pacific hake, merluccicus productus, off the west coast of the United States," Fish. Bull. **96**, 329–343 (1998).
[62]D. Weston, "Sound propagation in the presence of bladder fish," in *Underwater Acoustics*, edited by V. M. Albers (Plenum, New York, 1967), Vol. **2**, pp. 55–88.
[63]Z. Ye, "Low-frequency acoustic scattering by gas-filled prolate spheroids in liquids," J. Acoust. Soc. Am. **101**, 1945–1952 (1997).
[64]C. Feuillade and M. F. Werby, "Resonances of deformed gas bubbles in liquids," J. Acoust. Soc. Am. **96**, 3684–3692 (1994).
[65]E. Ona, "Physiological factors causing natural variations in acoustic target strength of fish," J. Mar. Biol. Assoc. U.K. **70**, 107–127 (1990).
[66]N. Gorska and E. Ona, "Modelling the effect of swimbladder compression on the acoustic backscattering from herring at normal or near-normal dorsal incidences," ICES J. Mar. Sci. **60**, 1381–1391 (2003).
[67]N. C. Makris, P. Ratilal, D. T. Symonds, S. Jagannathan, S. Lee, and R. W. Nero, "Fish population and behavior revealed by instantaneous continental shelf-scale imaging," Materials and methods are available as supporting material on *Science* online (2006).
[68]P. H. Dahl and O. A. Mathisen, "Measurement of fish target strength and associated directivity at high frequencies," J. Acoust. Soc. Am. **73**, 1205–1211 (1983).
[69]A. Leon-Garcia, *Probability and Random Processes for Electrical Engineering*, 2nd ed. (Addison-Wesley Longman, Reading, MA, 1994), p. 288.
[70]H. Bucker, "Sound propagation in a channel with lossy boundaries," J. Acoust. Soc. Am. **48**, 1187–1194 (1970).
[71]S. Gauthier and G. A. Rose, "Target strength of encaged Atlantic redfish (*Sebastes* spp.)," ICES J. Mar. Sci. **58**, 562–568 (2001).
[72]C. Stransky, "Geographic variation of golden redfish (*Sebastes marinus*) and deep-sea redfish (*S. mentella*) in the North Atlantic based on otolith shape analysis," ICES J. Mar. Sci. **62**, 1691–1698 (2005).

Spatial and temporal patterns of Risso's dolphin echolocation in the Southern California Bight

Melissa S. Soldevilla,[a] Sean M. Wiggins, and John A. Hildebrand
Scripps Institution of Oceanography, University of California, San Diego, 9500 Gilman Drive 0205, La Jolla, California 92093

(Received 17 February 2009; revised 28 September 2009; accepted 6 October 2009)

Geographical and temporal trends in echolocation clicking activity can lead to insights into the foraging and migratory behaviors of pelagic dolphins. Using autonomous acoustic recording packages, the geographical, diel, and seasonal patterns of Risso's dolphin (*Grampus griseus*) echolocation click activity are described for six locations in the Southern California Bight between 2005 and 2007. Risso's dolphin echolocation click bouts are identified based on their unique spectral characteristics. Click bouts were identified on 739 of 1959 recording days at all 6 sites, with the majority occurring at nearshore sites. A significant diel pattern is evident in which both hourly occurrences of click bouts and click rates are higher at night than during the day. At all nearshore sites, Risso's dolphin clicks were identified year-round, with the highest daily occurrence at the southern end of Santa Catalina Island. Seasonal and interannual variabilities in occurrence were high across sites with peak occurrence in autumn of most years at most sites. These results suggest that Risso's dolphins forage at night and that the southern end of Santa Catalina Island represents an important habitat for Risso's dolphins throughout the year.
 [DOI: 10.1121/1.3257586]

PACS number(s): 43.30.Sf, 43.80.Ka [WWA] Pages: 124–132

I. INTRODUCTION

Autonomous fixed-sensor, acoustic-based surveys offer a monitoring technique that can be used for surveying dolphin activity and geographical movements over long-periods in remote locations, during adverse weather conditions and during nighttime periods. Risso's dolphins (*Grampus griseus*) produce short duration (40 us) broadband echolocation clicks with peak frequencies around 50 kHz, centroid frequencies between 60 and 90 kHz, and source levels of 202–222 dB re 1 μPa (peak to peak) (Madsen *et al.*, 2004). Recent descriptions of species-specific spectral characteristics of southern Californian Risso's dolphin echolocation clicks indicate the presence of alternating peaks and notches within individual clicks such that spectral peaks occur at 22, 25, 31, and 39 kHz and spectral notches occur at 20, 28, and 36 kHz (Soldevilla *et al.*, 2008). These characteristics allow for the use of autonomous acoustic monitoring to study the spatial and temporal patterns of these dolphins' clicking activity. However, to interpret variability in recorded click activity in terms of behavioral activity and movement patterns, it is important to first understand echolocation click usage. All odontocetes are thought to produce echolocation clicks to investigate their environment for objects, prey, predators, conspecifics, and navigational cues (Au, 1993; Akamatsu *et al.*, 2005). While no previous studies have examined Risso's dolphin click usage patterns, concurrent behavioral and acoustical studies of numerous other odontocetes indicate higher click rates and occurrence during foraging behaviors, moderate click rates during traveling and socializing behaviors, and low click rates and occurrence during resting behaviors (Norris *et al.*, 1994; Barrett-Lennard *et al.*, 1996; Van Parijs and Corkeron, 2001; Nowacek, 2005). Assuming similar patterns for Risso's dolphin behavior and click activity, the variability in recorded click rates and occurrence can be used to differentiate periods of activity and rest. Understanding click usage is also important when studying movement patterns with clicks, as Mackenzie *et al.* (2006) noted that lack of acoustic detections may result from the presence of non-vocal animals and not necessarily the absence of animals. This can be especially important when call output varies seasonally, such as for seasonal reproductive displays (e.g., Van Parijs *et al.*, 1999; Oleson *et al.*, 2007). However, given the necessity for dolphins to feed daily (Smith and Gaskin, 1974; Lockyer, 1981) and the importance of echolocation in foraging, it is unlikely that echolocation clicks could exhibit such a seasonal cycle. Therefore, echolocation clicks are an ideal call type for examining seasonal movement patterns.

Passive acoustic techniques offer a monitoring method that allows researchers to examine variability in calling activity over a range of spatial and temporal scales. The ability to monitor through the night allows analysis of diel trends in dolphin acoustic activity that is not possible with visual monitoring and offers the potential of sampling animals that might not be available to daytime visual surveys. Autonomous acoustic instruments can sample over long time periods to obtain high-resolution sampling of dolphin presence across seasons and years at a given location, which allows analysis of seasonal and interannual trends. Multiple instruments can be deployed to increase spatial coverage and allow comparisons in temporal presence across locations to exam-

[a]Author to whom correspondence should be addressed. Electronic mail: melissa.soldevilla@duke.edu. Present address: 135 Duke Marine Lab Road, Beaufort, NC 28516.

ine long-term movement patterns. This study uses long-term passive acoustic recordings to examine spatial and temporal trends in Risso's dolphin echolocation behavior and movement patterns. Autonomous acoustic recording packages were deployed at six sites throughout the Southern California Bight (SCB) over a period of 2.5 years. To examine diel variability in echolocation activity levels, click rates and hourly presence of Risso's dolphin click bouts are compared between periods of daylight and darkness. To examine seasonal and interannual movement patterns, variability in daily presence is compared among seasons, years, and sites. Diel and seasonal trends in echolocation behavior are described and the implications for foraging and movements are discussed.

II. BACKGROUND

Risso's dolphins are a tropical to temperate water species and in the Eastern North Pacific Ocean, they range from the Gulf of Alaska to the equator (Leatherwood *et al.*, 1980). Risso's dolphins off California, Oregon, and Washington comprise a single population (Carretta *et al.*, 2004). While their range extends beyond the waters of these states, most population studies focus within this region. Abundance estimates indicate approximately 11 900 dolphins throughout the region with about 3400 dolphins found off Southern California (Barlow and Forney, 2007). Risso's dolphins feed nearly exclusively on cephalopods, especially squid species (Clarke and Pascoe, 1985; Clarke, 1996), and studies off California indicate that jumbo squid (*Dosidicus gigas*) and market squid (*Loligo opalescens*) are important prey items (Orr, 1966; Kruse, 1989).

While their importance to the SCB ecosystem is recognized, little is known about Risso's dolphin diel activity patterns and descriptions of their seasonal and interannual movement patterns based on ship-board visual surveys are complex and sometimes conflicting. Diel variability in behavioral activity levels of Risso's dolphins cannot be studied completely using visual survey methods because the animals cannot be seen at night. Studies of diurnal behavioral activity off California indicate variable behavioral states during the day (Kruse, 1989) but suggest foraging/feeding at night (Shane, 1995). Seasonal and interannual variabilities have received more attention from visual survey studies. Since the late 19th century, records indicate high decadal variability in Risso's dolphin occurrence off California (Norris and Prescott, 1961; Leatherwood *et al.*, 1980; Dohl *et al.*, 1981, 1983; Shane, 1995; Kruse *et al.*, 1999), with some researchers indicating correlations between occurrence and extended periods of warm water (Leatherwood *et al.*, 1980) while others find no relationship between abundance and water temperature across years (Barlow and Forney, 2007). Studies of seasonal movements based on visual surveys indicate (1) movements between Oregon/Washington in spring/summer and California in autumn/winter (Green *et al.*, 1992; Forney and Barlow, 1998), (2) high interannual variability in seasonal patterns off central and northern California (Dohl *et al.*, 1983), and (3) year-round residency within the SCB that include inshore/northward or offshore/southward movements in response to warm and cold waters, respectively (Dohl *et al.*, 1981). These results may not be mutually exclusive; discrepancies may reflect differing study areas or high interannual variability. A long-term monitoring program has the potential to clarify the temporal and spatial patterns of Risso's dolphin activity and movements.

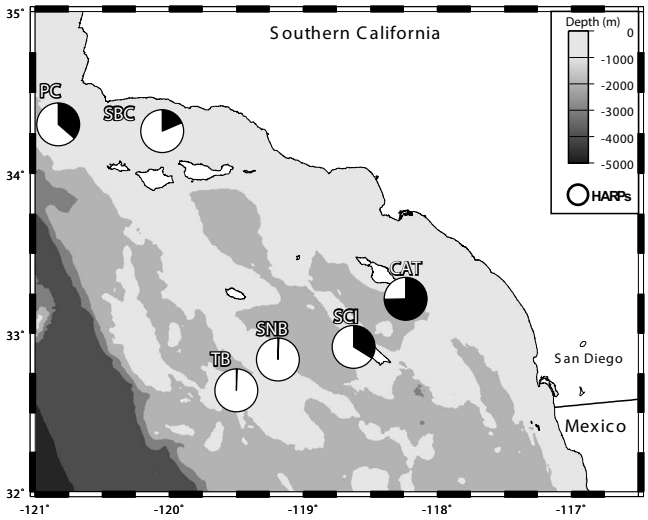


FIG. 1. Map of study area including locations of HARP deployments. Depths are indicated by colorbar. Sites: PC=Point Conception, SBC=Santa Barbara Channel, TB=Tanner Basin, SNB=San Nicholas Basin, SCI=San Clemente Island, and CAT=Santa Catalina Island. HARP locations are presented as pie charts in which the black area indicates the percentage of recording days with Risso's dolphin click bouts present. Risso's dolphins were detected most often at CAT and least often at TB and SNB.

III. METHODS

A. Instrumentation and data collection

High-frequency acoustic recording packages (HARPs) were deployed at six locations throughout the SCB between August 2005 and December 2007 at depths ranging between 300 and 1330 m (Figs. 1 and 2). A brief description of these autonomous seafloor-mounted acoustic recorders is provided here for clarity; see Wiggins and Hildebrand, 2007 for a detailed description of HARP design and capabilities. The HARP data-logging system includes a 16-bit analog to digital converter, up to 1.9 Tbytes of storage capacity, a hydrophone suspended 10 m above the seafloor, a release system, ballast weights, and flotation. The dataloggers are capable of sampling up to 200 kHz and can be set to record continuously or on a sampling schedule to accommodate variable deployment durations. This study includes data from 30 HARP deployments each lasting from 1 to 4 month durations. Temporal coverage at each of the six sites is variable due to research vessel availability and occasional instrument problems (Table I). Data from all deployments included in this study were sampled at 200 kHz, resulting in a recording bandwidth of 1 Hz–100 kHz. A variety of sampling schedules were used across deployments with 2/3 of deployments recorded on a continuous sampling schedule and the remaining 1/3 recorded on a sampling schedule of 5 min on followed by 5, 10, or 15 min off for a given deployment.

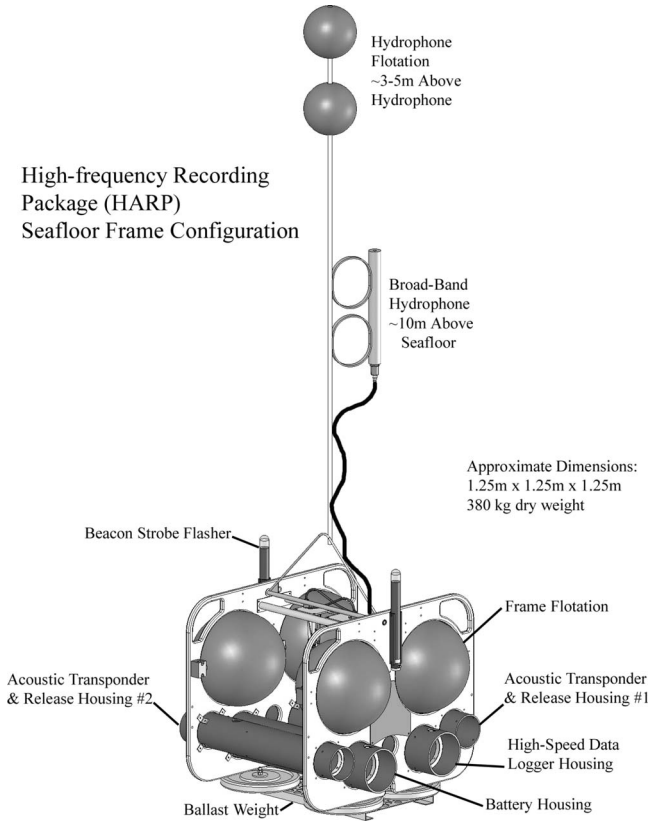


FIG. 2. Schematic representation of autonomous seafloor recording package.

B. Acoustical analysis

Acoustic data were analyzed with a custom software program, TRITON, developed in MATLAB (The MathWorks, Inc., Natick, MA). Raw acoustic HARP data were converted to XWAV format, a format similar to WAV format that incorporates instrument meta-data in an expanded header file, including recording start and stop times. This timing information is crucial when analyzing non-continuous data. Each HARP deployment resulted in 1.6–1.9 Tbytes of data, which is impractical to analyze manually in original form. Therefore, these data were compressed for visual overview by creating long-term spectral averages (LTSAs) (Wiggins and Hildebrand, 2007) from the XWAV files. LTSAs are effectively compressed spectrograms created using the Welch algorithm (Welch, 1967) by coherently averaging 500 spectra created from 2000-point, 0%-overlapped, Hann-windowed data and displaying these averaged spectra sequentially over time. The resulting LTSAs had resolutions of 100 Hz and 5 s in frequency and time, respectively.

Using LTSAs with 100 Hz and 5 s resolution, delphinid whistling and echolocation clicking bouts, rain bouts, ship passings, and other acoustic phenomenon can easily be distinguished from background noise (e.g., Wiggins and Hildebrand, 2007). Soldevilla *et al.* (2008) described distinct spectral banding patterns found in individual echolocation clicks of Risso's dolphins recorded during concurrent visual and acoustic ship-based surveys. Risso's dolphin clicks exhibit spectral peaks at 22, 25, 31, and 39 kHz with spectral notches at 20, 28, and 36 kHz. These spectral banding patterns are also found in many echolocation click bouts in autonomously recorded HARP data and the banding patterns are particularly striking as visualized in LTSAs (Fig. 3). By visually examining 30 min long LTSA segments, the start and end times of click bouts exhibiting the described spectral patterns were located and logged. Numerous delphinid species occur offshore southern California. Three of the most common species produce clicks that do not have consistent spectral banding patterns, while Pacific white-sided dolphins produce two click types with spectral banding at different frequencies than Risso's dolphins (Soldevilla *et al.*, 2008). Occasionally, click bouts without spectral banding patterns or with Pacific white-sided dolphin patterns overlapped Risso's dolphin click bouts. These bouts were labeled as Risso's dolphin mixed species groups. Click bout start and end times were used to calculate daily occurrence, hourly occurrence, and click rates for temporal analyses. Mixed species data were only included in temporal analyses of Risso's dolphin click occurrence and were removed from the remaining click rate analyses described in Sec. III C.

C. Click detection

To calculate click rates for analysis of diel patterns, an automatic call detection algorithm was developed. This algo-

TABLE I. Seasonal recording effort at each site across 3 years of study. Cells represent the number of week-long replicate samples available for statistical analysis at each site for each season of each year. Site abbreviations: PC=Point Conception, SBC=Santa Barbara Channel, TB=Tanner Basin, SNB=San Nicholas Basin, SCI=San Clemente Island, and CAT=Santa Catalina Island. TB and SNB are not included in the total samples for the seasonal analysis in this study because they were deployed during only part of the year in 2007 and did not sample across all seasons. Months included in each season are indicated by the first letter of the month.

	Winter (J F M)			Spring (A M J)			Summer (J A S)			Autumn (O N D)			
Site	2005	2006	2007	2005	2006	2007	2005	2006	2007	2005	2006	2007	Total
PC	0	3	10	0	0	13	7	5	13	13	4	3	71
SBC	0	4	8	0	0	3	7	0	12	13	7	5	59
TB	0	0	0	0	0	4	0	4	13	0	3	3	27
SNB	0	0	0	0	0	0	0	0	7	0	0	7	14
SCI	0	0	11	0	0	12	0	0	9	0	0	8	40
CAT	0	4	8	0	0	9	7	12	10	13	12	10	85
Total samples for seasonal analysis	0	11	37	0	0	37	21	17	44	39	23	26	255
		48			37			82			88		255

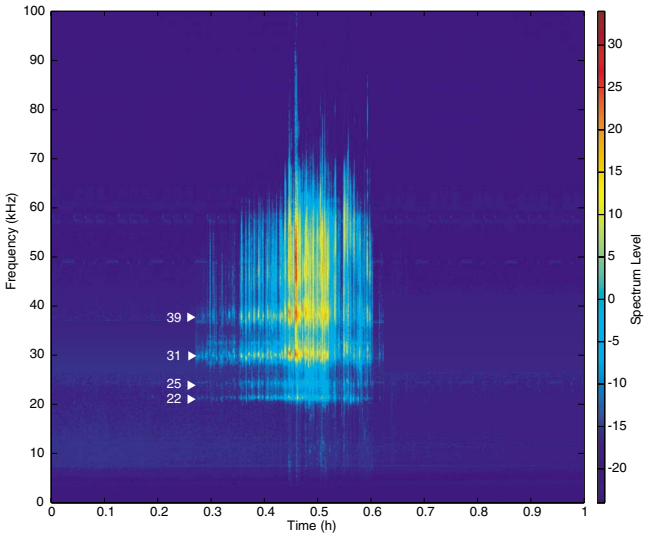


FIG. 3. (Color online) Example LTSA illustrating echolocation click bout containing the unique spectral peak and notch structure described for Risso's dolphins by Soldevilla *et al.* (2008). LTSA spectrogram has resolutions of 100 Hz and 5 s in frequency and time, respectively, and represents coherent averages of 500 spectra created using 2000-point, 0%-overlapped, Hann-windowed HARP data. Spectral peaks occur in the Risso's dolphin clicks at approximately 22, 25, 31, and 39 kHz.

rithm simultaneously detected broadband clicks and whistles in the spectral domain, although only the click results are presented here. To obtain the best results for both whistles and clicks, spectra were calculated using a 1024-point fast Fourier transform (FFT) with 50% overlap and a Hann window. Spectral-mean-subtraction was performed on each spectrum by subtracting the mean spectral vector of the surrounding 3 s of data. Individual spectra were selected as click candidates if a minimum percentage of frequency bins exceeded a minimum threshold within the bandwidth of interest. Values for minimum percentage, threshold, and bandwidth were set as 12.5%, 11 dB, and 15–95 kHz, respectively. These values were empirically determined to detect most high quality echolocation clicks while minimizing detection of other impulsive sounds, with a trade-off of missing lower amplitude echolocation clicks. For each click candidate, start and end times of 15 ms of data surrounding the click were extracted and overlapping segments were merged.

While the click detector can automatically detect clicks, it cannot automatically classify them to species. Therefore, individual click classifications were determined from the LTSA click bout classifications described in Sec. III B. All detected clicks, which occurred within start and end times of visually classified LTSA Risso's dolphin click bouts, were classified as Risso's dolphin clicks. Detected clicks that occurred in mixed species click bouts were not included in the diel click rate portion of the analysis. Mixed species groups accounted for 35% of the hourly Risso's dolphin click bout occurrences. No diel trend in occurrence was apparent for the mixed species groups removed from the diel click rate analysis.

Several potential biases of this click detection algorithm are worth considering. As noted above, if multiple species were calling during a Risso's dolphin click bout, and it was apparent in the LTSAs, these data were classified as mixed species and were not included in the analysis, biasing the counts to be low. However, if multiple species were present and clicking, but the click bout was classified as only Risso's dolphins, this would cause the click counts to be biased high for Risso's dolphins. This bias was minimized by making conservative decisions about species classification. Additional factors that could bias the click counts low include the following: (1) Many low amplitude clicks were present that did not exceed the thresholds, (2) during periods of intense clicking (multiple high amplitude overlapping click trains), the mean spectral intensity of the surrounding 3 s of data used for spectral-mean-subtraction would be relatively high resulting in some high amplitude clicks being missed by the detector, and (3) during periods with rapid click trains or click trains from multiple individuals in which the apparent interclick interval was greater than 15 ms, only one click was chosen per 15 ms. Overall, these biases would result in lower click rates during intense periods so that comparisons of diel click rate patterns are conservative.

D. Temporal analysis

Both click bout occurrence and click rate data were examined for diel patterns. At each site, daily sunrise, sunset, and twilight data were obtained from the U.S. Naval Observatory (2008) website to establish light and dark periods. Day and night were defined as the periods between nautical twilight, when the sun altitude was at $-12°$ from the horizon. Each click bout was assigned to either day or night. Click bouts that spanned multiple periods were segmented and each segment was assigned the appropriate dark or light period.

To examine diel variation in hourly click bout occurrence, ones and zeros were assigned to 1 h interval bins indicating presence or absence, respectively. Seasonal variation in duration of day and night periods was normalized by dividing the number of 1 h bins with clicks present per period per day by the total number of 1 h bins per period per day. Bins that crossed boundaries between periods were assigned to the period that contained the greater portion of the hour. Analysis of variance (ANOVA) was used to test whether variability in percentage of hours with clicks was significantly different between day and night periods (Zar, 1999).

To examine diel variation in Risso's dolphin click rates, click rates were calculated as the number of clicks detected divided by the total recording duration in minutes for each light and dark period of each day. To account for shorter recording durations caused by the sampling schedule in non-continuous data, total recording duration was calculated from recording start times and durations stored in the XWAV header files. Variability between days was accounted for by calculating a mean-adjusted calling rate in which the daily click rate was subtracted from the click rate for each light and dark period for each day. ANOVA was used to test whether variability in click rates was significantly different between day and night periods (Zar, 1999).

To examine seasonal variation in click bout occurrence, the metric total number of days with click detections present

TABLE II. Summary of number of days and hours with recording effort, number of days and hours with Risso's dolphin click bouts present, percentage of days and hours with Risso's click bouts present, and mean instrument depth for each of the six HARP sites. Site abbreviations: PC=Point Conception, SBC=Santa Barbara Channel, TB=Tanner Basin, SNB=San Nicholas Basin, SCI=San Clemente Island, and CAT=Santa Catalina Island.

	PC	SBC	TB	SNB	SCI	CAT	Total
No. of recording days	457	377	199	110	266	550	1 959
Days with Risso's click bouts	166	70	1	1	90	411	739
Days with Risso's click bouts (%)	36	19	1	1	34	75	38
No. of recording hours	10 499	8753	4647	2469	6277	12 862	45 507
Hours with Risso's click bouts	561	173	1	2	252	2 117	3 106
Hours with Risso's click bouts (%)	5	2	0	0	4	17	7
Instrument depth (m)	787	585	1013	1316	435	351	

per week was calculated and compared across seasons, sites, and years. Seasons were defined by quarters of the year and data were included only from sites with complete seasonal coverage. The two northern nearshore sites (Point Conception and Santa Barbara Channel) and the two southern nearshore sites (Santa Catalina Island and San Clemente Island) met the criteria for inclusion as they sampled all four seasons while data from the two southern offshore sites (Tanner Basin and San Nicolas Basin) did not (Table I). The metric total number of days with click detections present per week was chosen to minimize bias caused by differences in sampling effort across sites and seasons. Presence at 1-day resolution was chosen to minimize the bias due to sampling schedule as the mean probabilities of detecting Risso's dolphin click presence on a given day are 99%, 98%, and 97% for sampling schedules of 5 min on and 5, 10, and 15 min off, respectively (Soldevilla, 2008). Weeks were chosen for the sample duration to provide a reasonable number of replicates per season, site, and year, and because 83% of 255 week-long samples contained 7 complete days of recording effort. The remaining samples were normalized by the number of recording days to account for reduced effort. While this metric ensured that effort was comparable across samples, the number of replicate samples per season, site, and year varied (Table I), requiring the use of a statistical test that accounts for the unbalanced design. Therefore, the null hypotheses of equal means across seasons, years, and sites were tested using the generalized linear model (GLM) ANOVA function in SPSS 11.5 (SPSS Inc., Chicago, IL) with three-way full factorial design and type IV sum of squares to account for the unbalanced design. An ANOVA can only test if all means are equal or not; Tamhane's T2 *post hoc* test was used to determine which seasons, sites, or years were different (Zar, 1999; Garson, 2008).

IV. RESULTS

About 2000 recording days including over 45 500 h of data were analyzed from instruments at the six locations in the SCB, with the majority of sampling effort at the Santa Catalina Island, Santa Barbara Channel, Point Conception, and San Clemente Island sites (Table II). Risso's dolphin click bouts were identified at all six locations on a total of 739 recording days (38% of all recording days) and 3106 recording hours (7% of all recording hours). They were identified most often at the nearshore sites, in particular, at Santa Catalina Island where click bouts were identified on 75% of days and 17% of hours (Fig. 1, Table II). At the remaining three nearshore sites, Point Conception, Santa Barbara Channel, and San Clemente Island, Risso's dolphin click bouts were identified on 19%–36% of days and 2%–5% of hours. Risso's dolphin click bouts were rarely identified at the two southern offshore sites at Tanner Basin and San Clemente Basin (<1% of days and hours) (Fig. 1, Table II).

A distinct diel pattern was evident in the occurrence of Risso's dolphin echolocation click bouts across hours of the day. Across all sites, there was a sharp increase in click detections during the early part of the night, with a slight decrease in the middle of the night followed by another increase before sunrise (Fig. 4). This was followed by a sharp decrease after sunrise, though an additional moderate peak was evident in the late morning. Click bouts were detected least often in the afternoon (Fig. 4). Comparisons of the variability in click bout occurrence and click rate between day and night reveal that click bouts occurred significantly more often and that click rates are significantly higher at night than during the day (ANOVA: $F_{1,1472}=236$, $P<0.0005$, and $F_{1,1164}=93$, $P<0.0005$, respectively) (Fig. 5).

Risso's dolphin click bouts were identified throughout all four seasons at the four sites included in the seasonal analysis (Fig. 6). Variability in occurrence per week is high

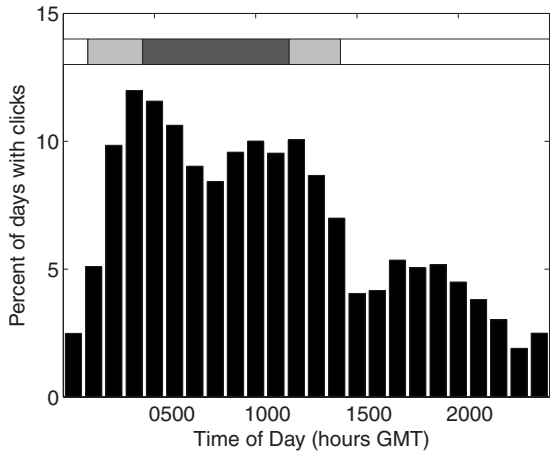


FIG. 4. Diel pattern of Risso's dolphin echolocation click bouts combined across the six HARP locations. Vertical bars represent the percentage of days that have clicks present in each 1 h time bin. Horizontal bar indicates periods of light (white) or darkness (dark gray). Light gray areas represent periods that may be light or dark depending on time of year.

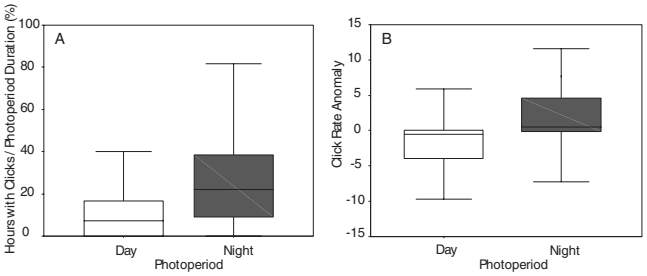


FIG. 5. Variation in Risso's dolphin click bout occurrence (a) and daily mean-adjusted click rate (b) between day and night periods. Central lines represent median value, boxes contain 25th to 75th percentiles, and whiskers contain 5th to 95th percentiles of data. Click bout occurrence and daily click rate anomaly are both significantly higher during the night than during the day.

between seasons, sites, and years. ANOVA results indicate that the main factors of season and site, as well as the interaction effects of season*site, season*year, and season*site*year, were significant sources of variation in mean occurrence of Risso's dolphin click bouts (Table III). Tamhane's T2 *post hoc* test demonstrates that mean occurrence was significantly higher in autumn than in winter, and that Risso's click bouts occurred significantly more often at Santa Catalina Island and significantly less often in the Santa Barbara Channel than at the other three sites, respectively. At

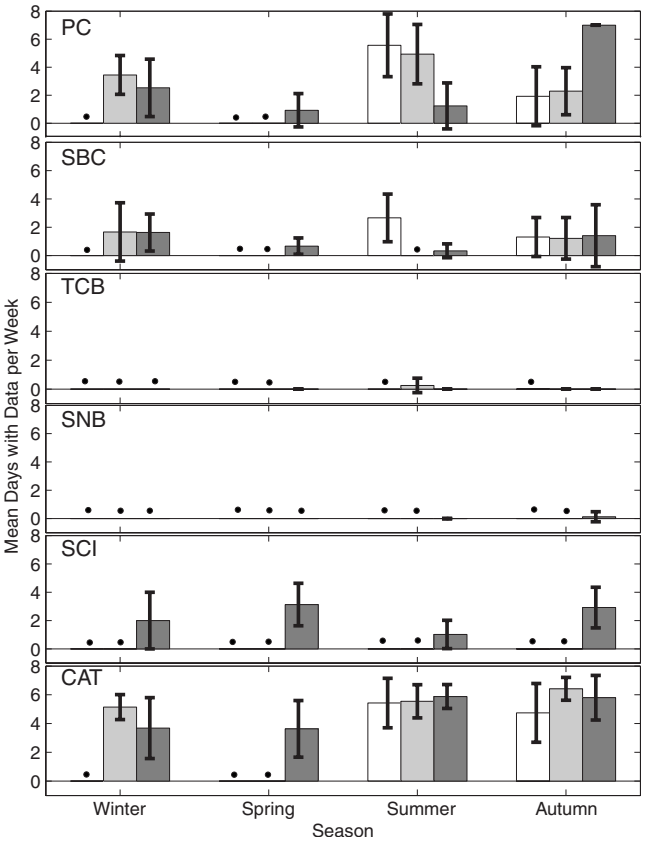


FIG. 6. Seasonal and annual variation in mean days per week with Risso's dolphin click bouts at the six HARP sites. Bar colors indicate year: white =2005, light gray=2006, and dark gray=2007. Error bars indicate standard deviation. Black dots indicate no data available in that season and year. Sites: PC=Point Conception, SBC=Santa Barbara Channel, TB=Tanner Basin, SNB=San Nicholas Basin, SCI=San Clemente Island, and CAT =Santa Catalina Island.

Santa Catalina Island, click bout occurrence was higher during summer and autumn in years 2006 and 2007 but decreased during autumn in 2005. At San Clemente Island, data were only available for 2007, in which there were peaks in click bout occurrence in spring and autumn. At Point Conception and in the Santa Barbara Channel, click bout occurrence was higher in winter and autumn during 2007 while occurrence was higher summer than autumn during 2005.

V. DISCUSSION

Recent discoveries that enable Risso's dolphin echolocation clicks to be identified to species have opened the door to using passive acoustic monitoring to examine temporal and spatial trends in Risso's dolphin echolocation click bouts. The ability of acoustic techniques to survey through the night has allowed the examination of diel patterns of click activity. Statistical comparisons of both click bout occurrence and click rates between day and night reveal a diel pattern in which both click activity metrics are higher at night than during the day. On a finer scale, hourly variability in Risso's dolphin click bout occurrence indicates high levels through the night, moderate levels during the morning, and low levels during the afternoon. In light of relationships found in prior studies of acoustic output and behavioral state (Norris *et al.*, 1994; Barrett-Lennard *et al.*, 1996; Van Parijs and Corkeron, 2001; Nowacek, 2005), the authors hypothesize that Risso's dolphins are generally foraging at night, traveling and/or socializing during the morning, and resting during the afternoon. This diel vocal and behavioral activity pattern is similar to that found for Hawaiian spinner dolphins at Kealakekua Bay, which echolocate at high rates while foraging at night, exhibit low vocal activity while resting during the morning, and exhibit moderate vocal activity while socializing in the afternoon (Norris *et al.*, 1994). Foraging at night is similar between the species and is consistent with Shane's (1995) hypothesis based on daytime visual surveys. However, these results indicate inverse behavior patterns between Risso's and Hawaiian spinner dolphins for morning and afternoon. A comparative visual and acoustic study of Risso's dolphin behavior could test these hypotheses about relationships between clicking and behavioral activities.

Additional support for the hypothesis that Risso's dolphins forage at night can be found in similarities between diel click occurrence and squid dive patterns. Nocturnal click bout occurrence is high throughout the night, in particular, during the early and late parts of the night with a dip in occurrence in the middle of the night (Fig. 4). Dive profiles of jumbo squid, a known Risso's dolphin prey item off California (Orr, 1966), indicate that they are found in deep waters during the daytime, rise to shallow waters at night, where they are readily available to Risso's dolphins, and often return to depth for a short-period in the middle of the night (Gilly, 2006). While detailed dive patterns of other squid prey species have not been described, many squid species found in the SCB are known to follow diel vertical migrations (Roper and Young, 1975) thereby increasing their availability to Risso's dolphins at night. Spinner and dusky dolphins have both been shown to follow the vertical move-

TABLE III. Results of three-way ANOVA for seasonal, annual, and site effects on Risso's dolphin occurrence. The main effects of season and site, and the interaction effects of season*site, season*year, and season*site*year all significantly influenced the occurrence of Risso's dolphin click bouts.

Source	Type IV sum of squares	*df*	Mean square	*F*	Sig.	
Corrected model	955.6	29	33.0	13.0	0.000	
Intercept	2083.1	1	2083.1	823.2	0.000	
Season	69.1	3	23.0	9.1	0.000	*
Site	218.5	3	72.8	28.8	0.000	*
Year	1.0	2	0.5	0.2	0.822	
Season*site	126.5	9	14.1	5.6	0.000	*
Season*year	106.2	3	35.4	14.0	0.000	*
Site*year	20.5	4	5.1	2.0	0.092	
Season*site*year	100.2	5	20.0	7.9	0.000	*
Error	569.32	225	2.5			
Total	3996.7	255				
Corrected total	1525.0	254				

ments of their diel-vertically migrating prey while in near-surface waters (Benoit-Bird and Au, 2003; Benoit-Bird *et al.*, 2004). Results for Risso's dolphins suggest a similar pattern in which Risso's dolphins are actively feeding and echolocating when squid are within a preferred shallow depth range.

Diel patterns of vocal activity have been described for several other odontocete species. A study of harbor porpoise click rates off Scotland revealed greater echolocation activity at dawn and night (Carlstrom, 2005), while off Ireland, bottlenose dolphin clicks exhibited no distinct diel pattern (Philpott *et al.*, 2007). Goold (2000) found that common dolphins (*Delphinus delphis*) off New Zealand produced significantly more whistle and click vocalizations during dusk and night, which he suggested may be related to greater foraging activity as common dolphins forage on diel-vertically migrating prey.

Geographically, Risso's dolphin click bouts were mostly confined to the four nearest-shore sites with rare detections occurring at the two offshore sites. This distribution is similar to that found by Forney and Barlow (1998) during visual line-transect surveys off California. Okutani and McGowan (1969) provided distributions for paralarvae of a variety of squid species found in the SCB. Market squid, a known prey item of Risso's dolphins off Monterey (Kruse, 1989), is the only squid with neritic spawning grounds and these grounds closely match the distribution of Risso's dolphin detections. Stomach content analyses off California are rare (e.g., Orr, 1966) so preferred prey in this area is unknown. These distribution results suggest that market squid is likely a preferred prey item. However, this study only includes six sites so it is possible that Risso's dolphins inhabit other regions outside of market squid spawning grounds. Additionally, sampling at the two offshore sites did not cover the entire year, so it is possible that they inhabit offshore regions during winter or spring.

While Risso's dolphin click bouts were identified throughout the year at all inshore sites, the temporal changes in Risso's dolphin distribution for the SCB show high seasonal and interannual variabilities at all sites. Acoustic occurrence was generally higher during autumn across all four inshore sites, particularly during 2007, although the peak occurred during summer in some years at the northern sites. Dohl *et al.* (1981) reported similar results from visual surveys of the entire SCB in which dolphins are present year-round with a peak in sightings occurring during September. Based on ship-board and aerial visual survey results, Green *et al.* (1992) and Forney and Barlow (1998) indicated that dolphins move from Oregon/Washington during spring/summer to California in autumn/winter. In this study, acoustic occurrence remained high during spring 2007 at San Clemente Island. Several possible explanations for this high occurrence include the following: The year 2007 was an anomalously cold year to the north of the SCB, only a portion of the population moves out of the region during spring and summer, or a distinct resident population remains in the region year-round. Risso's dolphin click bouts were identified on 75% of days at the Santa Catalina Island site. This is clearly an important habitat for these animals, but it remains unknown whether this represents a habitat of a resident population or if it is just a common feeding ground for animals moving over a larger area. Photo-identification studies, such as those conducted by Shane (1994) and Kruse (1989), or use of satellite tags could help answer this question.

A. Limitations

When considering the diel behavior of highly mobile species such as dolphins, it is unclear whether a lack of acoustic detections represents absence of the dolphins from the study site or presence of non-vocalizing animals. Dolphins are capable of moving into and out of a study site within 1 day, as shown by visual surveys of spinner dolphins off Hawaii that exhibit diel movements between inshore resting areas during the day to offshore feeding zones at night (Norris *et al.*, 1994; Benoit-Bird and Au, 2003). Dolphins may also produce more than one echolocation click type, as has been found in Pacific white-sided dolphins (Soldevilla *et al.*, 2008). Therefore, the diel variation in occurrence of Risso's dolphin clicks could either represent movements out of the area during the day and movements into the area during the night, or it could represent variability in echolocation

activity as a function of varying behavior state. Since Risso's dolphins exhibited greater echolocation activity at night at all sites it seems likely that this variation represents a changing behavioral state, for example, increased foraging, particularly when one considers the correspondence of increased click activity with squid diving behavior. Risso's dolphins have only been recorded producing one click type off southern California, but the sample size is small (Soldevilla *et al.*, 2008). Whether they produce other click types and, if so, whether click type production varies with behavior state deserve further study. The development of compact acoustic tags, similar to those used on larger whales, may provide more definitive answers to these questions.

Factors that may affect acoustic detection area and therefore the probability of detecting calling animals include directionality of the sound source, distance from hydrophone, sound propagation conditions, and acoustic masking. Echolocation clicks are highly directional (Au, 1993). Given a highly directional sound source combined with distance from the hydrophone, variability between light periods or sites could be caused by variability in probability of detection due to diel variability in dive depth of the animals or site variability in the depth of the HARPs if the dolphins' beams are rarely directed downward. Studies that examine how click detectability varies with depth on a vertical hydrophone array could answer this question. Sound propagation conditions may vary across sites and seasons. However, at the frequencies used for echolocation, attenuation severely limits detection range, and therefore the potential for significant variation owing to seasonal variations is reduced. Vessels, sonars, other animals, rain, wind, and waves may produce sounds that mask echolocation clicks so it is important to consider seasonality in these sources. Preliminary analyses indicate that vessel noise is consistent throughout the year at all sites except Santa Catalina Island where there is an increase in occurrence during the summer (May–September), while wind and rain noise within the frequency band of clicks occur minimally throughout the year. A more thorough investigation of seasonal variation related to these noise sources is the topic of another study.

VI. CONCLUSIONS

Temporal and spatial patterns are evident for Risso's dolphin click bout occurrence and click rates in the SCB. Diel patterns exist in which Risso's dolphins click more frequently and at higher rates during the night than day at all sites. These diel patterns provide insight into the behavioral ecology of these animals and add support to suggestions that Risso's dolphins are nighttime foragers. This type of data is important when designing passive acoustic monitoring surveys. Oftentimes, data sampling schedules are necessary so it is imperative that survey designs should either sample different periods of the day equally or at least sample at the same time each day. Seasonal and interannual variations among sites in the SCB were high as has been found during studies using visual survey methods. The year-round occurrence of Risso's dolphins in the SCB and high occurrence rates at Santa Catalina Island suggest the possibility of a resident population in addition to a population that moves seasonally between California and Oregon/Washington waters. This should be investigated further with photo-identification and satellite-tagging techniques. While the time series described in this study is not long enough to examine interannual and interdecadal changes such as those described in previous studies (e.g., Shane, 1995; Kruse *et al.*, 1999), this study is on-going and has the potential to answer such questions in the future especially when coupled with detailed environmental observations.

ACKNOWLEDGMENTS

The authors thank all the personnel who were involved in instrument development, deployment, recovery, and data acquisition including: Simone Baumann, Hannah Basset, Catherine Berchock, Greg Campbell, George Crocker, Chris Garsha, Martin Gassmann, Liz Henderson, Brent Hurley, Megan McKenna, Lisa Munger, Erin Oleson, Ethan Roth, Nadia Rubio, and Lynne Williams. Jay Barlow, Simone Baumann, Dave Checkley, Bruce Cornuelle, Liz Henderson, Lisa Munger, Erin Oleson, Bhaskar Rao, and Marie Roch provided insightful discussions and/or constructive critique of earlier versions of this manuscript. The authors thank the ARCS Foundation, Los Angeles chapter for supporting M.S.S., and Frank Stone and Ernie Young for supporting this work through The Chief of Naval Operations, division N45.

Akamatsu, T., Wang, D., Wang, K. X., and Naito, Y. (**2005**). "Biosonar behaviour of free-ranging porpoises," Proc. R. Soc. London, Ser. B **272**, 797–801.

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

Barlow, J., and Forney, K. A. (**2007**). "Abundance and population density of cetaceans in the California Current ecosystem," Fish. Bull. **105**, 509–526.

Barrett-Lennard, L. G., Ford, J. K. B., and Heise, K. A. (**1996**). "The mixed blessing of echolocation: Differences in sonar use by fish-eating and mammal-eating killer whales," Anim. Behav. **51**, 553–565.

Benoit-Bird, K. J., and Au, W. W. L. (**2003**). "Prey dynamics affect foraging by a pelagic predator (*Stenella longirostris*) over a range of spatial and temporal scales," Behav. Ecol. Sociobiol. **53**, 364–373.

Benoit-Bird, K. J., Wursig, B., and McFadden, C. J. (**2004**). "Dusky dolphin (*Lagenorhynchus obscurus*) foraging in two different habitats: Active acoustic detection of dolphins and their prey," Marine Mammal Sci. **20**, 215–231.

Carlstrom, J. (**2005**). "Diel variation in echolocation behavior of wild harbor porpoises," Marine Mammal Sci. **21**, 1–12.

Carretta, J. V., Forney, K. A., Muto, M. M., Barlow, J., Baker, J., and Lowry, M. (**2004**). "US Pacific Marine Mammal Stock Assessments: 2003", National Marine Fisheries Services, Southwest Fisheries Science Center, La Jolla, CA.

Clarke, M. R. (**1996**). "Cephalopods as prey. III. Cetaceans," Philos. Trans. R. Soc. London, Ser. B **351**, 1053–1065.

Clarke, M. R., and Pascoe, P. L. (**1985**). "The stomach contents of a Risso's dolphin (*Grampus griseus*) stranded at Thurlestone, South Devon," J. Mar. Biol. Assoc. U.K. **65**, 663–665.

Dohl, T. P., Guess, R. C., Duman, M. L., and Helm, R. C. (**1983**). "Cetaceans of central and northern California, 1980–1983: Status, abundance, and distribution," Pacific OCS Region Minerals Management Service, US Department of the Interior Contract No. 14-12-0001-29090, Santa Cruz, CA.

Dohl, T. P., Norris, K. S., Guess, R. C., Bryant, J. D., and Honig, M. W. (**1981**). "Summary of marine mammal and seabird surveys of the Southern California area, 1975–1978 part two. Cetacea of the Southern California Bight," in *Final Report to the Bureau of Land Management*, NTIS Report, Washington, DC, p. 4.

Forney, K. A., and Barlow, J. (**1998**). "Seasonal patterns in the abundance and distribution of California cetaceans, 1991–1992," Marine Mammal Sci. **14**, 460–489.

Garson, G. D. (**2008**). "Univariate GLM, ANOVA, and ANCOVA," in *Statnotes: Topics in Multivariate Analysis*, http:..www2.chass.ncsu.edu/garson/pa765/statnote.htm (Last viewed 5/30/2008).
Gilly, W. F. (**2006**). "Vertical and horizontal migrations by the jumbo squid *Dosidicus gigas* revealed by electronic tagging," Mar. Ecol.: Prog. Ser. **324**, 1–17.
Goold, J. C. (**2000**). "A diel pattern in vocal activity of short-beaked common dolphins, *Delphinus delphis*," Marine Mammal Sci. **16**, 240–244.
Green, G., Brueggeman, J. J., Grotefendt, R. A., Bowlby, C. E., Bonnell, M. L., and Balcomb, K. C. I. (**1992**). "Cetacean distribution and abundance off Oregon and Washington, Ch. 1," in *Oregon and Washington Marine Mammal and Seabird Surveys*, OCS Study 91-0093, Final Report prepared for Pacific OCS Region, Minerals 120 Management Service (U.S. Department of the Interior, Los Angeles, CA).
Kruse, S., Caldwell, D. K., and Caldwell, M. C. (**1999**). "Risso's dolphin *Grampus griseus* (G. Cuvier, 1812)," in *Handbook of Marine Mammals*, edited by S. H. Ridgway and R. Harrison (Academic, Cambridge), pp. 183–212.
Kruse, S. L. (**1989**). "Aspects of the biology, ecology, and behavior of Risso's dolphins (*Grampus griseus*) off the California Coast," Masters thesis, University of California, Santa Cruz, CA.
Leatherwood, S., Perrin, W. F., Kirby, V. L., Hubbs, C. L., and Dahlheim, M. (**1980**). "Distribution and movements of Risso's dolphin, *Grampus griseus*, in the eastern North Pacific," Fish. Bull. **77**, 951–963.
Lockyer, C. (**1981**). "Estimates of growth and energy budget for the sperm whale, Physeter catodon," in *Mammals in the Seas*, edited by J. Gordon Clark (Food & Agriculture Organization of the UN (FAO), Rome, Italy), pp. 489–504.
Mackenzie, D. I., Nichols, J. D., Royle, J. A., Pollock, K. H., Bailey, L. L., and Hines, J. E. (**2006**). *Occupancy Estimation and Modeling: Inferring Patterns and Dynamics of Species Occurrence* (Academic, San Diego, CA).
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. Exp. Biol. **207**, 1811–1823.
Norris, K. S., and Prescott, J. H. (**1961**). "Observations on Pacific cetaceans of Californian and Mexican waters," Univ. Calif. Publ. Zool. **63**, 291–402.
Norris, K. S., Wursig, B., Wells, R. S., and Wursig, M. (**1994**). *The Hawaiian Spinner Dolphin* (University of California Press, Berkeley, CA).
Nowacek, D. P. (**2005**). "Acoustic ecology of foraging bottlenose dolphins (*Tursiops truncatus*), habitat-specific use of three sound types," Marine Mammal Sci. **21**, 587–602.
Okutani, T., and McGowan, J. A. (**1969**). "Systematics, distribution, and abundance of the epiplanktonic squid (Cephalopoda, Decapoda) larvae of the California Current April, 1954-March, 1957," *Bulletin of the Scripps Institution of Oceanography* (Scripps Institution of Oceanography, La Jolla, CA), pp. 26–31.
Oleson, E. M., Wiggins, S. M., and Hildebrand, J. A. (**2007**). "Temporal separation of blue whale call types on a southern California feeding ground," Anim. Behav. **74**, 881–894.
Orr, R. T. (**1966**). "Risso's dolphin on the Pacific coast of North America," J. Mammal. **47**, 341–343.
Philpott, E., Englund, A., Ingram, S., and Rogan, E. (**2007**). "Using T-PODs to investigate the echolocation of coastal bottlenose dolphins," J. Mar. Biol. Assoc. U.K. **87**, 11–17.
Roper, C. F. E., and Young, R. E. (**1975**). "Vertical distribution of pelagic cephalopods," Smithson. Contrib. Zool. **209**, 1–51.
Shane, S. H. (**1994**). "Occurrence and habitat use of marine mammals at Santa Catalina Island, California from 1983-91," Bull. South. Calif. Acad. Sci. **93**, 13–29.
Shane, S. H. (**1995**). "Relationship between pilot whales and Risso's dolphins at Santa Catalina Island, California, USA," Mar. Ecol.: Prog. Ser. **123**, 5–11.
Smith, G. J., and Gaskin, D. E. (**1974**). "The diet of harbor porpoises (*Phocoena phocoena*) in coastal waters of eastern Canada, with special reference to the Bay of Fundy," Can. J. Zool. **52**, 777–782.
Soldevilla, M. S. (**2008**). "Risso's and Pacific white-sided dolphins in the Southern California Bight: Using echolocation clicks to study dolphin ecology," Ph.D. thesis, University of California, San Diego, La Jolla, CA.
Soldevilla, M. S., Henderson, E. E., Campbell, G. S., Wiggins, S. M., Hildebrand, J. A., and Roch, M. A. (**2008**). "Classification of Risso's and Pacific white-sided dolphins using spectral properties of echolocation clicks," J. Acoust. Soc. Am. **124**, 609–624.
Van Parijs, S. M., and Corkeron, P. J. (**2001**). "Vocalizations and behaviour of Pacific humpback dolphins *Sousa chinensis*," Ethology **107**, 701–716.
Van Parijs, S. M., Hastie, G. D., and Thompson, P. M. (**1999**). "Geographical variation in temporal and spatial vocalization patterns of male harbour seals in the mating season," Anim. Behav. **58**, 1231–1239.
Welch, P. D. (**1967**). "The use of fast Fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms," IEEE Trans. Audio Electroacoust. **15**, 70–73.
Wiggins, S. M., and Hildebrand, J. A. (**2007**). "High-frequency acoustic recording package (HARP) for broad-band, long-term marine mammal monitoring," in *International Symposium on Underwater Technology 2007 and International Workshop on Scientific Use of Submarine Cables and Related Technologies 2007* (Institute of Electrical and Electronics Engineers, Tokyo, Japan), pp. 551–557.
Zar, J. H. (**1999**). *Biostatistical Analysis* (Prentice-Hall, Upper Saddle River, NJ).

Localization of sperm whales in a group using clicks received at two separated short baseline arrays

Ryo Hirotsu and Masao Yanagisawa
Department of Computer Science, Waseda University, 3-4-1 Ohkubo, Shinjukuku, Tokyo 169-8555, Japan

Tamaki Ura, Masao Sakata, and Harumi Sugimatsu
Underwater Technology Research Center, Institute of Industrial Science, University of Tokyo, 4-6-1 Komaba, Meguro, Tokyo 153-8505, Japan

Junichi Kojima
KDDI Research and Development Laboratory, 2-1-5 Ohara Fujimino, Saitama 356-8502, Japan

Rajendar Bahl
Indian Institute of Technology Delhi, Hauz Khas, New Delhi 110016, India

(Received 16 June 2009; revised 21 October 2009; accepted 23 October 2009)

In this paper, a sperm whale click analysis scheme is proposed in order to calculate the position of individual sperm whales in a group using data received at two arrays deployed near the surface. The proposed method mainly consists of two parts: short baseline (SBL) with classification and long baseline (LBL) with class matching. In SBL with classification, a click is automatically detected, and its direction of arrival is calculated. The clicks are then classified based on their direction vectors. The class data are then sent together with direction data and matched to the other array's class data. LBL with class matching is used for localization. The classification algorithm can be used to estimate the number of whales clicking and to list potential candidates for LBL matching. As a result, the proposed method is able to localize the positions of the whales in a group. The performance of the proposed method is evaluated using data recorded off Ogasawara islands with two arrays near the surface. The three-dimensional underwater trajectories of six sperm whales are extracted to demonstrate the capability of the proposed method.
 [DOI: 10.1121/1.3268593]

PACS number(s): 43.30.Wi, 43.80.Ev [AIT] Pages: 133–147

I. INTRODUCTION

Many researchers have taken great interest in marine mammals in order to unlock the secrets of life in the deep sea. Sperm whales, *Physeter macrocephalus*, are known to emit impulsive sounds called clicks during their dives, which can reach more than 1000 m in depth (Levenson, 1974; Jaquet *et al.*, 2001), making it difficult to observe their underwater behavior. Furthermore, sperm whales are recognized as an endangered species and so investigation of their ecology in deep seas must be carried out without influencing their natural behavior. The observation of sperm whales was traditionally limited to visual observation from a boat, where pictures of their tails were used for identification.

Recent approaches to the study of sperm whales can be divided into two methods: bio-logging and acoustic analysis. In bio-logging (Hooker *et al.*, 2007), a tag containing several devices such as magnetic compass, inclinometer, depth indicator, and flash memory is attached to a target whale in order to estimate the motion characteristics during their dives (Amano and Yoshioka, 2003; Aoki *et al.*, 2007a, 2007b). The other method, acoustic analysis, typically uses multiple hydrophones to compare the time difference between the clicks arriving at each hydrophone to estimate the positions of the sperm whales or to analyze the characteristics of the clicks (Watkins and Schevill, 1977; Møhl *et al.*, 2000, 2001; Wahlberg, 2002; Thode, 2004; Tiemann *et al.*, 2006; Nosal *et al.*, 2007). Lately, work has been carried out, which combines the two methods by using a tag containing a hydrophone attached to a whale and a hydrophone array deployed on the surface. The clicks received by the two types of hydrophones are compared in order to estimate the influence of the whales' orientation on their vocalization (Watkins *et al.*, 1993; Johnson and Tyack, 2003; Zimmer *et al.*, 2003, 2005; Teloni *et al.*, 2008). However, the method using tags is only applicable to observe a limited number of whales, and it is not suitable to study whales in groups.

In our previous work, the co-authors introduced an acoustic observation system using an autonomous underwater vehicle (AUV) and passive sonar arrays (Ura *et al.*, 2004a, 2004b). The two arrays near the surface detect the sperm whale clicks, and the direction to the sound source is determined by a short baseline (SBL) system. Based on triangulation with a pair of separated SBL arrays, it is possible to calculate the location of the whale in real-time by a long baseline (LBL) system. The system recognizes the individual sperm whales so that the AUV can follow a specific sperm whale and possibly take pictures of it in the deep sea. However, sperm whales usually swim and dive in groups (Leatherwood, 1983), so each array captures thousands of clicks from multiple whales during a given period. Therefore, it is necessary to identify the corresponding clicks of each individual acquired by each array in order to triangulate its position. In the localization method using LBL, it is necessary to have a long baseline using a pair of arrays. So an array attached to an AUV would enable long baseline calculation,

while it follows a sperm whale for observation. However, the acoustic communication bandwidth between an AUV and the mother ship is limited, so it is impossible to transmit and correlate the raw acoustic data in real-time between two arrays. Therefore, we need a different method for triangulation that uses significantly less data exchange between the two arrays.

In this paper, an analysis scheme of sperm whale clicks is proposed using two hydrophone arrays in order to localize specific sperm whales in a group. In the proposed method, the classification of clicks is performed at each array considering the direction of arrival. The classification is used not only to estimate the number of whales but also to list the potential candidates for triangulation. Only the class data constructed both arrays are analyzed together with the direction data, and the class of one array is matched with the other array's classes to calculate the position of the whale. Thus, the proposed method is applicable for positioning the sperm whales in a group. The results of the proposed method using acoustic data recorded off Ogasawara with two arrays near the surface show that the three-dimensional (3D) trajectories of individual sperm whales in a group can be achieved using this approach.

The rest of this paper is organized as follows. The basic concept of the acoustic system with a pair of arrays is described in Sec. II. The proposed click analysis scheme is described in Sec. III. The results are shown in Sec. IV, followed by discussions in Sec. V, and conclusions in Sec. VI.

II. DEVELOPED ACOUSTIC SYSTEM

The designed acoustic system consists of two basic parts: a SBL system for direction calculation and a LBL system for position calculation. Figure 1(a) shows the overview of the acoustic system. Each SBL system consists of four hydrophones to calculate the relative direction of click arrival. The relative direction to the sperm whale is calculated from the time difference of the corresponding click between hydrophones. An inclinometer and a magnetic compass are attached to the array in order to convert the relative direction to the absolute direction. With a pair of arrays, the LBL system calculates the three-dimensional position of the individual sperm whale. The two directions calculated by the arrays and the absolute position of each array measured by a GPS on each of the boat are used for triangulation.

A. SBL system

1. Calculation of direction of arrival

The relative direction to the sperm whale is calculated from the time delay of a click's arrival at each hydrophone. Figure 1(b) is a diagram of the relationship between the direction and time delay. The time delay between the hydrophones i and j is defined as ΔT_{ij}, and the angle formed by the direction of arrival and the line connecting points of the hydrophones defines ϕ_{ij}. The distance between the hydrophones is d_{ij} and the underwater sound velocity is v_s. The relationship for ϕ_{ij} can be described as follows:

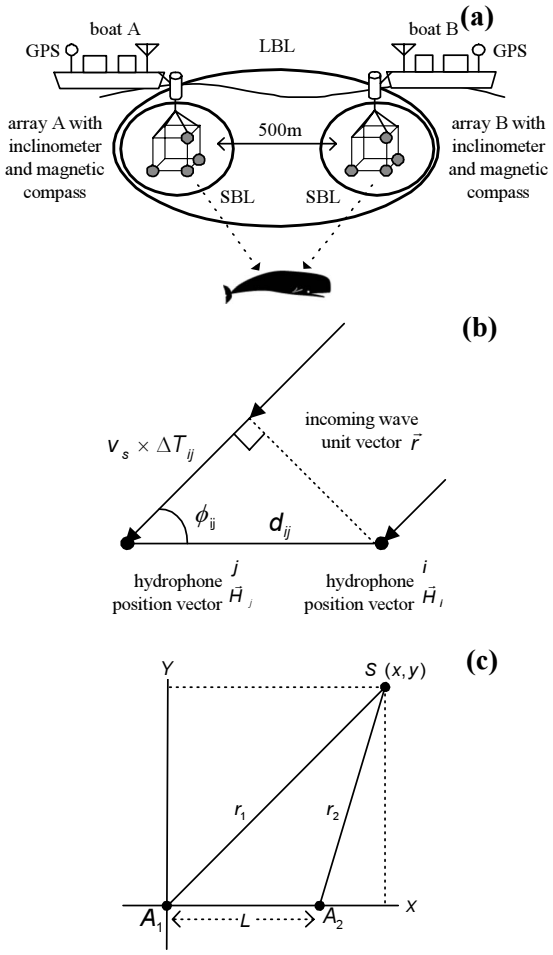


FIG. 1. Developed acoustic system: panel (a) shows the overview of the acoustic system; the middle panel (b) shows the principle of the direction calculation; and panel (c) shows the localization geometry using the designed acoustic system.

$$\cos \phi_{ij} = v_s \Delta T_{ij}/d_{ij}. \tag{1}$$

The position vector of hydrophone i is expressed as $\vec{H}_i$ in polar coordinates, and the three-dimensional unit vector of incoming wave is defined as $\vec{r}$. We can also write

$$d_{ij} \cos \phi_{ij} = (\vec{H}_j - \vec{H}_i) \cdot \vec{r}. \tag{2}$$

Combining Eqs. (1) and (2), we have

$$(\vec{H}_j - \vec{H}_i) \cdot \vec{r} = v_s \cdot \Delta T_{ji}. \tag{3}$$

Thus, the unit vector $\vec{r}$ can be calculated by solving the following simultaneous equation:

$$\begin{pmatrix} (\vec{H}_2 - \vec{H}_1)^T \\ (\vec{H}_3 - \vec{H}_1)^T \\ (\vec{H}_4 - \vec{H}_1)^T \end{pmatrix} \vec{r} = \begin{pmatrix} v_s \cdot \Delta T_{12} \\ v_s \cdot \Delta T_{13} \\ v_s \cdot \Delta T_{14} \end{pmatrix}. \tag{4}$$

Hence, it is possible to calculate the direction to the sound source using a minimum of three pairs of hydrophones.

2. Measurement of time delay

In order to measure the time delay between hydrophones, a cross correlation function is applied to the ex-

tracted segments containing a click. The signals received by the hydrophones i and j are defined as x_i and x_j, and the cross correlation function $X_{\mathrm{corr}(x_i,x_j)}$ is calculated using fast Fourier transform (FFT) and inverse fast Fourier transform (IFFT) as follows:

$$X_{\mathrm{corr}(x_i,x_j),k} = \mathrm{IFFT}(\mathrm{FFT}(x_i) \cdot \overline{\mathrm{FFT}(x_j)})_{\mathrm{real},k}. \tag{5}$$

The value of kth sample corresponding to the peak value in $X_{\mathrm{corr}(x_i,x_j)}$ is equal to the time delay between x_i and x_j. The five points surrounding the peak of the function $X_{\mathrm{corr}(x_i,x_j)}$ are approximated by a quadratic curve. The peak of the fitted quadratic curve is used as the estimated time delay.

3. Direction calculation with nonlinear least-squares method

The direction calculation is performed using the nonlinear Levenberg–Marquardt least-squares method applied on all possible pairs of hydrophone time delays. The inner product of direction vector of the sound source $\vec{r}_w$, the position vectors $\vec{H}_i$, and arrival vector $\vec{r}_\phi$ are used in the calculation. The observation vector $\vec{v}_{\mathrm{obs}}$ is defined as follows:

$$\vec{v}_{\mathrm{obs}} = (v_s \cdot \Delta T_{12} v_s \cdot \Delta T_{13} v_s \cdot \Delta T_{14} v_s \cdot \Delta T_{23} v_s \cdot \Delta T_{24} v_s \cdot \Delta T_{34})^T, \tag{6}$$

and the vector $\vec{v}_{\mathrm{cal}}$ is defined as

$$\vec{v}_{\mathrm{cal}} = ((\vec{H}_2 - \vec{H}_1) \cdot \vec{r}_\phi(\vec{H}_3 - \vec{H}_1) \cdot \vec{r}_\phi(\vec{H}_4 - \vec{H}_1) \cdot \vec{r}_\phi(\vec{H}_3 - \vec{H}_2) \cdot \vec{r}_\phi(\vec{H}_4 - \vec{H}_2) \cdot \vec{r}_\phi(\vec{H}_4 - \vec{H}_3) \cdot \vec{r}_\phi)^T. \tag{7}$$

The difference of the sum of squares $|\vec{v}_{\mathrm{obs}} - \vec{v}_{\mathrm{cal}}|^2$ is used as the convergence parameter. The optimization parameter is the three-dimensional vector component of $\vec{r}_w$, and the initial value of this vector is taken from Eq. (4) using lower-upper (LU) decomposition.

B. LBL system

Figure 1(c) shows the example of the location calculation in a plane. The position of the sound source S is defined as (x,y), and the two arrays A_1 and A_2 are located at (0,0) and $(L,0)$. The ranges to S from A_1 and A_2 are defined as r_1 and r_2, and the angles formed by S with the x-axis at A_1 and A_2 are defined as θ_1 and θ_2, respectively. From Fig. 1(c), the following relations can be obtained:

$$x = L \cos\theta_1 \sin\theta_2 / \sin(\theta_2 - \theta_1), \tag{8}$$

$$y = L \sin\theta_1 \sin\theta_2 / \sin(\theta_2 - \theta_1), \tag{9}$$

$$r_1 = L \sin\theta_2 / \sin(\theta_2 - \theta_1), \tag{10}$$

$$r_2 = L \sin\theta_1 / \sin(\theta_2 - \theta_1). \tag{11}$$

Therefore, the two-dimensional position of the sound source $S(x,y)$ and the ranges to the sound source, r_1 and r_2, can be calculated from the direction of arrival at each array, θ_1 and θ_2, respectively.

1. Location calculation with least-squares method

The 3D position of a whale is calculated as the intersection of the two direction vectors based on triangulation. However, the calculated direction vectors would inevitably have errors, and the calculated directions may not intercept in 3D space. Thus, a least-squares method is applied for the position calculation. The unit vectors to the whale, $\vec{v}_1$ and $\vec{v}_2$, measured by the two arrays are defined as follows:

$$\vec{v}_1 = (v_{1x} v_{1y} v_{1z})^T, \tag{12}$$

$$\vec{v}_2 = (v_{2x} v_{2y} v_{2z})^T. \tag{13}$$

The position vector of the two arrays in earth coordinates, $\vec{g}_1$ and $\vec{g}_2$, are described by Eqs. (14) and (15):

$$\vec{g}_1 = (g_{1x} g_{1y} g_{1z})^T, \tag{14}$$

$$\vec{g}_2 = (g_{2x} g_{2y} g_{2z})^T. \tag{15}$$

The true position vector of a whale $\vec{P}$ is defined as follows:

$$\vec{P} = (xyz)^T. \tag{16}$$

The calculation error vectors for each array are defined as $\vec{E}_1$ and $\vec{E}_2$, respectively, and are described by Eqs. (17) and (18):

$$\vec{E}_1 = \vec{P} - (\vec{g}_1 + k_1 \vec{v}_1), \tag{17}$$

$$\vec{E}_2 = \vec{P} - (\vec{g}_2 + k_2 \vec{v}_2), \tag{18}$$

where k_1 and k_2 represent the range to the position from array A and B, respectively. Here, the error E is defined as

$$E = |\vec{E}_1|^2 + |\vec{E}_2|^2. \tag{19}$$

From Eqs. (12)–(19), E can be described as a function of k_1, k_2, x, y, z and $\vec{v}_1, \vec{v}_2, \vec{g}_1, \vec{g}_2$. And E is differentiated by each parameters k_1, k_2, x, y, z. Then in order to minimize E, the following simultaneous equation is solved:

$$\vec{A}\vec{x} = \vec{b}, \tag{20}$$

where

$$\vec{A} = \begin{pmatrix} v_{1x}^2 + v_{1y}^2 + v_{1z}^2 & 0 & -v_{1x} & -v_{1y} & -v_{1z} \\ 0 & v_{2x}^2 + v_{2y}^2 + v_{2z}^2 & -v_{2x} & -v_{2y} & -v_{2z} \\ v_{1x} & v_{2x} & -2 & 0 & 0 \\ v_{1y} & v_{2y} & 0 & -2 & 0 \\ v_{1z} & v_{2z} & 0 & 0 & -2 \end{pmatrix}, \tag{21}$$

$$\vec{x} = \begin{pmatrix} k_1 & k_2 & x & y & z \end{pmatrix}^T, \tag{22}$$

$$\vec{b} = \begin{pmatrix} v_{1x} g_{1x} + v_{1y} g_{1y} + v_{1z} g_{1z} \\ v_{2x} g_{2x} + v_{2y} g_{2y} + v_{2z} g_{2z} \\ g_{1x} + g_{2x} \\ g_{1y} + g_{2y} \\ g_{1z} + g_{2z} \end{pmatrix}. \tag{23}$$

Hence, the position of the whale can be identified by solving the simultaneous equations using LU decomposition.

C. Localization error

The localization error models for calling marine mammals using widely spread hydrophones are described in detail by Spiesberger and Fristup (1990) and Wahlberg *et al.* (2001). In our localization method, the closely spaced hydrophones on the surface vessel are used to measure the time delay of arrival. Since the hydrophones in SBL array are separated by less than 1 m, the propagation conditions are identical, so the sound refraction does not contribute to the time delay measurement error, even though it is concerned to the localization error. The error factors of this localization are considered as follows: SBL orientation error, sound refraction error, and global positioning system (GPS) positioning error.

1. SBL orientation error

The SBL error is attributed to its estimation accuracy of direction of arrival in SBL system and the accuracy of inbuilt inclinometer and magnetic compass, and it affects both horizontal and vertical direction estimations. The calibration of the SBL orientation calculation was performed in the experimental pool. The length of the baseline in each hydrophone was 30 cm, and the inbuilt magnetic compass and inclinometers were attached to the array. The sound transmitter was set at the distance of 3 m from the array. The array was rotated about its horizontal and vertical axes, respectively. The transmitted signal was a burst signal of 10 kHz sine wave. The averaged angular directional error in the horizontal direction is 0.10°, and the averaged angular difference in the vertical direction is 0.607°.

2. Sound refraction error and sound velocity profile

The pathway of the sound from the source to the array on surface is not straight but bent. The calculated direction observed on the surface has a margin of error compared to the correct direction to the sound source due to the sound refraction in mainly vertical direction. To estimate the effects of sound refraction, the variation in the water temperature with depth in the experimental area in Ogasawara was measured using expendable bathythermograph (XBT). The calculated sound velocity profile is shown in Fig. 2(a). From this sound velocity profile, the angular difference caused by refraction is measured by Snell's law when the depth of the sound source is found.

3. Estimation of localization error in horizontal and vertical axes

From these results, the localization error can be estimated in both horizontal and vertical axes. In order to estimate the localization error in horizontal axis, the positions of the arrays are defined as $(x,y)=(-200,-150)$ and (200, 150), respectively. The position of the sound source is varied over 50 m, so the directional vector to each position of the source including the SBL orientation error in each array can be calculated. The GPS positioning error is defined as 10 m. Therefore, the localization error can be estimated, and the estimation result of localization error is described in Fig. 2(b).

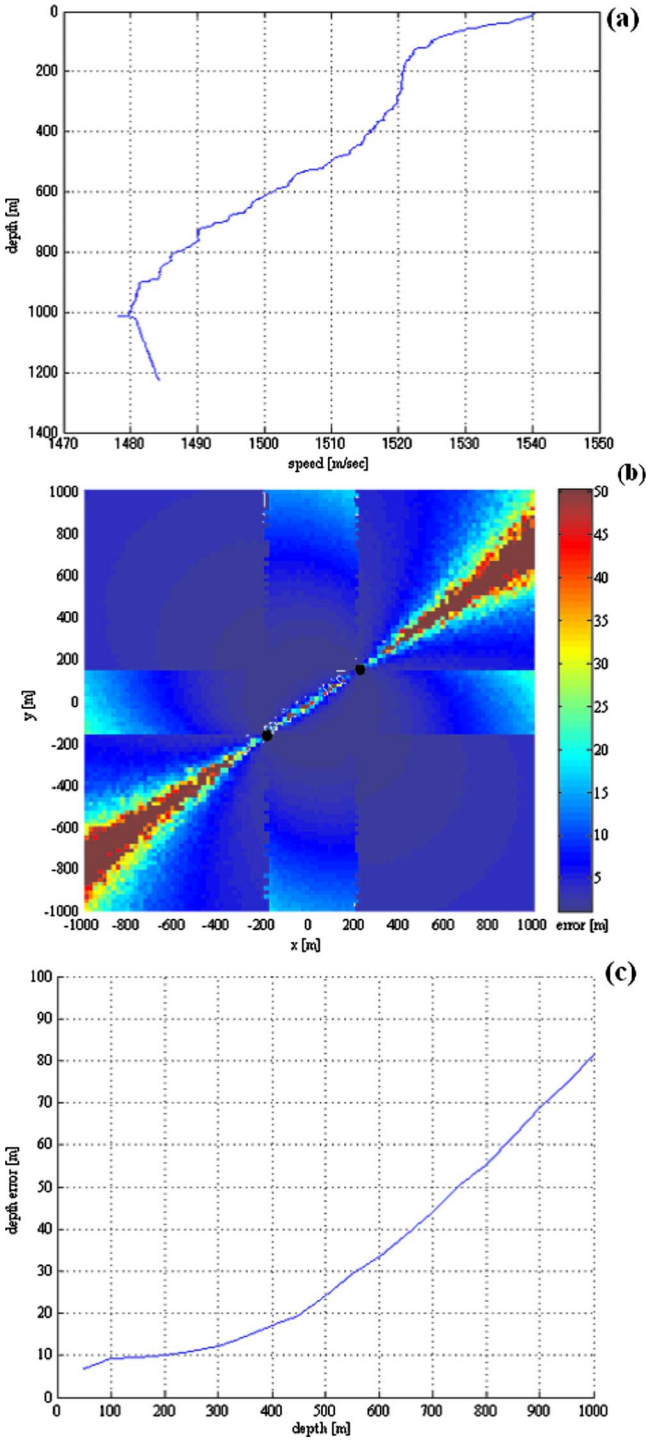


FIG. 2. (Color online) Estimation of localization error: panel (a) shows the sound velocity profile using data from XBT by the equation of Del Grosso (salinity is set to be constant at 35%); panel (b) shows the estimation results of localization error in horizontal plane, and in this panel, 50 m is set to the maximal distance error (the position of each array is indicated as a circle); panel (c) shows the estimation results of vertical error.

Thereafter, the position of the source is defined as (0, 0), and the depth is changed from 50 to 1000 m in 50 m increments. The direction including the SBL orientation error and sound refraction error can be calculated, and the results are described in Fig. 2(c).

4. Deployment method of the designed system

The localization error in vertical axis is larger than the error in horizontal axis because three of the four hydro-

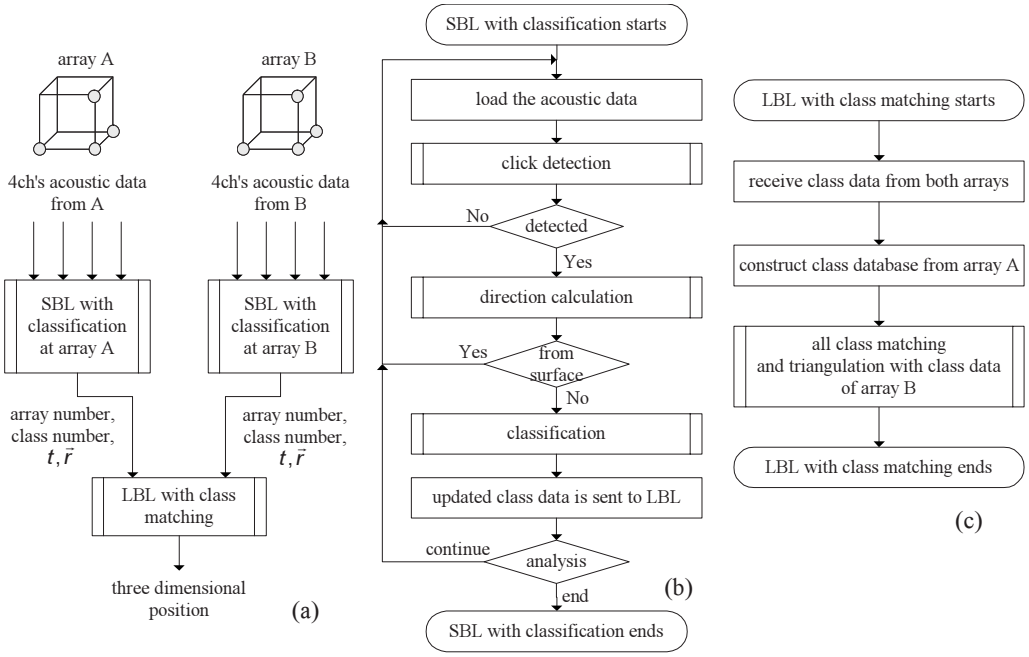


FIG. 3. Proposed analysis flow to calculate the position of the whale: panel (a) shows the block diagram of the data flow of the proposed analysis; the middle panel (b) shows the analysis flow of the SBL with classification; panel (c) shows the analysis flow of the LBL with class matching.

phones are designed to be on the same horizontal plane. The vertical direction has to also include the error due to refraction. From Fig. 2(b), the area around the straight line connecting two arrays have a large localization error in horizontal plane, so the array should be deployed at right angle to the sperm whale's directional movement. If a sperm whale swims straight though the line connecting two arrays, the localization expects to perform with sufficient accuracy for identification. From Fig. 2(c), the body length of sperm whale is less than 30 m, so it is difficult to separately identify sperm whales swimming together at the depth of more than 600 m.

III. PROPOSED ANALYSIS SCHEME

In this section, the proposed analysis scheme for individual sperm whale localization is described. Figure 3(a) shows the data flow of the proposed analysis. Figure 3(b) shows the analysis flow of the SBL, which performs click detection, direction calculation, and classification. Figure 3(c) shows the flow of the LBL, which performs class matching and triangulation to calculate the position of the sperm whale using the acoustic data of the two arrays. Further details of the algorithm in Fig. 3 are given in Figs. 4 and 5.

A. SBL with classification

The SBL in Fig. 3(b) is performed at each array, and so the click detection, direction calculation, and classification are performed independently in the two arrays. The first task of the analysis is to load the acoustic data from the hydrophones for each 5 ms segment. Then, click detection is performed, as shown in Fig. 4(a). If a click is detected, the direction calculation is performed, as shown in Fig. 4(b). After considering the false direction that may be caused by surface reflection, the classification is performed on the detected click. The class data are sent to the LBL with class matching part to calculate the position of the sperm whale.

1. Click detection

Figure 4(a) shows the block diagram of the analysis flow for click detection. The signals captured by each hydrophone may contain clicks from multiple whales. The duration of one click is typically less than 5 ms, and the interval between clicks for a whale is about 0.5–2 s. The data used for the click detection are the segmented acoustic data of 5 ms duration captured by one of the hydrophones.

The click is automatically detected, as described in Fig. 4(a) and based on the detected time of the hydrophone; other three channels' acoustic data containing the click are extracted to calculate the direction of arrival, as shown in Fig. 4(b). Since the clicks have significant energy compared to the transient noise, an energy-based thresholding is expected to perform better than simple level-based thresholding (Bahl *et al.*, 2003). The energy over each 5 ms segment is calculated and compared to a threshold, which is determined to be $\alpha_{\min}$ times the average energy of the preceding segments. On the other hand, if the energy is $\alpha_{\max}$ times the average energy, the energy is too large to be regarded as that of a click. Once a segment containing a click is detected, the peak time of this segment is marked in order to re-extract 5 ms data including the click from all four channels of the array for time delay estimation. Considering that the surface reflection comes after the direct click, the load pointer for the acoustic data is updated for α_t ms in order to prevent the surface reflected click from being analyzed.

2. Direction calculation

Figure 4(b) shows the block diagram of the analysis flow of the direction calculation. The first task of the direction calculation is to extract the four channels' 5 ms acoustic data

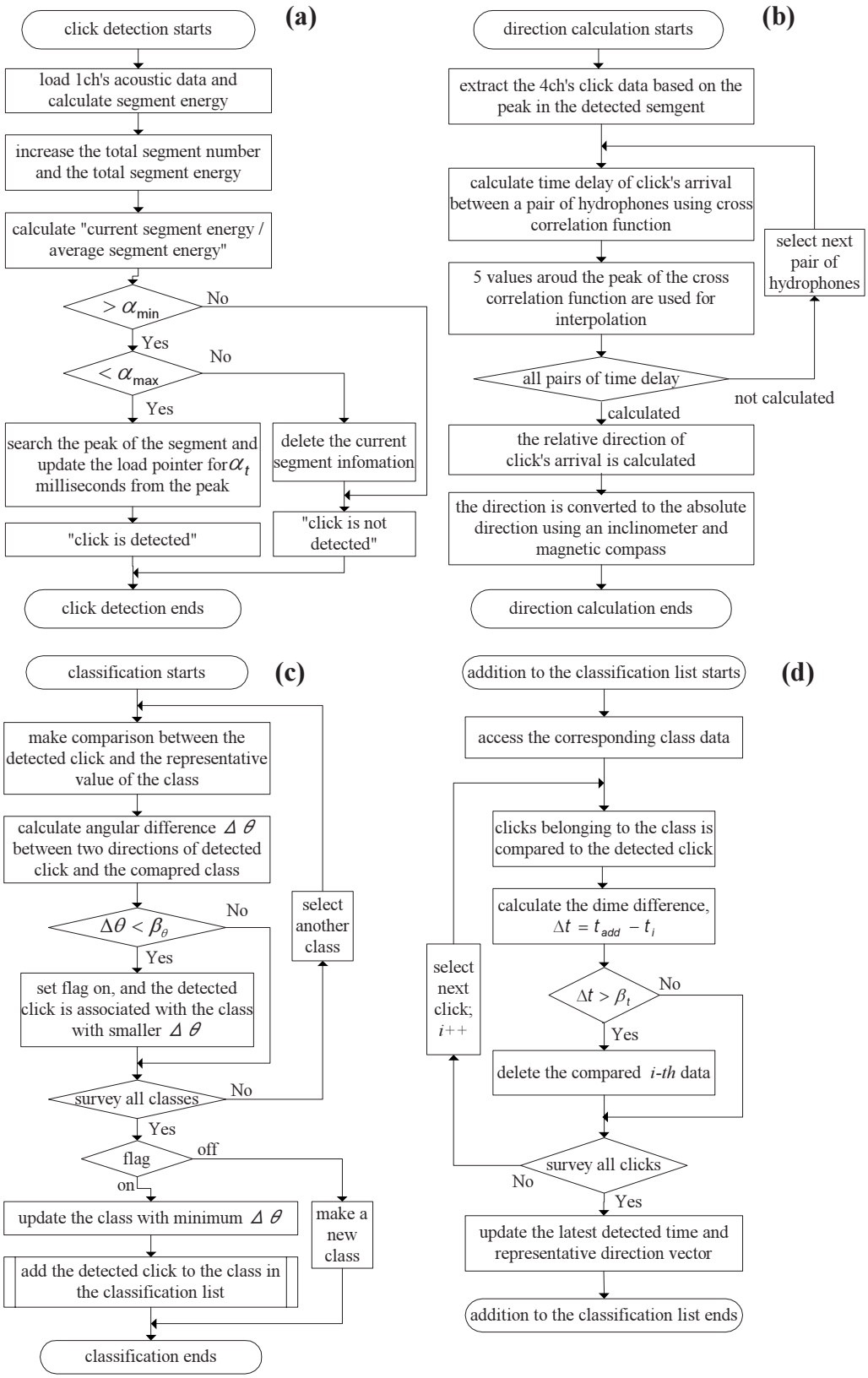


FIG. 4. Analysis flow of the click detection, direction calculation, and classification: panel (a) shows the analysis flow of the click detection; panel (b) shows the analysis flow of the direction calculation; panel (c) shows the analysis flow of the classification; panel (d) shows the analysis flow of the addition click data to the classification list.

starting from α_s m prior to the peak in the detected segment. After extracting the acoustic data, the time delay between all pairs of hydrophones is calculated using a cross correlation function, as described in Sec. II. After interpolating the peak of the cross correlation function, the relative direction of the click's arrival is calculated. Then, the relative direction is converted to the absolute direction using measurements of an inbuilt inclinometer and magnetic compass.

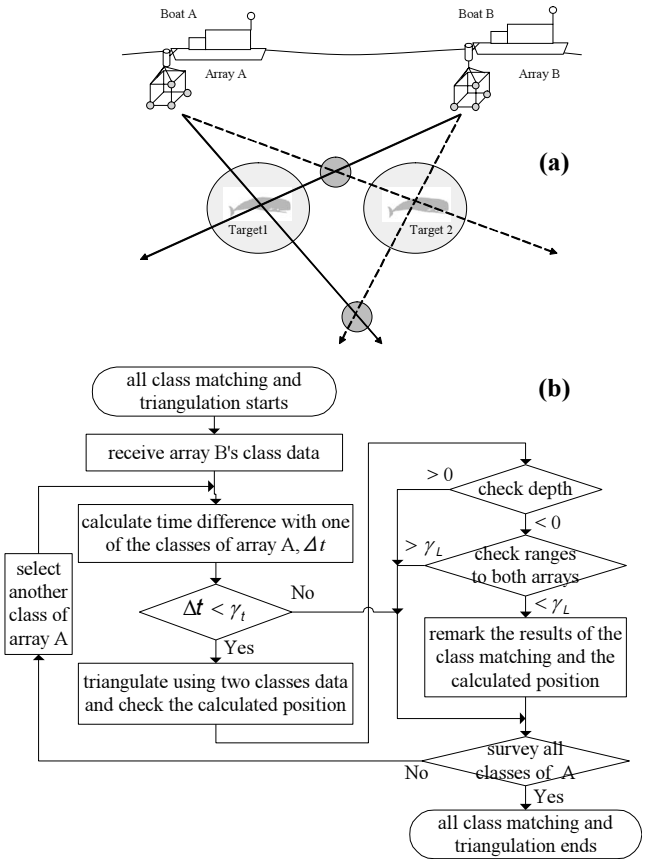


FIG. 5. All class matching: panel (a) shows the example of class mismatching and mislocalization, and panel (b) shows the analysis flow of all-class matching.

3. Classification

Figures 4(c) shows the block diagram of the analysis flow of classification. The classification is performed based on the results of direction calculation at a particular hydrophone array. The main purpose of the classification is to compress the data based on class features for matching candidate classes between the two arrays that leads to successful triangulation.

The class data constructed by classification are defined by C, and its elements, $\{n, T, t, \vec{R}, \vec{r}\}$, are defined as follows:

$$C = \{n, T, t, \vec{R}, \vec{r}\}, \tag{24}$$

$$T = \{t_i\}(i = 1,2, \ldots, n), \tag{25}$$

$$\vec{R} = \{\vec{r}_i\}(i = 1,2, \ldots, n). \tag{26}$$

In these equations, n is the number of clicks belonging to the class and represents the size of the data set. The time of each click's arrival is defined as T, and t is the last detected time. The set of direction vectors of each click is defined as $\vec{R}$, and $\vec{r}$ is the representative value of the direction vector, which is defined as follows:

$$\vec{r} = \sum_{i=1}^{n} \vec{R}_i / n. \tag{27}$$

Figure 4(c) shows the flow of the classification algorithm. It constructs a classification list that takes the list structure and stores the class data and the analyzed click data. When detecting and analyzing a new click, the click's direction vector and its detected time, $\{t_{\text{add}}, \vec{r}_{\text{add}}\}$, are compared to the existing classes in order to determine whether the click belongs to the existing classes or not. The angular difference between the detected click and the compared class $\Delta\theta$ is calculated as

$$\Delta\theta = \arccos(\langle \vec{r}, \vec{r}_{\text{add}} \rangle / (|\vec{r}||\vec{r}_{\text{add}}|)). \tag{28}$$

It marks those class numbers whose angular difference is within angular threshold β_θ. If no class is matched, a new class is made. Otherwise, the click is set to belong to the class having the smallest angular difference, as described in Fig. 4(c).

In order to add the click data to a class, the index of the corresponding class is accessed. The clicks belonging to the class are compared with the added click, and the time difference Δt between the added click and the compared click of the class is calculated; $\Delta t = t_{\text{add}} - t_i$. If Δt exceeds the time threshold β_t (in seconds), the ith click data are regarded to be too late for the added click; t_i and $\vec{r}_i$ are deleted and n is decremented. The index of the list, which stores the representative value of the class, is updated using $\vec{r}$ and t; t is updated to t_{add} and $\vec{r}$ is updated according to Eq. (27). The direction vectors calculated in Fig. 4(b) have fluctuations, so averaging is performed on the direction vector of the clicks belonging to the class with Eq. (27) in order to reduce the fluctuations. After the classification is complete, the array number, class number, and updated t and $\vec{r}$ are sent to the LBL analysis part in order to match the classes and then calculate the position of the whale.

B. LBL with class matching

As mentioned above, the proposed analysis of the LBL with class matching part is shown in Fig. 3(c). The first step of the analysis is to collect the class data from two arrays, arrays A and B. The class data from array A are used as the class database, and class data from array B are used to search for the matching class in this database.

1. Matching issues

With the direction vectors of the classes at two separated arrays, matching errors may occur. An example is shown in Fig. 5(a), where two arrays, A and B, detect two whales, targets 1 and 2. Two classes would be constructed in each array based on the direction of each whale. However, there are four matching candidates, implying that four candidate positions of the whales would be calculated. In order to avoid this mismatching, it is necessary to unambiguously identify the corresponding classes using parameters, such as the inter-pulse interval (IPI). The IPI can be used as an identification parameter that originates from the internal reflection of a click between the frontal and distal sacs (Goold, 1996) of the whale. However, the IPI depends on body length, so it is inadequate to discriminate between individuals having the same body length. Moreover, IPI does not consistently show the same value for the same specific whale

(Zimmer *et al.*, 2005a). So, some robust parameter for reliable identification needs to be considered to find the corresponding classes.

2. All class matching

We propose another method to find the corresponding classes between two arrays based on the changes in the direction vectors observed in each array. After all classes are constructed, we compare class characteristics such as appearance and disappearance time and the changes in direction vectors between the classes of the two arrays. So, whenever the class data of array B are investigated, the class is matched with all classes of array A. Only the matching results of the corresponding classes are extracted to obtain the underwater trajectories of the sperm whales.

Figure 5(b) shows the analysis flow of all-class matching. After receiving the class data of array B, the detected time difference Δt between the two classes of arrays A and B is calculated. If the time difference Δt exceeds the time threshold γ_t, the class of A is determined not to be matched to the received class of B. The triangulation is performed using the two matched classes' direction vectors and the positions of the boats obtained by GPS. In order to reject an abnormal calculated position, we check two parameters of the position: The first one is the depth to confirm that the calculated position is not above the surface, and the other parameter is the range from each boat. If the range exceeds the range threshold γ_L (in meters), it is regarded to be too far to determine that there is a whale at the position. Thereafter, the whale positions are marked.

All the classes in the class database of array A are compared. The candidate positions are calculated automatically by the proposed method without any manual intervention. However, manual processing is necessary to correctly select the corresponding classes and to extract the localization results.

IV. EXPERIMENTAL RESULTS

The results obtained from data recorded off Ogasawara islands on August 21, 2003 by the proposed analysis are described in this section. The proposed method is implemented in click analysis software developed using C++. The software is able to analyze real-time data from the hydrophone arrays and other apparatus such as GPS, inclinometer, and magnetic compass.

A. Recorded data and parameter settings

Two boats, A and B, were used for the ocean experiment held at Ogasawara islands in 2003 (Ura *et al.*, 2004a). Upon spotting a group of sperm whales, the boats approached the vicinity of the group and the engines were stopped. The two boats retained sufficient range between them, i.e., about 500 m. The length of the baseline in each SBL array was 30 cm. After the hydrophone arrays were deployed from each boat, as illustrated in Fig. 1(a), acoustic data of sperm whales' clicks were obtained. The duration of recording was about 40 min. The acoustic data were acquired by hydrophones with a sensitivity of -185 dB re 1 V/μPa at 1 m, sampled with a frequency of 50 kHz per channel, and recorded with a resolution of 16 bits (Ura *et al.*, 2004b).

The threshold values and parameters in Sec. III and Figs. 4 and 5 are set as follows. The threshold values of $\alpha_{\min}=2$, $\alpha_{\max}=20$, $\beta_\theta=15°$, and $\beta_t=10$ s are obtained empirically. The value of $\alpha_s=1$ ms to extract four channels' acoustic data from the detected peak time is chosen considering that each hydrophone is separated by 30 cm. The value of $\alpha_t=10$ ms is taken to prevent the surface reflected click from being analyzed and is chosen considering the depth of the array (5 m). This implies that the surface reflection arrives with a maximum delay of 7 ms from the direct click. The localization threshold $\gamma_L=2000$ m, the time threshold $\gamma_t=1$ s is chosen considering the measurement error of the designed LBL system and its time delay.

B. Evaluation of the click detector

The operating characteristics of the click detector are quantified using results of the proposed method. The recorded data from array B are used for this quantification, and checking the sound, waveform, and spectrum of the clicks in the acoustic data; the time of the clicks are noted. The detected time of the clicks by the proposed method are compared to the correct time of the clicks. So it can be quantified whether the proposed method detects the clicks correctly or it fails by detecting the noise as a click. The correct detection rate is defined as follows: the number of clicks detected accurately divided by the number of the existing clicks. The false detection rate is defined as follows: the number of detected non-clicks divided by the number of detected as clicks.

In the time frame of first 10 min, 1882 clicks are confirmed to exist in the acoustic data of array B. Then, the number of clicks detected accurately is 1434, and the number of detected non-clicks is 53. Thus, the quantification of the operating characteristics of the click detector is as follows: The correct detection rate is 76.2%, and the false detection rate is 3.56%. The misdetected clicks are rejected in the beginning of the analysis because of the rather large energy. The false detections are surface reflected clicks, bottom reflected clicks, and transient noise.

C. Results of direction calculation and classification

Figures 6(a)–6(d) show the results of direction calculation at each array. The angles are calculated from the direction vectors and are the absolute values after compensation using the inclinometer and magnetic compass at each array. The bearing angle is measured with 0° as north and 90° as east, while 0° of depression angle means horizontal and 90° means directly below the array. The time $t=0$ is the starting time of recording after both arrays are deployed from the boats. From these figures, trajectories of multiple whales can be observed in both arrays. From the results of bearing angles, there seem to be six whales in the observation area.

The results of the classification are shown in Fig. 7. Figures 7(a) and 7(b) show the classification results of array A, and Figs. 7(c) and 7(d) show the classification results of

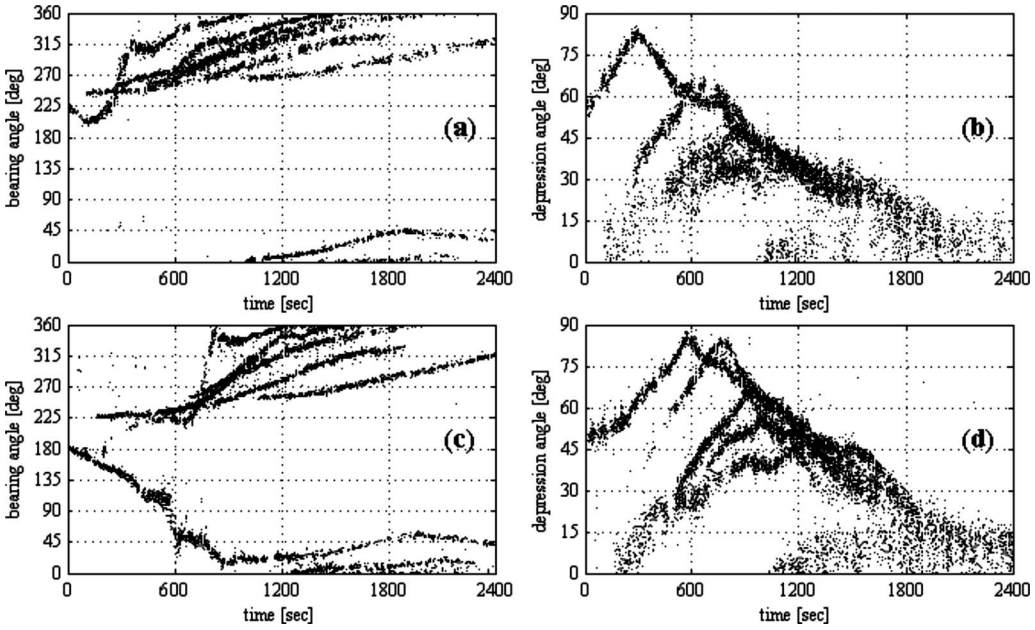


FIG. 6. Results of direction calculation: panel (a) shows the bearing angles of clicks at array A, and panel (b) shows the depression angles of clicks at array A. The lower left panel (c) shows the bearing angles of clicks at array B, and panel (d) shows the depression angles of clicks at array B.

array B. In these figures, only the classes containing over 100 clicks are shown. Sometimes classes that seem to be the same are disjointed; in Figs. 7(a) and 7(b), classes A0 and A19 should apparently be classified as the same class, but a short period that contained no clicks resulted in the classes being separated. In array A, 6599 clicks are detected out of which 6399 clicks (about 97%) belong to the classes with over 100 clicks. Similar to array B, 9641 clicks are detected out of which 9391 clicks (about 97%) belong to classes containing more than 100 clicks.

D. Extraction of corresponding classes

After the results of classification and candidate positions, manual processing is performed for rejection of any possible mismatches by checking the classes' characteristics: appearance time, disappearance time, and changes in direction vectors. At first, we look for the separated classes that belong to the same whale's clicks at each array. In array A [see Figs. 7(a) and 7(b)], classes A0, A19, A26, and A39 should be same; A0 appears from 0 to 650 s, A19 appears

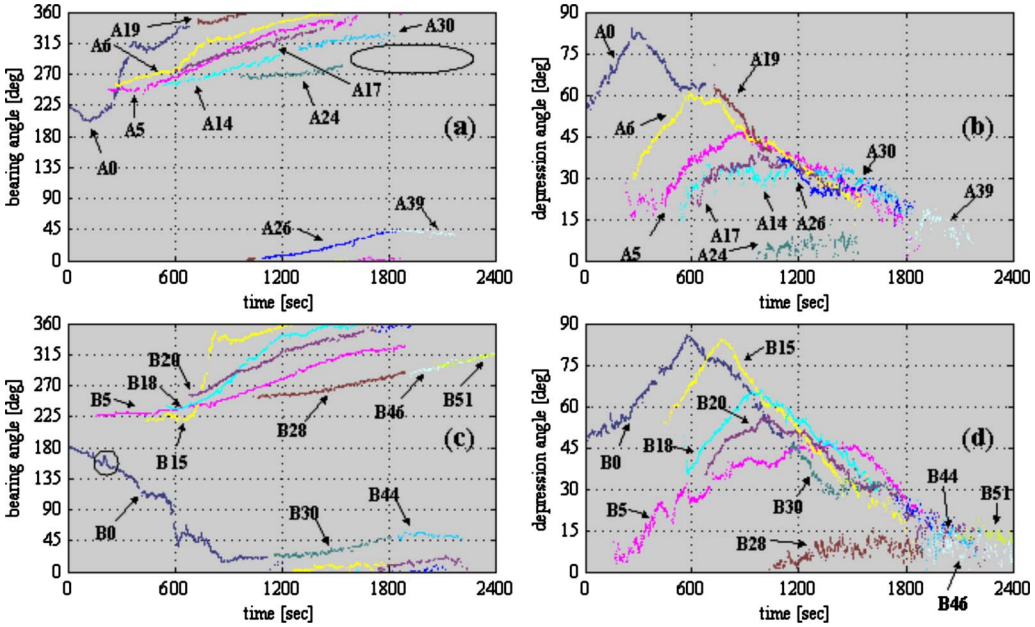


FIG. 7. (Color online) Results of classification: panel (a) shows the bearing angles of the classes at array A, and panel (b) shows the depression angles of the classes at array A. The lower left panel (c) shows the bearing angles of the classes at array B and panel (d) shows the depression angles of the classes at array B.

TABLE I. Characteristics of the classes of A and B; click time shows clicking time of the class, and shape shows the characteristics of the depression angle, as seen in Fig. 7. The time event of the shape and the corresponding value of the depression angle are also shown.

	Click time (s)		Characteristics of depression		
Class No.	Start	End	Shape	Time (s)	(deg)
			Classes of array A		
A0, A19, A26, A39	0	2167	Peak	303	82
A5	236	1867	Peak	915	46
A6	274	1556	Peak	611–763	58
A14, A30	547	1823	Dip	1012	30
A17	594	1434	Peak	1002	40
A24	982	1527	Flat	All	0–13
			Classes of array B		
B0, B30, B44	0	2002	Peak	585	85
B5	164	1891	Dip	1023	38
B15	458	1836	Peak	770	84
B18	586	1641	Peak	934	65
B20, B41	688	2217	Peak	1008	55
B28, B46, B51	1039	2386	Flat	All	0–15

after A0, A26 appears after A19, and A39 appears after A26. So, considering the class times and angles, the disjointed classes are merged as follows: (i) A0, A19, A26, and A39; (ii) A14 and A30; (iii) B0, B30, and B44; (iv) B28, B46, and B51; and (v) B20 and B44. The rest of the classes are solitary and do not need any merging.

E. Results of class matching

There are six matching candidates in each array. Considering that the depth of each array is equal, the matching criterion is the time of the event noted in the depression angle. Table I summarizes the characteristics of the classes of A and B. In the initial period of 0–300 s, each array has only one class, A0 and B0, which means that they should be matched as whale A. But the classes headed by A0 and B0 have large difference in the time of the event in the depression angle. The peak of A0 appears 303 s, and the peak of B0 appears in 585 s. The depression angle represents 0° as surface and 90° as directly below the array. However, depression angles that exceed 90°, for example, 120°, gets represented as 60°. So, if a whale dives past under the array, the depression angle first increases to 90° and subsequently decreases from 90°, even though the depth of the whale keeps decreasing. This is the reason why the large difference in the time event appears in the classes of A0 and B0. Again, at about 1000 s, there appear new classes, A24 and B28. The bearing and depression angles seem to correlate between these two classes, which implies that there is a whale near the surface, regarded as whale B.

Looking at rest four candidates' depression angles, three angles' peaks correlate with one another in the interval of 700–1000 s, and only one class has a dip in both arrays, A14 and B5. If they are said to be matched, then there is a problem that A14 has no data from 170 to 540 s, while B5 has data during this time. By comparing B5 and B18, B18 appears very close to B5 at about 600 s. Hence, it can be assumed that B5 before 600 s belongs to B18, and that B5 after 600 s matches with A14 as whale C. The motion and peak time of the depression angles of A6 match B15 in the time frame around 770 s, and get assigned as whale D. Similarly, we find that class A17 matches B20 as whale E and A5 matches B18 as whale F. Table II summarizes the matching classes.

TABLE II. Matching table.

Whale No.	Classes in array A	Classes in array B
Whale A	A0, A19, A26, A39	B0, B30, B44
Whale B	A24	B28, B46, B51
Whale C	A14, A30	B5
Whale D	A6	B15
Whale E	A5	B18
Whale F	A17	B20, B41

F. Trajectories

The whale matching shown in Table II can be used to estimate the trajectories of the six sperm whales in a group. Figures 8(a)–8(c) show the three-axis motion with respect to time. Figure 8(d) shows the movements of the six sperm whales and the two boats in the horizontal plane. In these figures, the origin, $x=0$ and $y=0$, is the initial position of the array A obtained by GPS, which is 27.0759° north latitude and 142.034° east longitude. Depth $z=0$ represents the surface and $z=-5$ is the depth of the arrays. Time $t=0$ is the start of the recordings. In Figs. 8(a)–8(c), the trajectories of whales A, B, C, D, E, and F are shown. From these figures, it can be seen that the movements of whales A and D were very similar, whales C, E, and F move together, and whale B moved independently from the other five whales and stayed near the surface. Moreover, the five whales appeared to swim in the same direction in the horizontal plane and repeatedly dived between 300 and 800 m depth.

V. DISCUSSIONS

The results of the estimated trajectories are discussed in this section. At first, the click's misdetection and the influence of multipath are discussed from the results of detection calculation and classification. The trajectories are smoothed in order to estimate the speed of the whales. Later, scope of further work on the proposed algorithm is discussed.

A. Misdetection and multipath

We have considered only the classes containing over 100 clicks for class matching and localization considering misdetection and multipath. If noise or nondirect clicks are analyzed, their direction of arrival is not constant so that the formed classes do not contain sufficient clicks. The clicks belonging to the classes are compared to the threshold in order to eliminate the false matching candidates. However, this also causes a problem in that the classes, which have insufficient clicks but are certainly constructed by sperm

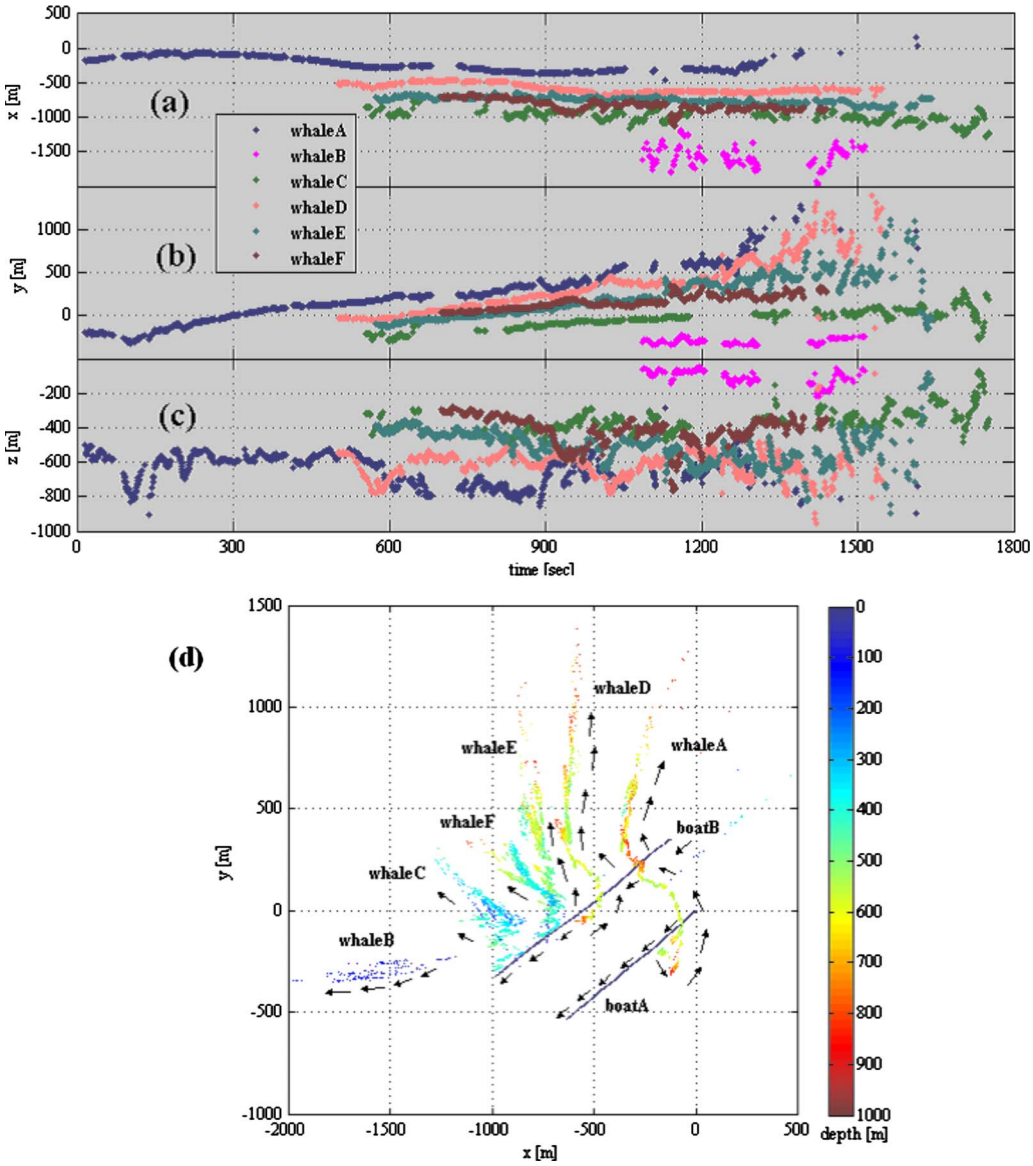


FIG. 8. (Color online) Triaxial movements: panel (a) shows the x-axis (east to west) motion; panel (b) shows the y-axis (north to south) motion; panel (c) shows the z-axis (depth) motion. The panel (d) shows the movements in the horizontal plane. In this panel, horizontal motions of six whales and two boats are described.

whale direct clicks, would not be used for matching. Figures 6(a) and 7(a) can be used to show an example of this problem. In this paper, the threshold for number of clicks is set to be 100. Looking at the click data in the time duration of 1500–1800 s, a sperm whale seems to exist at the bearing angle of 280° – 300°. But this class does not get recorded in the region described by the circle of Fig. 7(a) because the number of clicks is less than 100. The classification algorithm needs to be improved so that the disjointed classes from a specific whale can be classified as the same class in our future work.

B. Smoothing

The speeds of whales in a group are obtained by differentiating their positions with respect to time. Since the calculated positions have fluctuations, it is first necessary to smooth the trajectories. Linear interpolation is performed on the three-axis motion data, and the interpolated data are filtered using a Butterworth filter. The Butterworth filter is an order 1 low pass filter with normalized cutoff frequency of 0.001. Figure 9 shows the smoothed tri-axial motion of the six whales; Fig. 9(a) shows the smoothed x-axis motion of each whale, and Figs. 9(b) and 9(c) show the y-axis and z-axis motion, respectively.

1. Calculation of speed

The smoothed positions are used to calculate the speed of whales that is achieved by differentiating the smoothed positions with respect to time. The differentiation being sensitive to position fluctuations, so the calculated speed is applied to a Butterworth filter order 1 low pass filter with normalized cutoff frequency of 0.001.

Figure 10(a) shows the results of the filtered speed of six whales and Fig. 10(b) shows the vertical component of the speed. The estimated error of triangulation using two separated arrays mainly depends on the error in the estimated

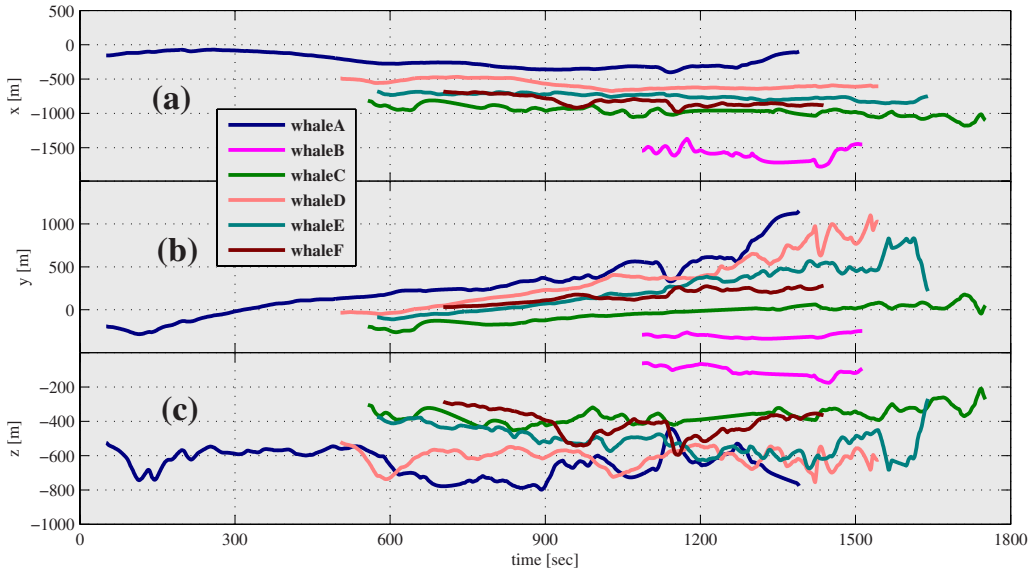


FIG. 9. (Color online) Smoothed triaxial motion: panel (a) shows the smoothed x-axis (east to west) motion; panel (b) shows the smoothed y-axis (north to south) motion; panel (c) shows the smoothed z-axis (depth) motion.

direction at each array, the length of baseline between two arrays, and the range to the target. Considering the typical range to the targets, only those positions that are within 1000 m from both arrays are shown in Figs. 10(a) and 10(b). It is to be cautioned that the calculated results of speed shown in Figs. 10(a) and 10(b) are the filtered values, which means that the filter parameters have an effect on the estimates of instantaneous speed. We show that the whales' speed varies between 1 and 4 m/s during diving, as shown in Fig. 10(a), and that vertical speed varies from −2 to +2 m/s, as shown in Fig. 10(b).

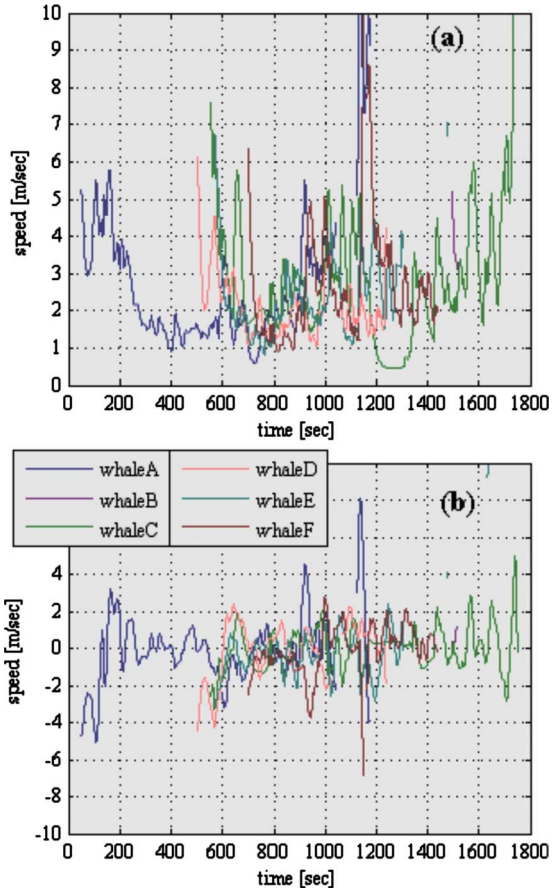


FIG. 10. (Color online) Speed of the whales: panel (a) shows filtered speed of the six whales, and panel (b) shows the vertical component of (a).

C. Relationships between the depth and ICI in a group

Figure 11 shows the relationships between the depth and ICI observed in array B of five whales. Figure 11(a) shows the results of whale A, and Figs. 11(b)–11(e) show the results of C, D, E, and F, respectively. In these figures, the smoothed depth and ICI with respect to time are described. From these figures, the dominant ICI of these five whales is about 0.4–0.45 ms, which is shorter than has been reported (Jaquet *et al.*, 2001; Zimmer *et al.*, 2005b). It has been reported that sperm whales adjust the ICI during a dive; a longer sonar range at the beginning of the dive and subsequent reduction in ICI when they approach sonar targets, e.g., prey or bottom (Madsen *et al.*, 2002). Sperm whales stay on the surface for breathing and dive for foraging, so at the beginning of the dives, they use the longer sonar range, hence, long ICI for search, and after diving, as they approach the targets, they decrease the ICI. However, from Fig. 11, the suggested adjustment of ICI is not well observed except in the time frame of 550–600 s from Fig. 11(c) of whale D. In other situations such as in Fig. 11(a) in the time frame of 800–900 s, Fig. 11(c) of 1100 s, and Fig. 11(d) of 1200 s, the descent may lead to the up and down variations in ICI. The depth of the five whales is almost constant from 300 to 800 m, so the adjustment of ICI from the surface during diving, as indicated by Madsen *et al.* (2002), would not be clearly seen from these results, but it can be seen that they adjust their ICI in short dive. Considering that the ICI of five whales is maintained at about 0.45 ms and the horizontal directions of these

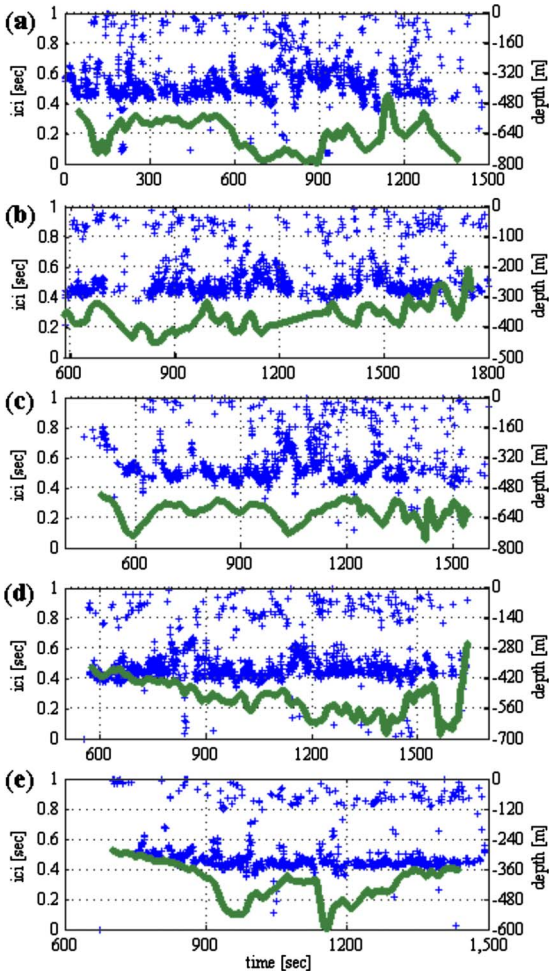


FIG. 11. (Color online) Relationships between the depth and ICI in a group: in these figures, the lines indicate the smoothed depth, and the cross markers indicate the ICI. The top panel (a) shows the results of whale A, and panels (b)–(e) show the results of whale C, D, E, and F, respectively.

whales are similar and smoothly spread, as shown Fig. 8, it may be hypothesized that they collaborate with each other in a group in search of targets.

D. Future work and issues

There are certain issues which should be carefully considered in the proposed algorithm, as shown in Sec. III. Some issues concerned with the process of classification and class matching are discussed below.

1. Classification

The proposed classification algorithm, as shown in Fig. 4, classifies the click based on its direction vector of arrival. The click is classified to belong to the class, whose angular difference between the compared click and the class lies within the classification threshold β_θ. If noise gets detected, then it too can inadvertently get classified to belong to the class if the calculated direction vector of noise is closer to that class. Thus, the direction vector of the class shows the outlier values because of noise. Figures 6(c) and 7(c) show an example of this problem. In Fig. 6(c), in the time frame of 150 s and bearing angle of 140° – 180°, there is a line caused by a sperm whale and also a cluster caused by noise. However, after the classification, as shown in the circle of Fig. 7(c), the line and the cluster get connected with a common line, which deflects it from the actual whale, as shown in Fig. 6(c).

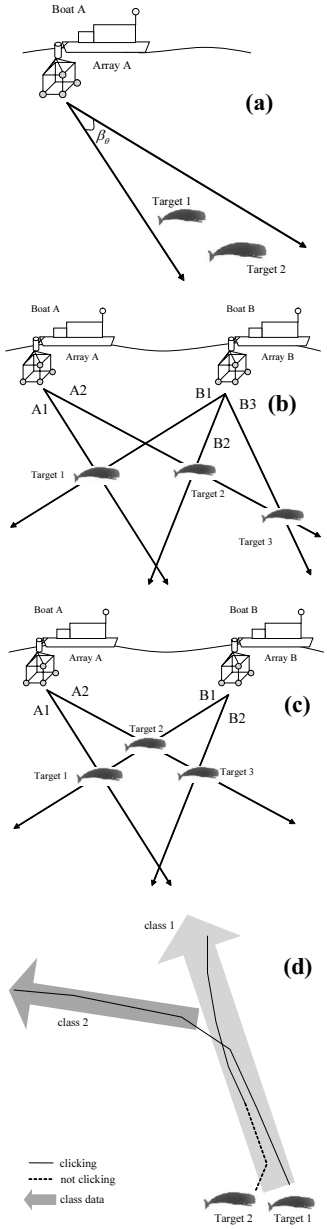


FIG. 12. Misclassification: panel (a) shows the example of misclassification; panel (b) shows the example of mislocalization; panel (c) shows yet another example; panel (d) shows the example of a class constructed by the different whales.

The proposed classification has another issue, as demonstrated by examples in Fig. 12. In Fig. 12(a), there are two whales, targets 1 and 2, detected by array A. In case the angular difference between two whales is within the angular threshold β_θ, these two whales are classified to belong to the same class. Hence, in the proposed classification, the same class may be constituted by multiple whales. To avoid these two types of misclassification, some discriminating parameter obtained by analyzing the sperm whale's click should be considered after the classification.

TABLE III. Example of misclassification.

	Class of A	Class of B	Position
	Matching cases of Fig. 12(b)		
Case 1	A1 (target 1)	B1 (target 1)	Target 1
	A2 (targets 2 and 3)	B2 (target 2)	Target 2
Case 2	A1 (target 1)	B1 (target 1)	Target 1
	A2 (targets 2 and 3)	B3 (target 3)	Target 3
	Matching cases of Fig. 12(c)		
Case 3	A1 (target 1)	B1 (targets 1 and 2)	Target 1
	A2 (targets 2 and 3)	B2 (target 3)	Target 3
Case 4	A2 (targets 2 and 3)	B1 (targets 1 and 2)	Target 2
	A1 (target 1)	B2 (target 3)	Miss

2. Class matching

Another issue may occur in the localization of class matching because of the misclassification, as shown in Fig. 12(b), where there are three whales, targets 1, 2, and 3, and two arrays, A and B. So each array should show three classes based on the direction vector of each whale. However, since the direction vectors to targets 2 and 3 from array A are very close, only two classes are constructed in array A; class A1 of target 1 and class A2 of targets 2 and 3, even though three classes are constructed in array B. So number of classes in array A is not equal to the number of classes of B because of misclassification. The matching cases are shown in Table III. For Fig. 12(b) in case 1, the position of target 1 is calculated using the classes A1 and B1, and target 2 is calculated with A2 and B2, but the position of target 3 is not calculated. In the other case 2, A1 and B1 are matched, and A2 and B3 are matched, but the position of target 2 is not calculated.

Figure 12(c) shows another example of misclassification due to class matching. There are three whales, targets 1, 2, and 3, and two arrays, A and B. In array A, the directions to targets 2 and 3 are very close to be classified as separated classes, while in array B, the directions of targets 1 and 2 are similar as well. Thus, only two classes are constructed in each array; class A1 of target 1 and class A2 of targets 2 and 3 in array A, and class B1 of targets 1 and 2 and class B2 of target 3. The four candidate positions are calculated in all-class matching analysis but only two matching results are extracted since there are only two classes at each array. Table III shows the matching patterns. In case 3, the position of target 1 is calculated using the classes of A1 and B1, and target 3 is calculated with A2 and B2, but the position of target 3 is not calculated. In another case 4, A2 and B1 are matched as target 2, and A1 and B2 are mismatched. The direction vector based classification is applicable, as shown in Figs. 6 and 7, but it is necessary to improve the classification algorithm to estimate such misclassifications.

3. Manual processing

The same relative arrival time pattern from a particular whale used in the class matching may not be observed as in the following cases; the sperm whale swims very far away from the array, or too many whales are clicking and diving simultaneously around. If a sperm whale swims too far away from the arrays, the changes in direction pattern may appear too small to determine the matched class.

The clicks are classified based on the arrival direction vector, which implies that a class may contain different whales' clicks, as shown in Fig. 12(d). From Fig. 12(d), if a sperm whale, target 2, begins to click very close to another whale, target 1, and then moves to different directions, two classes are constructed; class 1 constructed by the clicks of targets 1 and 2, and class 2 constructed by the clicks of target 1. Thus, it is necessary to check the classes' arrival time pattern very carefully to find the matching classes by manual proceedings. However, it is necessary to develop an autonomous matching algorithm as a future work.

4. Ground-truthing the system

The estimation of localization error is described before, but it is necessary to empirically ground-truth the system. Since visual observation for ground-truthing will be limited to surface observations, we intend to use other methods. The methods for ground-truthing that we plan to carry out are either by using a tag for independent measurement of the whale's depth or to use an AUV with a sound transmitter. The AUV can use its sensors to estimate its own position during the dive, while the arrays on the surface capture the sound transmitted from the AUV and estimate the position of the AUV in order to evaluate the localization accuracy of the system for ground-truthing.

VI. CONCLUSIONS

A methodology for sperm whale localization using two separated hydrophone arrays has been proposed. The method has been applied to actual sperm whale click data recorded off Ogasawara islands in 2003. The underwater trajectories of six sperm whales simultaneously moving in a group were successfully obtained after matching the whale classes recorded on the two arrays. With the results of the relationships between the depth and ICI, the possibility that the sperm whales in a group use short ICI to search for targets in collaboration with each other can be pointed out. And in order to let an AUV follow and investigate the underwater behavior of a specific sperm whale in real-time, it is necessary to obtain the position data of the individual targeted whale. This is now considered to be achievable by application of the proposed method. To achieve this goal in real-time, the whale classes between the two arrays should be matched automatically, considering the changes in the direction vectors. The next step would be to develop a fully autonomous real-time matching algorithm between two separated arrays.

ACKNOWLEDGMENTS

The authors would like to thank Ogasawara Whale Watching Association, Dr. Mori Kyoichi, System Giken, and Mr. Tetsuo Fukuchi. The project is supported by the Japan Society for Promotion of Science, Japan Science, and the author was supported by the Technology Agency and "Am-

bient SoC Global COE Program of Waseda University" of the Ministry of Education, Culture, Sports, Science and Technology, Japan.

Amano, M., and Yoshioka, M. (**2003**). "Sperm whale diving behavior monitored using a suction-cup-attached TDR tag," Mar. Ecol.: Prog. Ser. **258**, 291–295.

Aoki, K., Amano, M., Sugiyama, N., Muramoto, H., Suzuki, M., Yoshioka, M., Mori, K., Tokuda, D., and Miyazaki, N. (**2007b**). "Measurement of swimming speed in sperm whales," in Proceedings of the Underwater Technology '07, Tokyo, Japan, pp. 467–471.

Aoki, K., Amano, M., Yoshioka, M., Mori, K., Tokuda, D., and Miyazaki, N. (**2007a**). "Diel diving behavior of sperm whales off Japan," Mar. Ecol.: Prog. Ser. **349**, 277–287.

Bahl, R., Ura, T., and Fukuchi, T. (**2003**). "Techniques for segregation and classification of several vocalizing sperm whales," in Proceedings of the Oceans '03, San Diego, CA, pp. 457–463.

Goold, J. C. (**1996**). "Signal processing techniques for acoustic measurement of sperm whale body length," J. Acoust. Soc. Am. **100**, 3431–3441.

Hooker, B. M., McConnella, B. J., Millera, P. J. O., and Sparlinga, C. E. (**2007**). "Bio-loggings science: Logging and relaying physical and biological data using animal-attached tags," Deep-Sea Res., Part II **54**, 177–182.

Jaquet, N., Dawson, S., and Douglas, L. (**2001**). "Vocal behavior of male sperm whales: Why do they click?," J. Acoust. Soc. Am. **109**, 2254–2259.

Johnson, M., and Tyack, P. (**2003**). "A digital acoustic recording tag for measuring the response of wild marine mammals to sound," IEEE J. Ocean. Eng. **28**, 3–12.

Leatherwood, S., and Reaves, R. (**1983**). *The Sierra Club Handbook of Whales and Dolphins* (Sierra Club Books, San Francisco, CA).

Levenson, C. (**1974**). "Source level bistatic target strength of the sperm whale, Physeter catodon, measured from an oceanographic aircraft," J. Acoust. Soc. Am. **55**, 1100–1103.

Madsen, P. T., Payne, R., Kristiansen, U. K., Wahlberg, M., Kerr, I., and Møhl, B. (**2002**). "Sperm whale sound production studied with ultrasound time/depth-recording," J. Exp. Biol. **205**, 1899–1906.

Møhl, B., Wahlberg, M., and Heerfordt, A. (**2001**). "A large-aperture array of nonlinked receivers for acoustic positing of biological sound sources," J. Acoust. Soc. Am. **109**, 434–437.

Møhl, B., Wahlberg, M., Madsen, P. T., Miller, L. A., and Surlykke, A. (**2000**). "Sperm whale clicks: Directionality and source level revisited," J. Acoust. Soc. Am. **107**, 638–648.

Nosal, E., and Franzer, L. (**2007**). "Sperm whale three-dimensional track, swim orientation, beam pattern, and click levels observed on bottom-mounted hydrophones," J. Acoust. Soc. Am. **122**, 1969–1987.

Spiesberger, J. L., and Fristup, K. M. (**1990**). "Passive localization of calling animals and sensing of their acoustic environment using acoustic tomography," Am. Nat. **135**, 107–153.

Teloni, V., Johnson, M., Patrickc, M., and Peter, M. (**2008**). "Shallow food for deep divers: Dynamic foraging behavior of male sperm whales in a high latitude habitat," J. Exp. Mar. Biol. Ecol. **354**, 119–131.

Thode, A. (**2004**). "Tracking sperm whale (*Physeter macrocephalus*) dive profiles using a towed passive acoustic array," J. Acoust. Soc. Am. **116**, 245–253.

Tiemann, C., Thode, A., Straley, J., and O'Connell, V. (**2006**). "Three-dimensional localization of sperm whales using a single hydrophone," J. Acoust. Soc. Am. **120**, 2355–2365.

Ura, T., Bahl, R., Sakata, M., Kojima, J., Fukuchi, T., Ura, J., Mori, K., Nakatani, T., Nose, K., and Sugimatsu, H. (**2004a**). "Development of AUV-based system for acoustic tracking of diving sperm whales," in Proceedings of the Oceans '04, Kobe, Japan, pp. 2302–2307.

Ura, T., Sakata, M., Bahl, R., Kojima, J., Fukuchi, T., Ura, J., Nose, M., Sugimatsu, H., Mori, K., Nakatani, T., and Yanagisawa, M. (**2004b**). "Acoustic tracking of sperm whales using two sets of hydrophone array," in Proceedings of the Underwater Technology '04, Taipei, Taiwan, pp. 103–107.

Wahlberg, M. (**2002**). "The acoustic behavior of diving sperm whales observed with a hydrophone array," J. Exp. Mar. Biol. Ecol. **281**, 53–62.

Wahlberg, M., Møhl, B., and Madesen, P. T. (**2001**). "Estimating source position accuracy of a large-aperture hydrophone array for bioacoustics," J. Acoust. Soc. Am. **109**, 397–406.

Watkins, W. A., Daher, M. A., Fristrup, K. M., and Howald, T. J. (**1993**). "Sperm whales tagged with transponders and tracked underwater by sonar," Marine Mammal Sci. **9**, 55–67.

Watkins, W. A., and Schevill, W. E. (**1977**). "Sperm whale codas," J. Acoust. Soc. Am. **62**, 1485–1490.

Zimmer, W. M. X., Johnson, M. P., D'Amico, A., and Tyack, P. L. (**2003**). "Combining data from a multisensor tag and passive sonar to determine the diving behavior of a sperm whale," IEEE J. Ocean. Eng. **28**, 13–28.

Zimmer, W. M. X., Madsen, P. T., Teloni, V., Johnson, M. P., and Tyack, P. L. (**2005a**). "Off-axis effects on the multipulse structure of sperm whale usual clicks with implications for sound production," J. Acoust. Soc. Am. **118**, 3337–3345.

Zimmer, W. M. X., Tyack, P. L., Johnson, M. P., and Madsen, P. T. (**2005b**). "Three-dimensional beam pattern of regular sperm whale clicks confirms bent-horn hypothesis," J. Acoust. Soc. Am. **117**, 1473–1485.

Experimental validation of a time domain simulation of high frequency ultrasonic propagation in a suspension of rigid particles

Belfor Galaz
Laboratoire d'Imagerie Paramétrique, UPMC Univ Paris 6, and CNRS, LIP, UMR 7623, 15 rue de l'École de Médecine, 75006 Paris, France; and Department of Physics, Universidad de Santiago de Chile (USACH), Ecuador 3493, Santiago, Chile

Guillaume Haïat[a)]
CNRS, Laboratoire de Recherches Orthopédiques, Université Paris 7, UMR CNRS 7052 B2OA, 10 Avenue de Verdun, 75010 Paris, France

Romain Berti and Nicolas Taulier
Laboratoire d'Imagerie Paramétrique, UPMC Paris 6, and CNRS, LIP, UMR 7623, 15 rue de l'École de Médecine, 75006 Paris, France

Jean-Jacques Amman
Department of Physics, Universidad de Santiago de Chile (USACH), Ecuador 3493, Santiago, Chile

Wladimir Urbach
Laboratoire d'Imagerie Paramétrique, UPMC Paris 6, and CNRS, LIP, UMR 7623, 15 rue de l'École de Médecine, 75006 Paris, France; and Laboratoire de Physique Statistique de l'École Normale Supérieure de Paris, CNRS UMR 8550, 24 rue Lhomond, 75005 Paris, France

(Received 2 October 2008; revised 7 November 2009; accepted 10 November 2009)

Ultrasonic propagation in suspensions of particles is a difficult problem due to the random spatial distribution of the particles. Two-dimensional finite-difference time domain simulations of ultrasonic propagation in suspensions of polystyrene 5.3 μm diameter microdisks are performed at about 50 MHz. The numerical results are compared with the Faran model, considering an isolated microdisk, leading to a maximum difference of 15% between the scattering cross-section values obtained analytically and numerically. Experiments are performed with suspensions in through transmission and backscattering modes. The attenuation coefficient at 50 MHz (α), the ultrasonic velocity (V), and the relative backscattered intensity (I_B) are measured for concentrations from 2 to 25 mg/ml, obtained by modifying the number of particles. Each experimental ultrasonic parameter is compared to numerical results obtained by averaging the results derived from 15 spatial distributions of microdisks. α increases with the concentration from 1 to 17 dB/cm. I_B increases with concentration from 2 to 16 dB. The variation of V versus concentration is compared with the numerical results, as well as with an effective medium model. A good agreement is found between experimental and numerical results (the larger discrepancy is found for α with a difference lower than 2.1 dB/cm). © *2010 Acoustical Society of America.* [DOI: 10.1121/1.3270399]

PACS number(s): 43.35.Bf, 43.20.Hq, 43.20.Px, 43.35.Cg [CCC] Pages: 148–154

I. INTRODUCTION

The propagation of acoustic waves through particles dispersed in a fluid (suspension) is relevant to many applications, including the ultrasonic characterization of biological tissues such as blood (Haider *et al.*, 2000) or ultrasound contrast agents (Bleeker *et al.*, 1990) of emulsion (McClements, 1992), and to the acoustic propagation in underwater acoustics (for example, sandy suspensions) (Thorne and Campbell, 1992; Thorne *et al.*, 1993; Thorne and Buckingham, 2004). In addition, suspensions of elastic spheres have been proposed as acoustic calibration standards (Rhyne *et al.*, 1986). However, predicting the acoustic wave behavior of such systems remains difficult, due to possible multiple scattering phenomena between particles, mode conversion, and apparent disorder of the particles' spatial distribution. The propagation of sound waves through suspensions of different natures has been the subject of many studies since the pioneer work of Tindall (1875), Rayleigh (1872), and Sewell (1910). A comprehensive review of wave propagation in suspensions has been written by Temkin (2001).

Briefly, suspensions have been treated by some authors as a two-phase fluid (Temkin and Dobbins, 1966; Harker and Temple, 1988; Atkinson and Kytomaa, 1992), while others (Hovem, 1980; Ogushwitz, 1985; Gibson and Nafi Toksoz, 1989) have modeled suspensions as a porous solid whose rigidity can be varied by means of adjustable parameters,

[a)]Author to whom correspondence should be addressed. Electronic mail: guillaume.haiat@univ-paris-diderot.fr

which is a model appropriate for media where a solid skeleton exists, but seems limited for suspensions of free particles such as those treated here.

The scattering theory, also called Epstein, Carhart, Allegra, and Hawley (ECAH) theory, is based on the work of Epstein and Carhart (1953) and Allegra and Hawley (1972), which has been extended by others (Hay and Mercer, 1985; Hay and Schaafsma, 1989). The ECAH theory was originally developed, considering a superposition of each particle contribution, and therefore did not consider multiple scattering. In consequence, several different approaches have been followed to modify the ECAH theory in order to incorporate multiple scattering (see, for example, Foldy, 1945; Waterman and Truell, 1961; Berryman, 1980a, 1980b; Mobley *et al.*, 1999). The coupled phase theory (Harker and Temple, 1988; Evans and Attenborough, 1997; Baudoin *et al.*, 2007) is another but powerful approach based on the two-phase hydrodynamic equations. The main difference between the coupled phase and the scattering approaches is that coupled phase theory is self-consistent. The self-consistency is generated by the use of volume averaged field variables (Margulies and Schwartz, 1994). Finally, Temkin (1998, 2000) also developed a theoretical framework applicable to a wide frequency range to model the propagation in dilute suspensions.

Time domain numerical simulation tools have not been applied to model ultrasound propagation in such complex heterogeneous media. A potential advantage of a numerical approach is its ability to solve complex problems that may become rapidly intractable when following purely analytical approaches in the frequency domain. Another advantage of time domain numerical approaches is that it allows simulating rf signals directly in the time domain, avoiding reconstructions from the frequency domain (Insana *et al.*, 1990; Doyle, 2006). Working in the time domain is interesting because (i) it allows a better comparison with the experimental signals, which are obtained in the time domain, and (ii) ultrasonic velocities measured in the time domain using different markers (such as the first zero crossing velocity) have been shown to be adapted for velocity measurements in a dispersive media such as bone (Haiat *et al.*, 2006).

The aim of this work is to examine the problem of wave propagation at high frequency (50 MHz) in suspensions of solid particles immersed in water using two-dimensional (2D) finite-difference time domain (2D FDTD) numerical simulation tools. The main advantage of this numerical method is to deal with large populations of particles, allowing to simulate multiple scattering effects in the time domain. The originality of the present approach is to account for the spatial distribution of the particles to compute the ultrasonic response of the suspension. Specifically, the 2D FDTD simulation code is validated by comparing the results with (i) analytical results obtained from the Faran theory and (ii) experimental results obtained with a solution of latex polystyrene microspheres of 5.3 μm diameter.

II. MATERIALS AND METHODS

A. Two-dimensional numerical modeling

2D numerical simulations of ultrasonic wave propagation through randomly distributed microdisk solutions are performed using SIMSONIC, a FDTD simulation software. This software has been developed by the Laboratoire d'Imagerie Paramétrique, and its description and validation can be found elsewhere (Bossy *et al.*, 2004, 2005; Haiat *et al.*, 2007). Briefly, it uses an algorithm based on the spatial and temporal discretizations of the two following coupled first-order equations describing a 2D linear elastic wave propagation (Bossy *et al.*, 2004):

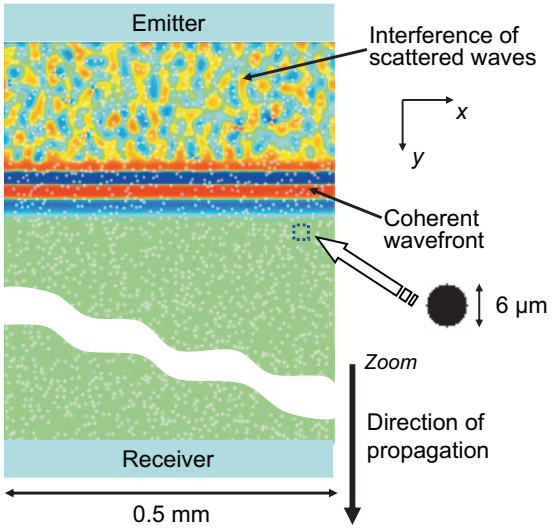


FIG. 1. Image illustrating the simulated propagation of ultrasonic wave in a suspension of disk-shaped polystyrene particles. The curved separation represents an arbitrary split of the simulation domain. In the upper part of the figure, the color codes the amplitude of the displacement as a function of position (direction y is parallel to the direction of propagation and x is perpendicular) at a given time. The coherent wavefront may be distinguished in red below the complicated wave field corresponding to interference of waves scattered by the particles. In the lower part, the figure displays the random distribution of the particles. On the right hand side, an isolated particle is shown where black pixels correspond to elastic polystyrene.

$$\frac{\partial v_i}{\partial t} = \frac{1}{\rho(\vec{r})}\frac{\partial \sigma_{ij}}{\partial r_j},$$

$$\frac{\partial \sigma_{ij}}{\partial t} = C_{ijkl}\frac{\partial v_k}{\partial r_l}, \tag{1}$$

where $\vec{r}$ is the 2D position vector, $\vec{v}$ is the displacement velocity, σ is the stress tensor, and C is the stiffness tensor. The discretization is performed following the "de Virieux" (Virieux, 1986; Graves, 1996) resolution scheme with a time step, which is automatically deduced from the required stability condition described in Virieux, 1986 and Graves, 1996. The main assumptions of the model are as follows: (i) All absorption phenomena are neglected, (ii) all heat transfer phenomena are neglected, and (iii) the temperature field is assumed to be constant and homogeneous. However, the model fully takes into account all reflection and refraction effects, as well as mode conversions.

The ultrasonic propagation is simulated in a 0.5 $\times$ 1 mm^2 rectangular domain, as shown in Fig. 1. The longest length is along the y-axis, which is also the direction of the propagation. A linear ultrasound pressure source of 0.5 mm length located at $y=0$ emits a broadband pressure pulse

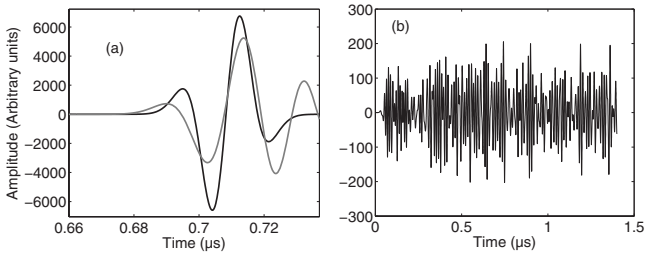


FIG. 2. Typical simulated rf signals: (a) The signal in water (in black) is identical in shape to the signal generated by the emitter. In gray: the rf signal transmitted in suspension made of 5.3 μm disk-shaped polystyrene particles immersed in water. The particle concentration is $C=13.4$ mg/ml. The value of the velocity (respectively, attenuation coefficient at 50 MHz) is equal to 1505 m/s (respectively, 4.46 dB/cm). (b) Backscattered rf signal obtained with the same suspension. The corresponding value of the relative backscattering intensity is equal to 10.6 dB.

shown in black in Fig. 2(a), with a center frequency equal to 50 MHz (bandwidth 35–62 MHz at −3 dB). To ensure plane wave propagation, symmetric boundary conditions are applied to the box sides located at $x=0$ and 0.5 mm, parallel to the y-axis. In order to avoid unphysical reflections due to the boundaries of the simulation mesh, perfectly matched layers (Collino and Tsogka, 2001) are positioned at $y=0$ and 1 mm.

Two linear receivers located at $y=0$ and 1 mm provide both backscattered and transmitted signals through a spatial average of the displacement over the entire transducer width (i.e., 0.5 mm). The chosen pixel size is equal to 0.25 μm, which is a compromise between an acceptable spatial resolution and a reasonable computational time.

Similarly to what was done previously (Bossy *et al.*, 2005; Haiat *et al.*, 2007, 2008), the numerical simulation was performed considering a plane wave propagation because this situation mimics what happens at the focus of a transducer such as the one used in the experiments.

An iterative probabilistic procedure was used to randomly insert N identical disk-shaped particles in the two-dimensional blank domain. During this procedure, the insertion of a particle is accepted only if the surface area of the particle is entirely included within the domain and if the particle surface area does not intercept the surface area of another previously inserted particle. If these conditions are not fulfilled, another location is randomly selected. This process is iterated until N particles are finally inserted in the domain.

We use this procedure to insert homogeneous polystyrene microparticles of 5.3 μm diameter. To define the region of the grid corresponding to such a homogeneous disk of radius R, having its center located at (x_0, y_0), all pixels having their center at a distance lower than R were identified and their material property was assigned to that of polystyrene. The right part of Fig. 1 shows the inserted disk-shaped particle obtained by this procedure, where the black pixels correspond to polystyrene. Figure 1 also shows an image corresponding to the ultrasonic propagation in this heterogeneous medium where the color codes the amplitude of the displacement as a function of space.

FDTD simulations require as input parameters mass densities and stiffness coefficients of all materials used in the simulation. One major difficulty is to find the most accurate parameters since they will influence the quality of the simulation predictions. We could not measure the transverse and longitudinal wave speeds in our polystyrene spheres; therefore, we started our simulations with values for polystyrene found in literature (Adjadj *et al.*, 2003; Wear, 2005). Then, we considered a variation of the longitudinal velocity in polystyrene ($V_{L,\mathrm{PS}}$) in order to better fit our experimental data. Best agreement was achieved for a value of $V_{L,\mathrm{PS}}$ 20%, lower than the value found in the literature. In consequence, all simulations were performed with $V_{L,\mathrm{PS}}=2000$ m/s and $V_{T,\mathrm{PS}}=1155$ m/s, the value of $V_{T,\mathrm{PS}}$ corresponding to Poisson's ratio equal to 0.25.

B. Experimental measurements

Measurements were performed in solutions of microspheres of 5.3 μm diameter, made of latex polystyrene (Corpuscular Inc., Cold Spring, NY). The solution was stabilized at 25.0 ± 0.1 °C using an external control temperature system. The ultrasonic pulses were generated by transducers coupled with a 33250A Agilent pulse generator (Santa Clara, CA). The emitted and received signals were digitalized at a frequency of 200 MHz by a Lecroy oscilloscope (Chestnut Ridge, NY).

Backscattering measurements were performed using a 3 mm deep cylindrical aluminum cell filled with 1.5 ml of microsphere solution. The exposure chamber has an internal diameter of 2.5 cm. A polyvinyl chloride (PVC) film is placed on top of the solution to hermitically separate it from water added on top of the film. Agitation was maintained inside the solution; thanks to a small off-centered stirring bar. A 50 MHz polyvinylidene fluoride broadband Panametrics (model PI57-1) transducer (15.8 mm focal length and a lateral resolution of 0.9 mm at the focal point) was immersed in water and focused using an acoustic lens in the polystyrene solution at approximately 1 mm below the PVC film to avoid any contribution coming from the film. A series of sinusoidal pulses of 0.2 μs was sent by the amplifier to the transducer every 10 ms, leading to a broadband ultrasonic pulse of center frequency of 50 MHz, generated in the solution. Signals diffused by microspheres in solution were collected by the same transducer.

To measure attenuation and ultrasonic velocity, classical transmission measurements were performed using planar transducers similar to those previously described for the receiver and emitter, except that no acoustical lenses were used. A time marker (the time of the first zero crossing) was used for the reference signal and the signal transmitted through the suspension in order to derive the ultrasonic velocity. This time of flight method is described in more detail in Haiat *et al.*, 2005, 2006. The emitter and receiver were coaxially aligned and operated in transmission. They were immersed in the microsphere solution and separated by a distance of 6.5 mm. Each rf signal was estimated by performing an average of over 100 signals in order to reduce the effect of noise. Knowing the distance between the two transducers from preliminary measurements (using distilled water, which has a known sound velocity), the speed of sound was estimated with a reproducibility better than 0.1 m/s.

C. Comparison between numerical approach and experiments

In order to compare 2D simulations with three-dimensional (3D) experiments, a 2D concentration equivalent to the experimental one needed to be determined. We chose a method based on the mean distance between a particle and its closest neighbor since it is a meaningful parameter in two and three dimensions. For N randomly distributed particles within a sufficiently large 2D domain of surface area S, the mean distance D_2 between the center of one particle and the center of its closest neighbor is given by

$$D_2 = \sqrt{\frac{S}{N}}. \tag{2}$$

In the case of a 3D distribution, the inter-particle distance is given by

$$D_3 = \sqrt[3]{\frac{M}{C}}, \tag{3}$$

where M is the mass of a particle and C is the particle concentration in the solution (in mg/ml). The assumption used herein is that the distance between two particles must be the same in two and three dimensions; i.e., $D_2 = D_3$. This leads to the following relation between N and C:

$$C = M\left(\frac{N}{S}\right)^{3/2}, \tag{4}$$

which is used to relate N to the corresponding value of the concentration in three dimensions. The mass density of polystyrene was taken equal to 1.05 g/ml, which leads to a value of M equal to 8.18×10^{-11} g for a 5.3 μm diameter particle.

D. Determination of the ultrasonic parameters

Basic signal processing techniques were applied to the simulated and experimental rf signals to retrieve the ultrasound properties of the 2D domain. An example of simulated rf signal obtained in transmission and in backscattering is shown in Figs. 2(a) and 2(b), respectively. Three parameters were extracted from the simulations and experimental measurements: the speed of sound V, the attenuation coefficient $\alpha(f)$, and the apparent backscattered coefficient $ABC(f)$. Since absorption is neglected in simulations, the attenuation obtained *in silico* is only due to scattering phenomena, and the corresponding $\alpha(f)$ value, in decibel, is given by D'Astous and Foster (1986) as follows:

$$\alpha(f) = \frac{20}{L}\log\left(\frac{S_{\text{ref}}(f)}{S_{\text{sol}}(f)}\right), \tag{5}$$

where L is the path length of sound propagation. In polystyrene microparticule solutions, $L \sim 6.5$ mm and *in silico*, $L = 1$ mm. $S_{\text{ref}}(f)$ and $S_{\text{sol}}(f)$ are, respectively, the power spectrum densities (PSDs) of the signal transmitted in water and in the solution, obtained by using a fast Fourier transform. The apparent backscattered coefficient $ABC(f)$, expressed in decibel, is given by Chaffaï *et al.* (2000) as follows:

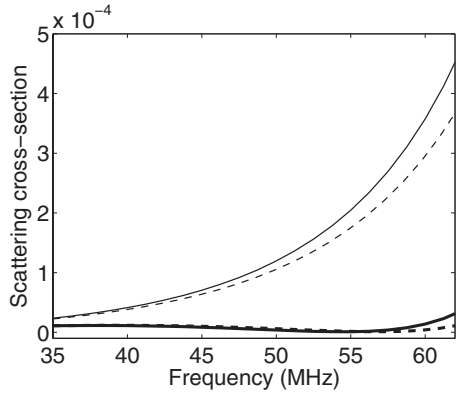


FIG. 3. Scattering cross-section obtained with the Faran model (dashed lines) and the numerical simulation (solid lines) at angles of 180° (thick lines) and 90° (thin lines) from the incident beam.

$$ABC(f) = 20\log\left(\frac{S_b(f)}{S_{\text{ref}}(f)}\right), \tag{6}$$

where $S_b(f)$ is the PSD of the backscattered signal recorded by the transducer located at $y=0$. The relative backscattered intensity I_B (which does not depend on the frequency) is then computed by averaging the apparent backscattered coefficient $ABC(f)$ over the frequency bandwidth of interest (35–62 MHz). The experimental precision on $\alpha(f)$ and I_B is equal, respectively, to 4 dB/cm and 1 dB.

III. RESULTS AND DISCUSSION

In order to validate our numerical approach in the framework of particle suspensions, we first carried out simulations in a 2D domain containing a single homogenous polystyrene disk with a diameter equal to 6 μm. This kind of system has been analytically solved by Faran (1951) for any angle of observation and of incidence. Figure 3 shows the comparison between the scattering cross-section (at 180° and 90°) obtained (i) with the analytical Faran model applied to a lossless isotropic elastic cylinder immersed in water and (ii) by computing the square of the amplitude ratio of the finite-difference time domain numerically simulated spectra of the scattered signal to the incident signal. The discrepancy between the analytical and simulated results, which increases up to 15% as the frequency reaches 62 MHz, is due to the (constant) spatial discretization (0.25 μm) used in the simulation code and to the fact that the scatterers are not strictly circular, as shown in Fig. 1. Decreasing the pixel size (down to 0.1 μm) leads to a decrease in the discrepancy between analytical and numerical results down to 5%. However, the choice of the value of the resolution equal to 0.25 μm corresponds to a compromise between reasonable computation time and memory requirements and an acceptable discrepancy between analytical and numerical results.

Our investigation was extended to the case of a distribution of similar homogenous polymer disks. For this case, to the best of our knowledge, there is no analytical model available in the time domain, so we confronted our simulation results with experimental values measured on solutions of polystyrene microspheres of the same diameter.

TABLE I. Number N of particles accounted for in the simulation domain and the corresponding concentration.

N	250	500	1000	1500	2000	2500
C (mg/ml)	0.92	2.59	7.3	13.4	20.7	28.9

The ultrasonic properties (ultrasonic velocity, attenuation, and relative backscattered intensity) of a given solution constructed using the procedure described in Sec. II A strongly depend on the random distribution of particles, which means that the ultrasonic parameters may differ for two different solutions constructed using the same parameters. For each concentration, 15 different solutions corresponding to 15 cases with a different random placement of particles were considered and the ultrasonic parameters were averaged over the 15 solutions. This value (15) corresponds to a compromise between a reasonable computation time and an acceptable convergence of the ultrasonic parameters. An increase in the number of simulated solutions from 15 up to 25 induces a relative change in the averaged attenuation coefficient equal to 9% of the averaged signal to noise ratio of 3% and no change in the averaged velocity. All these three values are significantly lower than the standard deviation of the results, which corresponds to the variation, due to the random spatial distribution of the particles in the simulation domain.

We have investigated the effect of concentration on the speed of sound, attenuation, and backscattered intensity. Eight concentrations, ranging from 0.9 to 28.9 mg/ml, were used herein and are listed in Table I, with the corresponding values of N obtained from Eq. (4).

Ultrasonic attenuation results from a combination of scattering and absorption phenomena. The complete characterization of absorption requires the knowledge of a large number of thermophysical parameters that are, in practice, hard to quantify. That is why we do not account for absorption in the simulation. Even so, a good agreement is obtained in Fig. 4(a) between simulated and measured values of the attenuation coefficient at 50 MHz, as the averaged difference between experimental and numerical results is equal to 2.1 dB/cm.

In the experiments, the relative backscattered intensity is normalized by the electronic noise level, which is not possible in the simulations where no noise is present. Therefore, a modified backscattered intensity was defined in the simulation. We added 38.5 dB to the simulated relative backscattered intensity, so that the extrapolated value at zero concentration in microspheres falls down to zero, as shown in Fig. 4(b). Therefore, the results shown in Fig. 4(b) only compare the variation of backscattered energy versus the concentration of particles and not its absolute value.

The variation of the speed of sound (V) versus concentration is displayed in Fig. 4(c). An increase in V is obtained for both experimental and simulated values as the particle concentration increases. The velocity of the ultrasonic waves is influenced by the material properties and the density of the particles contained in the solution and to the amount of various phases present. In order to get further insight on the behavior of the velocity as a function of the concentration, an

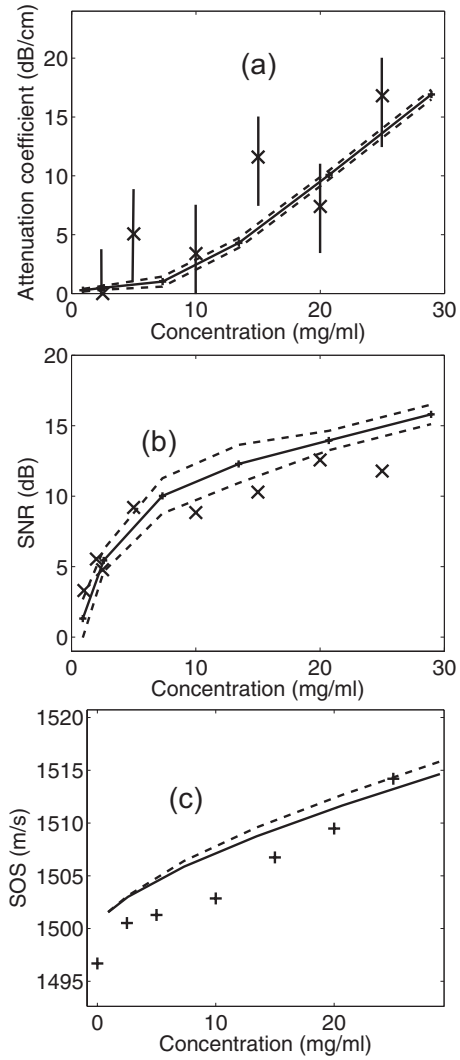


FIG. 4. (a) Mean attenuation coefficient at 50 MHz, (b) signal to noise ratio, and (c) speed of sound as a function of concentration of polystyrene microspheres. The black solid lines correspond to numerical results. The two dashed lines in (a) and (b) represent the sum and the subtraction of the mean and of the standard deviation of each quantity. The dashed line in (c) indicates the results obtained with an effective medium model (Yang and Mal, 1994). The crosses correspond to experimental results. The vertical lines in (a) indicate the experimental standard deviations. The experimental standard deviation in (b) and (c) is equal to 1 dB and 0.1 m/s, respectively.

effective medium model (Yang and Mal, 1994) was applied to the configuration of interest. This 2D model considers elastic circular scatterers immersed in an elastic matrix and takes into account multiple scattering effects. Similarly as what was done in Haïat *et al.*, 2008 in another context, we considered a value of shear wave velocity Vs in water equal to 5 m/s. This value was chosen arbitrarily and we verified that decreasing the value of Vs down to 0.1 m/s does not modify the phase velocity, which indicates that shear wave modes propagating in the matrix do not significantly affect the phase velocity. Readers are referred to Haïat *et al.*, 2008 for further details on the approach considered. A good agreement is obtained between the effective medium model and the numerical model, as the difference between the two results is lower than 1 m/s. This last result constitutes an additional further validation of the numerical model.

The advantage of the approach using FDTD simulations is to be able to provide a reasonable estimation of ultrasound properties of suspensions based on the chemical and physical properties of the materials. However, there are several limitations to the approach carried out in the present study. First, we use a 2D model, and 3D models would be more accurate to model the propagation in suspensions, which is possible using the SIMSONIC software (Bossy *et al.*, 2005; Haïat *et al.*, 2007). The propagation in a 2D medium of disks is different than in a 3D medium of spheres, which might explain possible discrepancies between experiments and simulations. However, the present 2D approach is a first step in studying these phenomena, and further studies are required to tackle the difficult 3D problem, which would require significantly longer time of computation and memory. Second, absorption effects are not taken into account and only the part of attenuation that depends on scattering phenomena is considered here. Third, the proposed model is simplistic because we only considered similar particles (monodisperse suspension), whereas they are, in reality, different in terms of diameter. We choose to consider monodisperse suspensions in order to study the effect of concentration independently of the diameter distribution, which corresponds to a first step in the study of such suspensions. Moreover, the precise distribution of the diameter of the polystyrene particles remains unknown and we thus did considered monodispersed suspensions. Fourth, we did not account for diffraction effects, which might modify the results obtained in backscattering since we assume a planar wave propagation in the simulation.

It should be emphasized that this work could have been carried out at any center frequency without any limitation. We choose to work around about 50 MHz because our long term goal is to provide a numerical model in order to study wave propagation through liquid filled ultrasound contrast agent used for high frequency applications such as microcirculation, ophthalmic disease diagnosis and small animal imaging, or biomicroscopy applications, which are currently investigated in the framework of our research project.

ACKNOWLEDGMENTS

Authors acknowledge financial support from Agence Nationale de la Recherche (ANR), ACUVA No. NT05-3_42548 and from EC-FP6-project DiMI, LSHB-CT-2005-512146.

Adjadj, L., Storti, G., and Morbidelli, M. (**2003**). "Ultrasound attenuation in polystyrene latexes," Langmuir **19**, 3953–3957.

Allegra, J. R., and Hawley, S. A. (**1972**). "Attenuation of sound in suspensions and emulsions: Theory and experiments," J. Acoust. Soc. Am. **51**, 1545–1564.

Atkinson, C. M., and Kytomaa, H. K. (**1992**). "Acoustic wave speed and attenuation in suspensions," Int. J. Multiphase Flow **18**, 577–592.

Baudoin, M., Thomas, J. L., Coulouvrat, F., and Lhuillier, D. (**2007**). "An extended coupled phase theory for the sound propagation in polydisperse concentrated suspensions of rigid particles," J. Acoust. Soc. Am. **121**, 3386–3397.

Berryman, J. (**1980a**). "Long-wavelength propagation in composite elastic media I. Spherical inclusions," J. Acoust. Soc. Am. **68**, 1809–1819.

Berryman, J. (**1980b**). "Long-wavelength propagation in composite elastic media II. Ellipsoidal inclusions," J. Acoust. Soc. Am. **68**, 1820–1831.

Bleeker, H. J., Shung, K. K., and Barnhart, J. L. (**1990**). "Ultrasonic characterization of AlbunexR, a new contrast agent," J. Acoust. Soc. Am. **87**, 1792.

Bossy, E., Padilla, F., Peyrin, F., and Laugier, P. (**2005**). "Three-dimensional simulation of ultrasound propagation through trabecular bone structures measured by synchrotron microtomography," Phys. Med. Biol. **50**, 5545–5556.

Bossy, E., Talmant, M., and Laugier, P. (**2004**). "Three-dimensional simulations of ultrasonic axial transmission velocity measurement on cortical bone models," J. Acoust. Soc. Am. **115**, 2314–2324.

Chaffaï, S., Roberjot, V., Peyrin, F., Berger, G., and Laugier, P. (**2000**). "Frequency dependence of ultrasonic backscattering in cancellous bone: Autocorrelation model and experimental results," J. Acoust. Soc. Am. **108**, 2403–2411.

Collino, F., and Tsogka, C. (**2001**). "Application of the PML absorbing layer model to the linear elastodynamic problem in anisotropic heterogeneous media," Geophysics **66**, 294–307.

D'Astous, F. T., and Foster, F. S. (**1986**). "Frequency dependence of ultrasound attenuation and backscatter in breast tissue," Ultrasound Med. Biol. **12**, 795–808.

Doyle, T. E. (**2006**). "Iterative simulation of elastic wave scattering in arbitrary dispersions of spherical particles," J. Acoust. Soc. Am. **119**, 2599–2610.

Epstein, P. S., and Carhart, R. R. (**1953**). "The absorption of sound in suspensions and emulsions. I. Waterfog in air," J. Acoust. Soc. Am. **25**, 553–565.

Evans, J. M., and Attenborough, K. (**1997**). "Coupled phase theory for sound propagation in emulsions," J. Acoust. Soc. Am. **102**, 278–282.

Faran, J. J. (**1951**). "Sound scattering by solid cylinders and spheres," J. Acoust. Soc. Am. **23**, 405–418.

Foldy, L. L. (**1945**). "The multiple scattering of waves. I. General theory of isotropic scattering by randomly distributed scatterers," Phys. Rev. **67**, 107–119.

Gibson, R. L. J., and Nafi Toksoz, M. (**1989**). "Viscous attenuation of acoustic waves in suspensions," J. Acoust. Soc. Am. **85**, 1925–1934.

Graves, R. (**1996**). "Simulating seismic wave propagation in 3D elastic media using staggered-grid finite differences," Bull. Seismol. Soc. Am. **86**, 1091–1106.

Haïat, G., Lhemery, A., Renaud, F., Padilla, F., Laugier, P., and Naili, S. (**2008**). "Velocity dispersion in trabecular bone: Influence of multiple scattering and of absorption," J. Acoust. Soc. Am. **124**, 4047–4058.

Haïat, G., Padilla, F., Barkmann, R., Dencks, S., Moser, U., Gluer, C. C., and Laugier, P. (**2005**). "Optimal prediction of bone mineral density with ultrasonic measurements in excised human femur," Calcif. Tissue Int. **77**, 186–192.

Haïat, G., Padilla, F., Cleveland, R. O., and Laugier, P. (**2006**). "Effects of frequency-dependent attenuation and velocity dispersion on in vitro ultrasound velocity measurements in intact human femur specimens," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **53**, 39–51.

Haïat, G., Padilla, F., Peyrin, F., and Laugier, P. (**2007**). "Variation of ultrasonic parameters with trabecular bone properties: A three-dimensional model simulation," J. Bone Miner. Res. **12**, 12.

Haider, L., Snabre, P., and Boynard, M. (**2000**). "Rheo-acoustical study of the shear disruption of reversible aggregates. Ultrasound scattering from concentrated suspensions of red cell aggregates," J. Acoust. Soc. Am. **107**, 1715–1726.

Harker, A. H., and Temple, J. A. (**1988**). "Velocity and attenuation of ultrasound in suspensions of particles in fluids," J. Phys. D **21**, 1576–1588.

Hay, A., and Mercer, D. (**1985**). "On the theory of sound scattering and viscous absorption in aqueous suspensions at medium and short wavelengths," J. Acoust. Soc. Am. **78**, 1761.

Hay, A., and Schaafsma, A. (**1989**). "Resonance scattering in suspensions," J. Acoust. Soc. Am. **85**, 1124–1138.

Hovem, J. M. (**1980**). "Viscous attenuation of sound in suspensions and high porosity sediments," J. Acoust. Soc. Am. **67**, 1559–1563.

Insana, M. F., Wagner, R. F., Brown, D. G., and Hall, T. J. (**1990**). "Describing small-scale structure in random media using pulse-echo ultrasound," J. Acoust. Soc. Am. **87**, 179–192.

Margulies, T. S., and Schwartz, W. H. (**1994**). "A multiphase continuum theory for sound wave propagation through dilute suspensions of particles," J. Acoust. Soc. Am. **96**, 319–331.

McClements, D. J. (**1992**). "Comparison of multiple scattering theories with experimental measurements in emulsions," J. Acoust. Soc. Am. **91**, 849.

Mobley, J., Waters, K. R., Hall, C. S., Marsh, J. N., Hughes, M. S., Bran-

denburger, G. H., and Miller, J. G. (**1999**). "Measurements and predictions of the phase velocity and attenuation coefficient in suspensions of elastic microspheres," J. Acoust. Soc. Am. **106**, 652–659.
Ogushwitz, P. R. (**1985**). "Applicability of the Biot theory. II. Suspensions," J. Acoust. Soc. Am. **77**, 441–452.
Rayleigh, L. (**1872**). "Investigation of disturbance produced by a spherical obstacle on the waves of sound," Proc. London Math. Soc. **4**, 253–383.
Rhyne, T. L., Sagar, K. B., Wann, S. L., and Haasler, G. (**1986**). "The myocardial signature: Absolute backscatter, cyclical variation, frequency variation, and statistics," Ultrason. Imaging **8**, 107–120.
Sewell, C. J. T. (**1911**). "On the extinction of sound in a viscous atmosphere by small obstacles of cylindrical and spherical form," Philos. Trans. R. Soc. London **210**, 239–270.
Temkin, S. (**1998**). "Sound propagation in dilute suspensions of rigid particles," J. Acoust. Soc. Am. **103**, 838–849.
Temkin, S. (**2000**). "Attenuation and dispersion of sound in dilute suspensions of spherical particles," J. Acoust. Soc. Am. **108**, 126–146.
Temkin, S. (**2001**). *Elements of Acoustics* (American Institute of Physics, Melville, NY).
Temkin, S., and Dobbins, R. A. (**1966**). "Attenuation and dispersion of sound by particulate relaxation processes," J. Acoust. Soc. Am. **40**, 317–324.
Thorne, P., and Buckingham, M. (**2004**). "Measurements of scattering by suspensions of irregularly shaped sand particles and comparison with a single parameter modified sphere model," J. Acoust. Soc. Am. **116**, 2876–2889.
Thorne, P., and Campbell, S. C. (**1992**). "Backscattering by a suspension of spheres," J. Acoust. Soc. Am. **92**, 978–986.
Thorne, P., Manley, C., and Brimelow, J. (**1993**). "Measurements of the form function and total scattering cross section for a suspension of spheres," J. Acoust. Soc. Am. **93**, 243–248.
Tindall, J. (**1875**). *Sound* (Longmans, Green, and Co., London).
Virieux, J. (**1986**). "P-SV wave propagation in heterogeneous media: Velocity-stress finite-difference method," Geophysics **51**, 889–901.
Waterman, P. C., and Truell, R. (**1961**). "Multiple-scattering of waves," J. Math. Phys. **2**, 512–537.
Wear, K. A. (**2005**). "The dependencies of phase velocity and dispersion on trabecular thickness and spacing in trabecular bone-mimicking phantoms," J. Acoust. Soc. Am. **118**, 1186–1192.
Yang, R. B., and Mal, A. K. (**1994**). "Multiple-scattering of elastic waves in a fiber-reinforced composite," J. Mech. Phys. Solids **42**, 1945–1968.

Efficient frequency-domain finite element modeling of two-dimensional elastodynamic scattering

Paul D. Wilcox[a] and Alexander Velichko
Department of Mechanical Engineering, University of Bristol, Queen's Building, University Walk, Bristol BS8 1TR, United Kingdom

(Received 1 April 2009; revised 14 October 2009; accepted 10 November 2009)

A frequency-domain finite element technique is presented that enables the complete characterization of a finite-sized scatterer using a minimum number of separate model executions and a relatively small spatial modeling domain. The technique is implemented using a commercial finite element package. A certain forcing profile is applied at a set of points surrounding the scatterer to cause a uni-modal plane wave to be incident on the scatterer from a specified direction. The scattered field is recorded and decomposed first into modes and then into far-field scattering coefficients in different directions. The data obtained from the model are represented in a scattering matrix that describes the far-field scattering response for all combinations of incident and scattering angles. The information in the scattering matrix can be efficiently represented in the Fourier domain by another matrix containing a finite number of Fourier coefficients. It is shown how the complete scattering behavior in both the near- and far-field can be extracted from the matrix of Fourier coefficients. Modeling accuracy is examined in various ways, including a comparison with the analytical solution for a circular cavity, and guidelines for the selection of modeling parameters are given.
 [DOI: 10.1121/1.3270390]

PACS number(s): 43.35.Cg, 43.20.Gp [LLT] Pages: 155–165

I. INTRODUCTION

This paper is concerned with numerically modeling the elastodynamic scattering process that occurs when an elastodynamic wave encounters a local change in material properties or a geometric discontinuity.

Mathematical modeling of the elastodynamic scattering process is challenging and exact analytical solutions only exist for certain geometrically simple shapes of scatterer such as spherical and cylindrical inclusions or voids.[1] Approximate solutions also exist that are valid under certain conditions. For example, if a scatterer is large and its shape slowly varying compared to the elastodynamic wavelength, the Kirchhoff approximation may be applied.[2–4] Alternatively, in the case of scattering from an inclusion whose properties are similar to those of the host medium, the Born approximation may be used.[5] Unfortunately, there are many scatterers of practical interest for which there is no analytical model or appropriate approximate model, and the scattering in these cases can only be modeled via direct numerical approaches such as finite elements (FE), finite differences (FD), boundary elements, etc.

It is only within the last few decades that the necessary computational power has become readily available to numerically solve scattering problems. There is considerable literature on numerical methods for solving acoustic problems (see, for example, the review by Thompson[6]) and somewhat less on elastodynamic problems. Much of the literature in both fields describes the implementation of complete numerical formulations with some aspect of the formulation (e.g., elements and boundaries) optimized for a particular problem. For example, Su *et al.*[7] developed a circular FE model of a scatterer and the local portion of the host medium and coupled this with an eigenfunction expansion of the far-field outside the FE domain to obtain solutions to acoustic scattering problems. Temple[8] developed a three-dimensional (3D) FD scheme to model elastic wave scattering from voids in heterogeneous, anisotropic material. Koshiba *et al.*[9] developed a two-dimensional (2D) FE model to model the scattering of guided elastic waves in plates from cracks. Kishore *et al.*[10] implemented a 2D FE model with absorbing boundaries to study the scattering of Rayleigh and bulk waves from cracks and circular holes. A recent study by Frehner *et al.*[11] compared different implementations of FE and FD methods for an elastodynamic bulk wave scattering problem in the context of seismology. The general conclusion from this work is that (in the spatial domain) FE is superior to FD in terms of the computational time needed to achieve a given accuracy. The authors suggest that this is due to the ability of FE to handle an unstructured mesh, which enables a more efficient representation of arbitrary scatterer geometry for a given number of nodes.

While these are important academic contributions, it is particularly the case in elastodynamic scattering that the implementation of even a basic FE framework is in itself a major programming task. Given that commercial FE packages capable of performing elastodynamic modeling (either in the frequency-domain or time-domain) are now widely available, there is considerable motivation to solve scattering problems within the framework of such packages.

One area in which commercial FE packages have been used extensively for elastodynamic scattering is guided waves. 2D models through the cross-section of a waveguide

[a]Author to whom correspondence should be addressed. Electronic mail: p.wilcox@bristol.ac.uk

have been used to examine the transmission and reflection of Lamb waves past various geometries of notches and cracks (e.g., Refs. 12 and 13) and Rayleigh wave diffraction around surface breaking cracks.[14] Of particular relevance here are 2D studies in the plane of the plate (which are mathematically identical to the 2D elastodynamic bulk wave case).[15,16] These models use explicit time-marching FE code and the modeling domain is designed to be sufficiently large for boundary reflections to be separated in the time-domain. Recently full 3D studies of Lamb wave scattering have also been reported[17,18] that use absorbing boundaries to reduce the size of the spatial modeling domain.

The work described in the current paper is motivated by the simulation of data from ultrasonic arrays in nondestructive evaluation (NDE) applications.[19] Each transmitting element in an array causes waves to be incident on a particular scatterer from a different angle, and similarly each receiving element detects the scattered field at different angles. Therefore in order to simulate the data associated with a scatterer, it is necessary to compute the amplitude and phase of scattered signals over a range of different incident and scattering angles. Furthermore, an array may well be in the near-field of a scatterer, in which case the scattered field also depends in a non-trivial manner on the radial as well as angular positions of the transmitter and receiver elements relative to the scatterer.

An observation about many of the numerical models reported in the literature is that they have been developed to investigate rather specific scattering cases. For example, Rajagopal and Lowe[16] investigated the scattered near-field for the case of SH_0 generated by a point source a finite distance away from a through-thickness crack oriented perpendicular to the incident direction. To extend this approach to simulating array data requires a separate model to be executed for each transmitter position relative to the scatterer. While this ultimately provides the required information, the results are not transferable; for example, to model the behavior of the same scatterer in a different position or orientation relative to an array requires a completely new set of models to be executed. Thus a better solution is to execute only the minimum number of models necessary to completely characterize the behavior of the scatterer, and record the data in such a way that it can be used to predict the scattering response from any position or orientation. For a typical ultrasonic array that contains of the order of 100 elements, this approach often actually requires a smaller number of model executions than directly predicting the scattering response separately for each transmitting element position.

In the current paper, an efficient FE procedure is presented for predicting the complete 2D scattering behavior of an arbitrary-shaped scatterer embedded in an otherwise homogeneous, isotropic host medium. Section II describes the theoretical basis and practical implementation of the modeling procedure in a commercial FE package. The accuracy of the procedure is investigated in Sec. III and example results are presented in Sec. IV. Some concluding remarks and suggestions for future avenues of investigation are made in Sec. V.

II. THEORY AND IMPLEMENTATION

A. Problem statement and overview of solution

The objective of the modeling process is to obtain a sufficient number of variables that enable the scattered field at any point from a scatterer in any incident wave field to be described. All quantities, including the scatterer geometry, are assumed to be invariant in the dimension perpendicular to the 2D plane of interest, and waves are therefore restricted to those propagating in this plane. In the current paper only the longitudinal mode and shear mode polarized in the plane of interest are considered. The 2D nature of the problem means that the scattering of the shear mode polarized perpendicular to the plane of interest is completely decoupled from the other two modes (the scattering of this mode being equivalent to an acoustic rather than elastodynamic problem). Although scattering of this mode is not considered, the same procedure as presented here could be slightly modified to do this.

The modeling procedure is summarized as follows. The scatterer and the immediately surrounding host medium are meshed, the outermost region of the mesh being an absorbing boundary. The modeling is performed in the frequency-domain at a single frequency. A system of harmonic point forces surrounding the scatterer is specified with the necessary amplitude and phases such that a plane wave of a specified mode is caused to be incident on the scatterer from a specified angle. The scattered field is monitoring at a series of points encircling the scatterer and decomposed first into modes and then into far-field values in different angular directions. This procedure is repeated for multiple incident angles equi-spaced over 360° and, if required, for both possible incident modes. As demonstrated below, the data so obtained are sufficient to recreate the scattered field at a location for any incident field. Furthermore, the number of incident angles required to completely specify the scattering behavior is finite and hence so is the required number of model executions.

B. Definition of the *S*-matrix and its relationship to the general scattering problem

To encapsulate the behavior of a scatterer, the concept of a scattering matrix or *S*-matrix is used. For consistency with previous work in the authors' research group,[20] the *S*-matrix is defined in the physically intuitive manner proposed by Schmerr[3] (p. 306), although Schmerr referred to the *S*-matrix as the scattering amplitude. It should be noted that the data in the *S*-matrix can be mapped to the many alternative scattering descriptions used by other authors, such as the transition or *T*-matrix,[21,22] the far-field pattern,[23] the scattering amplitude,[24] and the directivity.[25] The *S*-matrix is the 2D function describing far-field scattered amplitude as a function of incident, θ_1, and scattering, θ_2, angle for plane incident waves of unit amplitude. The angle definitions are as shown in the upper portion of Fig. 1, where the coordinate origin lies at the nominal center of the scatterer. It should be noted that here the angle of the incident wave θ_1 is defined by its angle of incidence on the scatterer, not the angle of its wavevector (equal to $\theta_1+\pi$) as used by some authors. For a

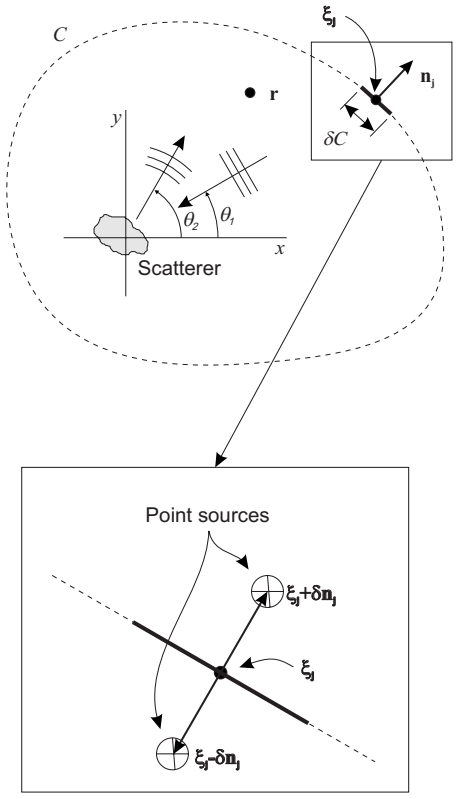


FIG. 1. Schematic of geometry showing scatterer, axes, wave angle definitions, and encircling contour. Inset shows details of numerical implementation of Helmholtz–Kirchhoff integral using point sources.

given scatterer, the S-matrix is also a function of frequency and the incident and scattered modes. Thus the complete 2D scattering behavior is described by the set of S-matrices $S^{\alpha\beta}(\theta_1,\theta_2,\omega)$ where ω is frequency and superscripts α and β refer to the type of incident and scattered modes (the possible types considered here being either longitudinal, denoted by $\alpha,\beta=L$, or in-plane polarized shear waves, denoted by $\alpha,\beta=S$). In the following discussion, subscripts are used to distinguish properties of the incident field (subscript=1) and scattered field (subscript=2). In the far-field, the scattered displacement amplitude decays with $1/\sqrt{r_2}$ (where r_2 is distance from the nominal center of the scatterer); thus the far-field values in the S-matrix must be specified at a prescribed distance, which is defined[20] as being at one wavelength $2\pi/k_\beta$ of the scattered mode, where k_β is the wavenumber of the scattered mode.

From the S-matrix definition the scattered far-field displacement u_2 is given by

$$u_2(\theta_1;r_2,\theta_2)=u_1S^{\alpha\beta}(\theta_1,\theta_2)\sqrt{\frac{2\pi}{k_\beta r_2}}\exp(ik_\beta r_2), \tag{1}$$

where u_1 is the amplitude and phase of the displacement of the plane wave incident from direction θ_1 at the scatterer location. The S-matrix can be written as a 2D Fourier series

$$S^{\alpha\beta}(\theta_1,\theta_2)=\sum_{m=-\infty}^{\infty}\sum_{m=-\infty}^{\infty}S_{mn}^{\alpha\beta}\exp(im\theta_1)\exp(in\theta_2), \tag{2}$$

with complex coefficients $S_{mn}^{\alpha\beta}$. By following a similar approach to that described by Cabayan *et al.*,[26] it is possible to show that the general expression for the scattered displacement field $\mathbf{u}_2=[u_r,u_\theta]^T$ at (r_2,θ_2) due to the application of a harmonic point source $\mathbf{f}_1=[f_r,f_\theta]^T$ at (r_1,θ_1) can be written in the form

$$\mathbf{u}_2=\frac{1}{D}\mathbf{M}_2^{\beta}(\mathbf{M}_1^{\alpha})^TC(r_1,\theta_1;r_2,\theta_2)\mathbf{f}_1, \tag{3}$$

where D_α is a constant that depends on the incident mode ($D_L=\lambda+2\mu$ and $D_S=\mu$ where λ and μ are the material Lamé constants), and the partial derivative vectors $\mathbf{M}_j^{\alpha,\beta}$ are computed according to the type of incident, $j=1$, or scattered, $j=2$, mode as follows:

$$\mathbf{M}_j^L=\begin{pmatrix}\partial/\partial r_j\\ 1/r_j\,\partial/\partial\theta_j\end{pmatrix},\quad \mathbf{M}_j^S=\begin{pmatrix}1/r_j\,\partial/\partial\theta_j\\ -\partial/\partial r_j\end{pmatrix} \tag{4}$$

and

$$C(r_1,\theta_1;r_2,\theta_2)=\frac{\pi\exp(-i\pi/4)}{4k_\alpha k_\beta}\times\sum_{m=-\infty}^{\infty}\sum_{n=-\infty}^{\infty}S_{mn}\exp\left[im\left(\theta_1+\frac{\pi}{2}\right)\right]\exp\left[in\left(\theta_2+\frac{\pi}{2}\right)\right]H_m^{(1)}(k_\alpha r_1)H_n^{(1)}(k_\beta r_2), \tag{5}$$

where k_α is the wavenumber of the incident mode. For example, in the L-L case, the resulting expression is

$$\mathbf{u}_2=\frac{1}{\lambda+2\mu}\begin{pmatrix}\partial^2/\partial r_1\,\partial r_2 & 1/r_1\partial^2/\partial\theta_1\,\partial r_2\\ 1/r_2\partial^2/\partial r_1\,\partial\theta_2 & 1/(r_1r_2)\partial^2/\partial\theta_1\,\partial\theta_2\end{pmatrix}C(r_1,\theta_1;r_2,\theta_2)\mathbf{f}_1. \tag{6}$$

The necessary partial derivatives of C are calculated using the recurrence relations for Hankel functions.[27]

Expression (3) is theoretically exact as long as r_1, $r_2>r_{\max}$ where $r_{\max}$ is the radius of a circle that just encloses the scatterer. However, in the immediate vicinity of the scatterer, infinitesimal contributions from all terms in the infinite series become significant and in practice numerical instability prevents accurate reconstruction. However, the extent of the unstable region is typically less than a wavelength (i.e., much less than the near-field region of the scatterer), and the inability to accurately predict scattering here is not a problem for the applications of interest.

It can be shown that Eq. (3) is consistent with the original S-matrix definition (1) by considering its behavior in the far-field limit as $k_1 r_1$ and $k_2 r_2$ become large. Here, the appropriate asymptotic forms of the Hankel functions enable Eq. (1) to be written as

$$C(r_1,\theta_1;r_2,\theta_2) \approx -i\frac{\exp(-i\pi/4)}{2k_1k_2}\frac{\exp(ik_\alpha r_1)}{\sqrt{k_\alpha r_1}}\frac{\exp(ik_\beta r_2)}{\sqrt{k_\beta r_2}} \times \sum_{m,n} S_{mn}^{\alpha\beta}\exp(im\theta_1)\exp(in\theta_2). \quad (7)$$

The last term in the above expression is the Fourier representation of $S^{\alpha\beta}(\theta_1,\theta_2)$ given by Eq. (2), while the second and third terms correspond to the spatially decaying incident and scattered fields. By using the far-field expressions[28] for a point source at (r_1,θ_1) to deduce the amplitude of the incident wave displacement at the scatterer u_1 and substituting the far-field expression for C into Eq. (3), the original expression for far-field displacement in terms of $S^{\alpha\beta}(\theta_1,\theta_2)$ is regained.

It can be shown[29] that the S-matrix of a finite-sized scatterer can be represented to very high accuracy by a finite number of complex coefficients $S_{mn}^{\alpha\beta}$. Once these coefficients are known, the scattered field at any position due to a point force at any position (and by extension the scattered field due to any incident field) can be computed. The goal of the FE modeling procedure is therefore to deduce the necessary number of $S_{mn}^{\alpha\beta}$ coefficients as efficiently as possible.

C. Numerical implementation of the Helmholtz–Kirchhoff integral

In order to deal with the issue of different wave modes, special types of point source and receiver are used in the FE model that allow the two modes to be selectively excited and detected, as described in Sec. II D. Effectively, the point sources and receivers act on the potential function of the desired mode (i.e., the scalar potential function φ for L waves and the ψ_z component of the vector potential $\boldsymbol{\psi}$ for S waves polarized in the x-y plane). Because the FE modeling is performed wholly in the frequency-domain, the notation in the following analysis is therefore that of a scalar scattering problem at a single frequency with the variable Φ used to denote the appropriate scalar potential function.

The Helmholtz–Kirchhoff integral for scalar waves provides an expression for the wave field Φ at any point $\mathbf{r}$ within a closed contour C, as a function of an arbitrary field Φ_{inc} on the boundary. The general expression is

$$\Phi(\mathbf{r}) = \int_C (G(\mathbf{r},\boldsymbol{\xi})\,\nabla\,\Phi_{\text{inc}}(\boldsymbol{\xi}) - \Phi_{\text{inc}}(\boldsymbol{\xi})\,\nabla\,G(\mathbf{r},\boldsymbol{\xi}))\cdot\mathbf{n}\,dC_{\boldsymbol{\xi}}, \quad (8)$$

where $\mathbf{n}$ is the outward surface normal and G is the appropriate Green's function. The physical interpretation is that a desired field Φ can be synthesized by the superposition of appropriate monopole, G, and dipole, ∇G, sources on the contour C. In the context of modeling 2D elastodynamic scattering behavior, the surface integral is a closed line integral around a region containing the scatterer. Given that a uni-modal point source can be realized in the FE model, the uni-modal dipole in the Kirchhoff integral can also be realized via a pair of uni-modal monopoles in anti-phase, separated by a small distance 2δ in the desired polarization direction, i.e.,

$$\nabla G(\mathbf{r},\boldsymbol{\xi})\cdot\mathbf{n} \approx \frac{G(\mathbf{r},\boldsymbol{\xi}+\delta\mathbf{n}) - G(\mathbf{r},\boldsymbol{\xi}-\delta\mathbf{n})}{2\delta}. \quad (9)$$

For reasons of symmetry in the numerical implementation, the incident field amplitude is based on the mean of its value at the same two positions:

$$\Phi_{\text{inc}}(\boldsymbol{\xi}) \approx \frac{\Phi_{\text{inc}}(\boldsymbol{\xi}+\delta\mathbf{n}) + \Phi_{\text{inc}}(\boldsymbol{\xi}-\delta\mathbf{n})}{2}. \quad (10)$$

Likewise, the gradient of the incident field is approximated as

$$\nabla\Phi_{\text{inc}}(\boldsymbol{\xi})\cdot\mathbf{n} \approx \frac{\Phi_{\text{inc}}(\boldsymbol{\xi}+\delta\mathbf{n}) - \Phi_{\text{inc}}(\boldsymbol{\xi}-\delta\mathbf{n})}{2\delta}, \quad (11)$$

and the required monopole source is actually realized via two in-phase monopole sources:

$$G(\mathbf{r},\boldsymbol{\xi}) \approx \frac{G(\mathbf{r},\boldsymbol{\xi}+\delta\mathbf{n}) + G(\mathbf{r},\boldsymbol{\xi}-\delta\mathbf{n})}{2}. \quad (12)$$

The overall result is that the field within the closed boundary S can be approximated by

$$\Phi(\mathbf{r}) \approx \frac{1}{2\delta}\int_C \Phi_{\text{inc}}(\boldsymbol{\xi}+\delta\mathbf{n})G(\mathbf{r},\boldsymbol{\xi}-\delta\mathbf{n}) - \Phi_{\text{inc}}(\boldsymbol{\xi}-\delta\mathbf{n})G(\mathbf{r},\boldsymbol{\xi}+\delta\mathbf{n})\,dC_{\boldsymbol{\xi}}. \quad (13)$$

The next stage in the implementation is to discretize the integral into small elements of length δC and represent each element by two monopole sources. This is shown schematically in the inset in Fig. 1. In order to generate the S-matrix for a particular incident mode, plane waves are required to be incident on the scatterer from a prescribed angle θ_1. The integral can thus be written as a summation of the source contributions

$$\Phi(\mathbf{r},\theta_1) \approx \frac{\delta C}{2\delta}\sum_j w_j^+(\theta_1)G(\mathbf{r},\boldsymbol{\xi}_\mathbf{j}-\delta\mathbf{n}_\mathbf{j}) - w_j^-(\theta_1)G(\mathbf{r},\boldsymbol{\xi}_\mathbf{j}+\delta\mathbf{n}_\mathbf{j}), \quad (14)$$

where $\zeta_\mathbf{j}$ now represents the point on the boundary at the mid-position of the jth pair of monopole sources, $\mathbf{n}_\mathbf{j}$ represents the surface normal at that point, and the weighting functions are

$$w_j^\pm(\theta) = \exp[i\mathbf{k}\cdot(\boldsymbol{\xi}_\mathbf{j} \pm \delta\mathbf{n}_\mathbf{j})], \quad \mathbf{k} = [k\cos\theta \quad k\sin\theta]^T. \quad (15)$$

By reciprocity, the measured field Φ_m is recorded at the same points as the monopole sources around the contour C in the FE model can be decomposed into plane waves by applying exactly the same weightings. This plane wave decomposition contains the superposition of incident and scattered fields; thus to extract the scattered field alone it is necessary to subtract the incident field. This can be achieved by subtract-

ing the theoretical incident field, but superior results are obtained if the incident field is obtained from a separate FE model with identical parameters but without a scatterer present. Once the scattered field is obtained in terms of its plane wave decomposition it needs to be converted into far-field values consistent with the definition of the S-matrix. By expressing the far-field as an integral of plane wave contributions and by applying the method of stationary phase it is found that the S-matrix values are simply related to the plane wave decomposition of the scattered field by a factor of $\exp(i\pi/4)/(4\pi)$. Thus the S-matrix can be written in terms of the FE model output as

$$S(\theta_1,\theta_2)=\frac{\exp(i\pi/4)}{4\pi}\times\left[\frac{\delta C}{2\delta}\sum_j w_j^+(\theta_2)\Phi_m^{(\mathrm{tot})}(\boldsymbol{\xi}_\mathbf{j}-\delta\mathbf{n}_\mathbf{j},\theta_1)\right.$$
$$-w_j^-(\theta_2)\Phi_m^{(\mathrm{tot})}(\boldsymbol{\xi}_\mathbf{j}+\delta\mathbf{n}_\mathbf{j},\theta_1)$$
$$-\frac{\delta C}{2\delta}\sum_j w_j^+(\theta_2)\Phi_m^{(\mathrm{inc})}(\boldsymbol{\xi}_\mathbf{j}-\delta\mathbf{n}_\mathbf{j},\theta_1)$$
$$\left.-w_j^-(\theta_2)\Phi_m^{(\mathrm{inc})}(\boldsymbol{\xi}_\mathbf{j}+\delta\mathbf{n}_\mathbf{j},\theta_1)\right], \tag{16}$$

where the superscripts (tot) and (inc) refer to measurements made with and without the presence of a scatterer, respectively.

D. Point transmitters and receivers

In Sec. II C, it was assumed that uni-modal point sources and receivers could be realized at any spatial position in the FE model. In this subsection, the details of this realization are presented.

In the FE model, excitation is ultimately via the application of uni-directional harmonic forces applied at prescribed nodes in the mesh. Similarly, the detection of wave motion must ultimately come from measured nodal displacements. Previously, uni-modal point sources and receivers with epicenters located precisely at nodes in a regular mesh of identical square elements were synthesized by applying a number of radial or tangential point forces to the surrounding nodes.[20] However, the numerical implementation of the Helmholtz–Kirchhoff integral requires sources to be placed with epicenters at arbitrary spatial locations that may not coincide with nodes. One possible solution is to use the same method[20] and to simply approximate the desired location of a source to that of the nearest available node. Figure 2(a) shows the radial forces that are used to excite the longitudinal mode by this method when the desired location is the point marked with a cross. In this figure, a square element size ε of $\lambda_L/40$ is used as a representative example. The desired source position is intentionally chosen as being at the center of an element, as this represents the worst-case scenario with a distance of $(\sqrt{2}/2)\varepsilon$ to the nearest node. The resulting displacement fields shown in Fig. 2(b) exhibit relative fluctuations in longitudinal wave amplitude of $\pm 8\times 10^{-4}$, a relative fluctuation in longitudinal wave phase of $\pm 6°$, and a relative shear wave amplitude of 0.02. The phase fluctuation and unwanted shear wave amplitude have been found to be too large to be acceptable. For this reason the approach described below is used that allows the simulation of a uni-modal source at an arbitrary spatial position.

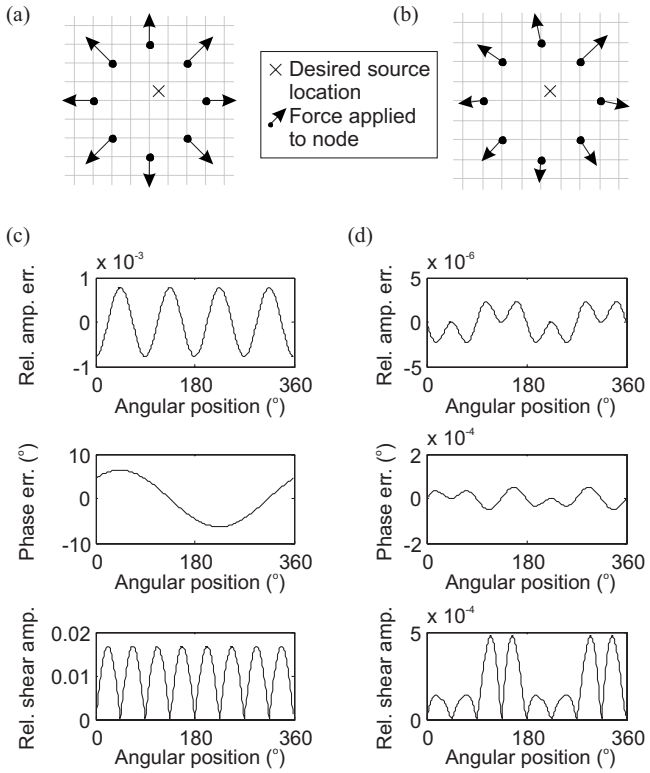


FIG. 2. Nodal forces to synthesize a point source of longitudinal waves by (a) approximation of its position to the nearest node and (b) by the method described in Sec. II D. The properties of the resulting fields with respect to the desired source location [indicated by an x in (a) and (b)] for the two cases are shown in (c) and (d), respectively.

A coordinate origin **O** is defined at the desired source location. In the vicinity of **O** is a nodal location in the FE mesh **p** at which a harmonic point force with orientation and amplitude given by the vector **d** can be applied. The resulting wave field from such a source contains displacements from longitudinal and shear waves. At a distant point **r** in the far-field (i.e., $|\mathbf{r}|\gg|\mathbf{p}|$) the longitudinal displacement u_L is entirely in the radial direction and the shear displacement u_S is entirely in the tangential direction. These are given by[28]

$$u_L(\mathbf{r})=\frac{\hat{\mathbf{r}}\cdot\mathbf{d}}{\lambda+2\mu}\frac{\exp(ik_L|\mathbf{r}|)}{(8\pi k_L|\mathbf{r}|)^{1/2}}\exp\left[i\left(\frac{\pi}{4}-k_L\hat{\mathbf{r}}\cdot\mathbf{p}\right)\right]$$
$$=C_L\hat{\mathbf{r}}\cdot\mathbf{d}\,\exp(-ik_L\hat{\mathbf{r}}\cdot\mathbf{p}), \tag{17}$$

$$u_S(\mathbf{r})=\frac{\hat{\mathbf{r}}_\perp\cdot\mathbf{d}}{\mu}\frac{\exp(ik_S|\mathbf{r}|)}{(8\pi k_S|\mathbf{r}|)^{1/2}}\exp\left[i\left(\frac{\pi}{4}-k_S\hat{\mathbf{r}}\cdot\mathbf{p}\right)\right]$$
$$=C_S\hat{\mathbf{r}}_\perp\cdot\mathbf{d}\,\exp(-ik_S\hat{\mathbf{r}}\cdot\mathbf{p}),$$

where λ and μ are the Lame constants, C_L and C_S are constants, $\hat{\mathbf{r}}$ is a unit vector in the direction of **r**, and $\hat{\mathbf{r}}_\perp$ is a unit vector equal to $\hat{\mathbf{r}}$ rotated by 90°.

The field from n such sources at nodal locations $\mathbf{p_1}\cdots\mathbf{p_n}$ surrounding **O** and with orientation and amplitude $\mathbf{d_1}\cdots\mathbf{d_n}$ is now considered. The total longitudinal and shear displacement at m field points $\mathbf{r_1}\cdots\mathbf{r_m}$ equally spaced around **O** at an arbitrarily large fixed distance are written in terms of the

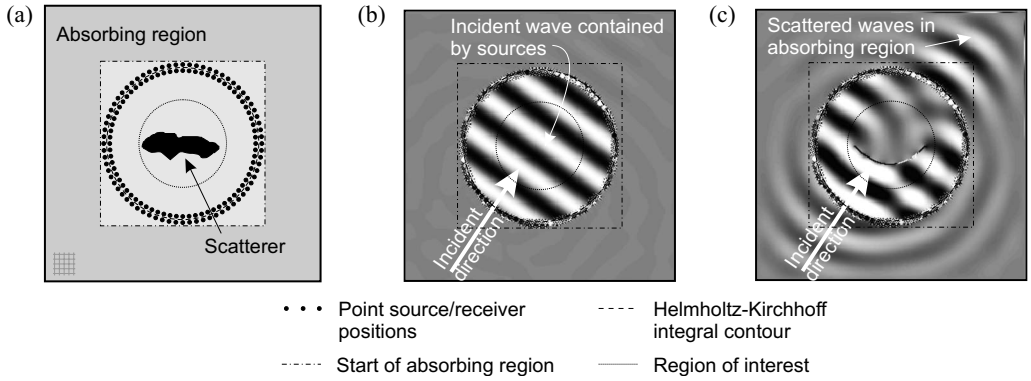


FIG. 3. (a) Schematic of FE model geometry and snapshots of the L wave potential when a plane longitudinal wave is injected in direction shown (b) in the absence of a scatterer and (c) in the presence of a curved crack.

unknown forces $\mathbf{d_1}\cdots\mathbf{d_n}$ and known physical forcing locations $\mathbf{p_1}\cdots\mathbf{p_n}$. This yields

$$\begin{bmatrix} u_L(\mathbf{r}_1) \\ \vdots \\ u_L(\mathbf{r_m}) \\ u_S(\mathbf{r}_1) \\ \vdots \\ u_S(\mathbf{r_m}) \end{bmatrix} = \begin{bmatrix} C_L\hat{\mathbf{r}}_\mathbf{1}^T \exp(-ik_L\hat{\mathbf{r}}_\mathbf{1}\cdot\mathbf{p_1}) & \cdots & C_L\hat{\mathbf{r}}_\mathbf{1}^T \exp(-ik_L\hat{\mathbf{r}}_\mathbf{1}\cdot\mathbf{p}_n) \\ \vdots & & \vdots \\ C_L\hat{\mathbf{r}}_\mathbf{m}^T \exp(-ik_L\hat{\mathbf{r}}_\mathbf{m}\cdot\mathbf{p_1}) & \cdots & C_L\hat{\mathbf{r}}_\mathbf{m}^T \exp(-ik_L\hat{\mathbf{r}}_\mathbf{m}\cdot\mathbf{p}_n) \\ C_S\hat{\mathbf{r}}_\mathbf{1}^T \exp(-ik_S\hat{\mathbf{r}}_\mathbf{1}\cdot\mathbf{p_1}) & \cdots & C_S\hat{\mathbf{r}}_\mathbf{1}^T \exp(-ik_S\hat{\mathbf{r}}_\mathbf{1}\cdot\mathbf{p}_n) \\ \vdots & & \vdots \\ C_S\hat{\mathbf{r}}_\mathbf{m}^T \exp(-ik_S\hat{\mathbf{r}}_\mathbf{m}\cdot\mathbf{p_1}) & \cdots & C_S\hat{\mathbf{r}}_\mathbf{m}^T \exp(-ik_S\hat{\mathbf{r}}_\mathbf{m}\cdot\mathbf{p}_n) \end{bmatrix} \begin{bmatrix} \mathbf{d_1} \\ \vdots \\ \mathbf{d_n} \end{bmatrix}. \tag{18}$$

The desired field values u_L and u_S are then specified (unity for the desired mode and zero for the undesired mode) and the system of equations solved.

The number of scalar unknowns is $2n$; hence the number of field points must be at least $2n$ (if greater, then the equations are solved in a least-squares sense using the Moore–Penrose pseudo-inverse). In the current implementation, forces are applied at 8 nodes and 32 field points (16 for the desired mode and 16 for the undesired mode) are used. An example of the forcing pattern required to synthesize a source at a non-nodal location is shown in Fig. 2(c) and the resulting field in Fig. 2(d). It can be seen that the regularity and modal purity of the resulting field are very high (relative fluctuation in longitudinal mode amplitude $\pm 2\times 10^{-6}$, phase fluctuation in longitudinal mode $\pm 5\times 10^{-5}$ deg, and relative shear mode displacement amplitude 5×10^{-4}). Although not rigorously proven, numerical evidence suggests that this procedure seems to work as long as the element size ε is smaller around $\lambda_L/10$. In passing it is noted that the same technique can be applied without modification to forcing nodes in non-uniform meshes too, although in the current paper modeling is restricted to regular meshes composed of identical square elements.

E. Implementation

The FE model is implemented in ABAQUS (now branded as Simulia, Providence, RI) using square, four-noded, plane-strain elements (ABAQUS type CPE4). For ease of meshing, the complete mesh is also square as shown in Fig. 3(a), although a circular mesh is somewhat more efficient. The material damping C in elements in the central region of the model is zero but increases from zero to a maximum value at the outside of the model in the absorbing region indicated. If L is the thickness of the absorbing region expressed in elements and $l=1\cdots L$ is the position of an element within the thickness of the absorbing region (with $l=1$ being innermost) then the damping of the element is specified as

$$C = C_0\left(\frac{l}{L}\right)^2, \tag{19}$$

where C_0 is the maximum material damping. Previous research[30] has suggested that this quadratic increase in damping gives a high absorption of waves of both modes over a large range of incident angles. The model input file is generated in MATLAB (MathWorks Inc., Natick, MA), which specifies the mesh, including the forcing (and monitoring) locations, as well as the actual forces to apply for each incident wave direction and mode based on the weightings given by Eq. (15).

Example snapshots of the complete model are shown in Figs. 3(b) and 3(c). It can be seen in Fig. 3(b) that, in the absence of a scatterer, the nature of the forcing largely contains the excited field to be entirely within the excitation points and the absorbing layer is somewhat redundant. However, when a scatterer is introduced [Fig. 3(c)], the scattered field passes through the excitation points and the absorbing layer is necessary to prevent it from being reflected back in and contaminating the measured displacements. After execu-

TABLE I. Modeling parameters.

Description	Dimension	Effect on number and size of models	Default value
Element size, ε	Length	Significant (size proportional to ε^{-2})	$(1/40)\lambda_L$
ROI diameter	Length	Significant contribution to size if large	$2\lambda_L$
Separation between ROI and nearest excitation/monitoring node	Length	Major contribution to model size if ROI is small	λ_L
Separation between monopoles within a pair	Length	Negligible	$(1/10)\lambda_L$
Separation between pairs of monopoles around a contour	Length	None	$(1/5)\lambda_L$
Step between incident angles $\delta\theta_1$	Angle	Significant (number of models proportional to $\delta\theta_1^{-1}$)	$\delta\theta_{1,2}=\frac{2\pi}{4n}\leq\frac{\lambda_L}{3D}$
Step between scattering angles $\delta\theta_2$	Angle	None	
Thickness of absorbing layer	Length	Major contribution to model size if ROI is small	$1.5\lambda_L$
Maximum damping in absorbing layer	Time^{-1}	None	$2\pi c_L/\lambda_L$ (c_L=bulk longitudinal wave velocity)

tion, the model output file is post-processed in MATLAB, which decomposes the nodal displacements into modes and then plane waves using expression (15). After executing models with and without a scatterer present, the *S*-matrix is computed using expression (16).

III. CONVERGENCE AND ACCURACY ESTIMATES

The FE-based modeling procedure is subject to the usual errors associated with numerical modeling. In the following investigations, an imaginary material is used with Poisson's ratio=1/3, resulting in a longitudinal to shear wavelength ratio of precisely 2. All linear dimensions are specified in terms of longitudinal wavelength λ_L. The parameters that affect FE accuracy and a qualitative description of their effect on computation time are listed in Table I. The term region of interest (ROI) is introduced, which refers to a circular region that completely bounds the scatterer. The Helmholtz–Kirchhoff integration contour is intentionally somewhat larger than the ROI in order to avoid the ROI from being contaminated by near-field effects associated with the finite-sized, discrete point sources used for excitation.

Due to the number of variables involved and the coupling between them, a complete analysis of all combinations of parameter values is challenging. For this reason, quantitative analysis has been restricted to those parameters that significantly affect the size and number of FE models, and here only limited results are presented. For other parameters, the default values specified in Table I are used unless otherwise specified.

Two parameters that warrant further explanation are the number of incident and scattering angles considered. The number of separate models that must be executed to characterize a given scatterer (at one frequency) is equal to the number of incident angles. However, the number of incident angles required to obtain all the necessary information is finite. (The same is true of the number of scattering angles into which the scattered field is decomposed although this does not affect the number of models that must be executed.) The rule for setting the number of incident and scattering angles specified in Table I is based on two criteria: (a) An integer number of angles cover 90° (this allows the exploitation of two planes of symmetry if appropriate) and (b) the arc length corresponding to the step between angles on the perimeter of the region of interest is a least three times shorter than the shear wavelength (and thus $<\lambda_L/6$). The latter may be loosely interpreted as satisfying the Nyquist sampling criteria (with a safety factor of 1.5) for shear waves.

Section III A investigates the ability of the FE modeling procedure to generate a uniform incident field of the desired mode in the absence of a scatterer. Section III B uses the FE modeling procedure to predict the scattering from a circular cavity and compares the results obtained to the analytical solution.

A. Uniformity and modal purity of incident field

The FE model is executed without a scatterer present, and the complete displacement field within the ROI is monitored during excitation for each incident angle of plane waves for each mode. The metrics used to assess irregularities in the incident field are the mean and standard deviation of its amplitude at all points within the ROI. However, before these metrics can be computed, the measured displacements within the ROI are decomposed into longitudinal and shear wave contributions. The decomposition is based on the following identities to obtain the longitudinal and shear potential functions, φ and ψ_z, from the displacement field $\mathbf{u}$:

$$\phi=-\frac{1}{|k_L|^2}\nabla\,\mathbf{u},\quad \psi_z=-\frac{1}{|k_S|^2}\nabla\times\mathbf{u}. \tag{20}$$

Numerically these are evaluated for each element in the ROI, by numerical differentiation of the displacements at the four corner nodes

$$\phi\approx-\frac{1}{|k_L|^2}\frac{[(u_3-u_4)+(u_2-u_1)]+[(v_3-v_2)+(v_4-v_1)]}{2\varepsilon},$$

$$\psi_z\approx-\frac{1}{|k_S|^2}\frac{[(v_2-v_1)+(v_3-v_4)]-[(u_4-u_1)+(u_3-u_2)]}{2\varepsilon}, \tag{21}$$

where ε is the element side length, u and v are the measured displacements in the x and y directions, and the subscripts indicate the local node number counted anti-clockwise around the element starting at the lower left corner. It can be shown that this gives a relative accuracy of the order of

$(\varepsilon/\lambda_{L,S})^2$, which is included in the error bars associated with the error estimates in the following analyses.

Metrics to assess the regularity and mode purity of the incident field within the ROI are computed as follows, where the parameters Φ_d and Φ_u refer to the potential functions (i.e., either φ or ψ_z) of the desired and undesired modes, respectively. The mean amplitude m_{da} and phase m_{dp} of the desired mode are given by

$$\begin{Bmatrix} m_{da} \\ m_{dp} \end{Bmatrix} = \frac{1}{An_\theta} \sum_{\theta_1} \int_{\mathrm{ROI}} \begin{Bmatrix} \mathrm{mod} \\ \mathrm{rel\ arg} \end{Bmatrix} \Phi_d(\theta_1, \mathbf{r}) d\mathbf{r}, \tag{22}$$

where A is the area of the ROI and n_θ is the number of incident angles considered. The function rel arg is the difference between the measured and theoretical phase at each point

$$\mathrm{rel\ arg}\ \Phi(\mathbf{r}) = \arg\ \Phi(\mathbf{r}) - \arg \exp(i\mathbf{kr}), \tag{23}$$

where $\mathbf{k}$ is the wavevector of the relevant mode. In the case of the phase calculation, due care is taken to avoid artifacts due to phase wrapping. The ideal values for these parameters are $m_{da}=1$ and $m_{dp}=0°$. The standard deviations of the amplitude s_{da} and phase s_{dp} of the desired mode are given by

$$\begin{Bmatrix} s_{da} \\ s_{dp} \end{Bmatrix} = \sqrt{\frac{1}{An_\theta} \sum_{\theta_1} \int_{\mathrm{ROI}} \left[\begin{Bmatrix} \mathrm{mod} \\ \mathrm{rel\ arg} \end{Bmatrix} \Phi_d(\theta_1, \mathbf{r}) - \begin{Bmatrix} m_{da} \\ m_{dp} \end{Bmatrix} \right]^2 d\mathbf{r}}. \tag{24}$$

The metrics s_{da} and s_{dp} describe the uniformity of the incident field, both within the ROI and as the incident angle is altered. Leakage of energy into the undesired mode is assessed by the mean m_{ua} and standard deviation s_{ua} of its potential function over the ROI and over all incident angles

$$m_{ua} = \frac{1}{An_\theta} \sum_{\theta_1} \int_{\mathrm{ROI}} \mathrm{mod}\ \Phi_u(\theta_1, \mathbf{r}) d\mathbf{r} \tag{25}$$

and

$$s_{ua} = \sqrt{\frac{1}{An_\theta} \sum_{\theta_1} \int_{\mathrm{ROI}} [\mathrm{mod}\ \Phi_u(\theta_1, \mathbf{r}) - m_{ua}]^2 d\mathbf{r}}. \tag{26}$$

The effect of element size on these metrics is considered, using 20, 40, 60, or 80 elements per longitudinal wavelength and with all other parameters set to their default values listed in Table I. The results are shown in Fig. 4, which show the mean amplitude error and phase error associated with the mode of interest (target amplitude=1 and phase=0°) and the mean amplitude of the unwanted mode (target amplitude=0). The error bars are the standard deviation of the amplitude or phase for all points in the ROI for all incident angles modeled plus the $(\varepsilon/\lambda_{L,S})^2$ error associated with the extraction of the longitudinal or shear displacement field. Although results are not presented, the effect of mesh anisotropy has been examined separately and found to introduce a systematic variation in amplitude of the form $\cos 4\theta_1$, which is approximately double that due to the $(\varepsilon/\lambda_{L,S})^2$ error. In summary, there is a monotonic improvement in all errors as the element size is decreased. A model with 40 elements per longitudinal wavelength results in amplitude errors of $<2\%$ for both

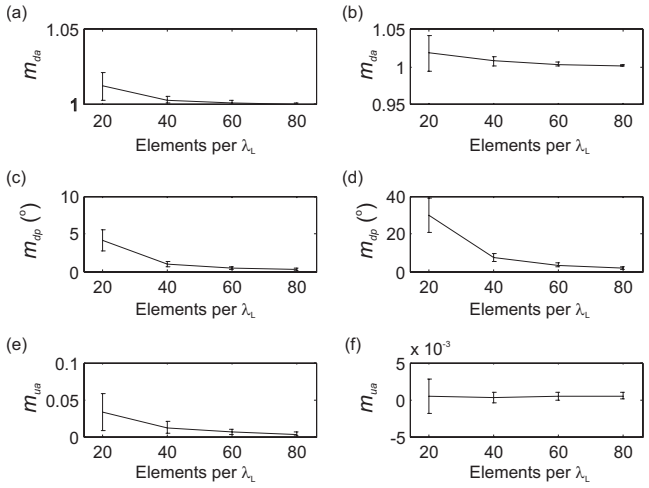


FIG. 4. Uniformity of incident field as a function of number of elements per longitudinal wavelength: [(a) and (b)] amplitude of desired mode for longitudinal and shear wave incidence, respectively; [(c) and (d)] phase of desired mode for longitudinal and shear wave incidences, respectively; [(e) and (f)] amplitude of undesired mode for longitudinal and shear wave incidences, respectively.

modes (i.e., uniformity of amplitude of desired mode between 0.98 and 1.02 and amplitude of unwanted mode less than 0.02) and phase consistencies of $<1.5°$ for longitudinal incidence and $<10°$ for shear incidence. These values are based on the mean error plus one standard deviation and are thus equal to the extrema of the error bars in Fig. 4.

The results of varying the separation between the ROI and the forcing/monitoring points have also been investigated using the same metrics. The results are not presented here, but show that there is no significant improvement in the uniformity and purity of the incident field once the gap is greater than $\lambda_L/4$. In other words, the near-field effect of the forcing/monitoring points falls off very rapidly.

B. Comparison with analytical solution for circular cavity

The ultimate purpose of the FE model is to predict scattering behavior, so it is appropriate to assess its accuracy in the presence of a scatterer. There are a limited number of cases where an exact analytical solution exists for comparison, and one of these is scattering by a circular cavity. In addition to the consequences of the errors arising from irregularities in the incident field discussed previously, the following sources of error are anticipated:

(1) imperfect description of scatterer geometry by square elements,
(2) artifacts due to the proximity of the scatterer to source and monitoring points, and
(3) imperfect absorption of scattered waves by absorbing boundaries causing energy to be reflected back into model, contaminating decomposition of scattered field.

The far-field scattering behavior of a circular cavity of radius a on which a plane longitudinal wave is incident from the $\theta_1=\pi$ direction is given originally by White.[31] Here, the results in the paper by Lopez-Sanchez *et al.*[4] are used, as

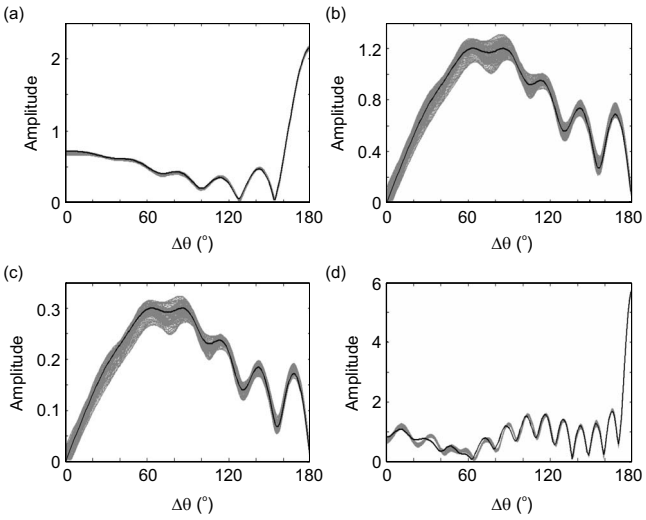


FIG. 5. Example of scattered fields from circular cavity of radius λ_L as a function of the difference between incident and scattering angles $\Delta\theta$ for mode combinations: (a) *L-L*, (b) *L-S*, (c) *S-L*, and (d) *S-S*. The multiple gray lines on each graph are FE model predictions for different incident angles and the black lines are the analytical solutions.

these provide the complete solutions for shear as well as longitudinal wave incidence on a circular cavity. A circular cavity of radius $a=\lambda_L$ is modeled. Due to the axial symmetry of the cavity, the true scattered field is only a function of the difference between incidence and scattering angles. A FE-generated *S*-matrix for a particular mode combination $S(\theta_1,\theta_2)$ can therefore be mapped to a series of functions $S'(\Delta\theta|\theta_1)=S(\theta_1,\theta_1+\Delta\theta)$, for different incident angles θ_1. These should theoretically be identical and variations between them can be attributed to effects such as the discretization of the scatterer, mesh induced anisotropy, angular variation of the absorbing region, etc. Example results using the default mesh parameters in Table I are shown in Fig. 5. For the cases where the incident and scattered modes are the same [Figs. 5(a) and 5(d)], it can be seen that there is very little variation in the FE predictions with incident angle (i.e., all the gray lines lie very close together), while in the mixed mode cases [Figs. 5(b) and 5(c)] the variation appears to be much greater. The reason for this is not understood, although it is noted that in terms of energy only 25% (in the *L* incident case) and 10% (in the *S* incident case) of the scattered field undergo mode conversion. The qualitative agreement between FE predictions and the analytical solution appears very good in the *L-L* case and somewhat poorer in the other cases, with systematic errors (i.e., for all incident angles) present in the *S-S* at low values of $\Delta\theta$.

In order to quantify the accuracy further and allow comparison of different modeling parameters, the mean amplitude error e_a is defined as

$$e_a=\frac{\sqrt{\frac{1}{4\pi^2}\int_0^{2\pi}\int_0^{2\pi}[|S_{\mathrm{FE}}(\theta_1,\theta_2)|-|S(\theta_1,\theta_2)|]^2 d\theta_1 d\theta_2}}{\frac{1}{4\pi^2}\int_0^{2\pi}\int_0^{2\pi}|S(\theta_1,\theta_2)|d\theta_1 d\theta_2}, \tag{27}$$

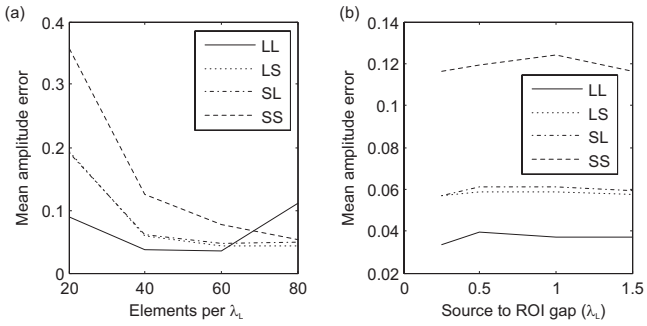


FIG. 6. Mean amplitude error for a circular cavity of radius λ_L as a function of (a) elements per longitudinal wavelength and (b) the gap between the sources and ROI.

where S_{FE} is the FE prediction for a particular mode combination and *S* is the corresponding analytical result. Physically, e_a is the rms error between S_{FE} and *S*, normalized to the mean amplitude of *S*. The metrics are evaluated for different model parameters, and two important results are shown in Fig. 6. From Fig. 6(a) it can be observed that the mean amplitude error diminishes with increasing numbers of elements per wavelength (i.e., smaller elements), although there is an anomalous point for the *L-L* case for the smallest element size. This is believed to be due to the limited precision of the data in the FE output file, which at large wavelength to element sizes begins to introduce numerical significant differentiation noise (due to the rounding errors) during the modal decomposition in post-processing. Discounting this anomaly, it can be observed that the *S-S* error is always the largest and this is expected due to the shorter wavelength. From Fig. 6(b) it can be observed that there is no clear trend between the mean amplitude error and the separation between the sources and the ROI. This suggests that the effect of any artifacts due to the proximity of the scatterer to the source and monitoring points are negligible compared to other sources of error. Therefore a source to ROI separation of 0.25 longitudinal wavelengths appears to be sufficient.

A more general observation from Fig. 6 is that, even though the calculation is not quite equivalent, the errors in the final *S*-matrix prediction are significantly larger than those associated with the generation and reception of plane waves discussed in Sec. III A. This suggests that the limiting problem is either the discretization of the scatterer itself, or problems associated with the absorbing regions that are only manifest when a scatterer is present. More work is required to investigate both these possibilities using different meshing techniques or different absorbing region parameters.

In summary, it is suggested that 40–60 square linear elements per longitudinal wavelength is an appropriate number for practical modeling. Smaller elements than this provide at best only marginal improvements and may, due to rounding, actually lead to significantly larger errors. It is quite possible that other element types (e.g., triangular, quadratic, etc.) may provide superior accuracy for a given total number of degrees of freedom in a model. This was observed by Frehner *et al.*,[11] who found that an unstructured FE mesh enabled better matching to scatterer geometry, and conse-

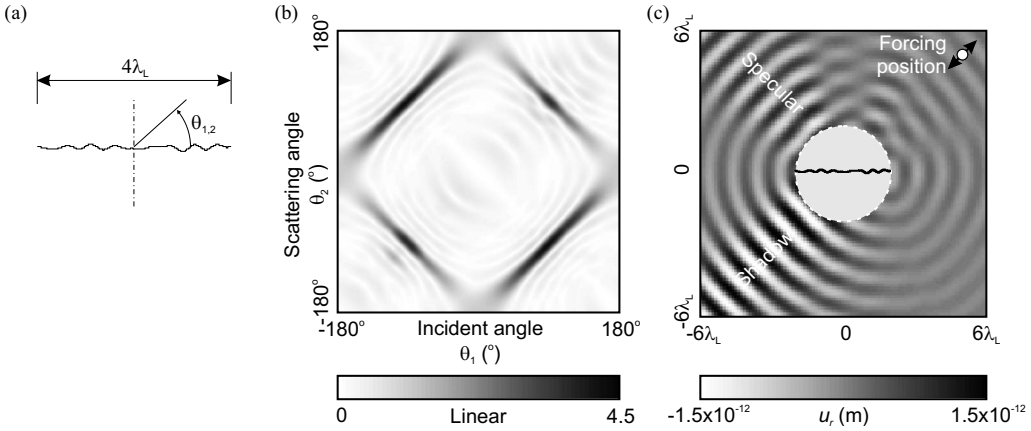


FIG. 7. (a) Realization of irregular crack in FE model, (b) *S*-matrices for *L-L* case, and (c) snapshot of scattered field (radial displacement u_r) reconstructed from data in *S*-matrix.

quently better accuracy, with fewer elements than a regular mesh similar to that described here. This remains to be investigated further.

IV. EXAMPLE RESULTS FOR AN IRREGULAR CRACK

To illustrate the generation of an *S*-matrix for an irregular defect geometry, the crack shown in Fig. 7(a) is considered. The crack is modeled in FE by disconnecting the appropriate nodes along the closest possible path, resulting in the "digitized" version shown in the figure. The ROI diameter is $4\lambda_L$ and the separation between ROI and nearest excitation/monitoring node is $0.25\lambda_L$. The other modeling parameters are as per the default values listed in Table I. The model size is such that excitation from 92 different incident angles equi-spaced over 360° is used for both longitudinal and shear modes. To execute such a model without exploiting the symmetry of the scatterer (i.e., solving 184 load cases for results over 92 incident angles for both modes) takes around 1 h of processing time on a machine with a 22.8 GHz dual-core processor (AMD Opteron) and 8 GB random access memory (RAM), which is part of a Beowulf cluster. The output is the set of *S*-matrices for all four mode combinations at one frequency. For economy of space, only the results for the case of longitudinal incidence and scattering are considered here. The corresponding *S*-matrix is shown in Fig. 7(b). The dark diagonal regions in the *S*-matrix where the scattering amplitude is highest correspond to specular reflections (i.e., $\theta_2 = 180° - \theta_1$) and shadowing (i.e., $\theta_2 = 180° + \theta_1$). The equivalent *S*-matrix for a straight crack of similar length[20] exhibits a similar structure except that the appearance of the specular and shadowing regions is identical. In the *S*-matrix for the irregular crack shown here, the peak in the shadowing region is of greater amplitude and extends over a larger angular range than that in the specular region.

Finally, expression (3) is used to compute the complete scattered longitudinal wave field for the specific case when a point harmonic force of amplitude $[1\ \mathrm{N}, 1\ \mathrm{N}]^T$ located at $(5\lambda_L, 5\lambda_L)$ excites longitudinal waves. Figure 7(c) shows the radial displacement component u_r of the resulting scattered field (the incident field is not shown). Note that the data in this image have been computed using expression (3) from only the information in the *S*-matrix shown in Fig. 7(b), which is accurately represented in the Fourier domain by approximately $60 \times 60 = 3600$ Fourier coefficients, i.e., the S_{mn} values in expression (2). The gray circle in the center of Fig. 7(c) has a diameter $4\lambda_L$, which is equal to the maximum dimension of the scatterer and indicates the region within which the scattered field reconstruction is not valid. The most dominant scattered signal is that labeled shadow, which corresponds to forward scattering. The signal here is in antiphase to the incident wave and corresponds to a shadow region of low amplitude displacement if the total field (i.e., incident plus scattered) is considered. The next largest scattered signal is in the expected direction of the specular reflection from the illuminated surface of the crack. This is a little weaker than the shadow signal as the irregular surface of the crack causes the specular reflection to be somewhat diffused.

V. CONCLUSION

A FE method for generating the elastodynamic *S*-matrices associated with a finite-sized scatterer embedded in an otherwise homogeneous, isotropic medium has been presented.

An important feature of the FE method described here is that only the scatterer and the immediate vicinity of the surrounding medium are modeled. This means that for a scatterer with maximum dimension *D*, the size of a model in terms of degrees of freedom is of the order D^2, and the number of times the model must be run to generate the *S*-matrix is of the order *D* (corresponding to the number of incident angles required to generate the *S*-matrix). The overall computational load is thus of order D^3. In a previous work,[20] the *S*-matrix was generated using source and receiver positions that were in the far-field of the scatterer. As the far-field is of the order D^2, that model size was of the order D^4 and the overall computational load of that method was therefore of the order D^5.

In the examples given here, a FE mesh composed of identical-sized square elements has been used. The major disadvantage of such a mesh is poor matching to any nonrectilinear scatterer shape and a lesser one is that the mesh induces a weak anisotropy with fourfold symmetry. An important advantage of square elements is that the only difference between a model containing a scatterer and a reference model without a scatterer is the scatterer itself. Thus when the results from two such models are subtracted to yield the *S*-matrices, the data in the *S*-matrices are guaranteed to come

from the presence of the scatterer alone. This is not necessarily the case if an automated meshing algorithm is used to generate the meshes with and without a scatterer present. In this case, artifacts in the *S*-matrices may be generated due to changes in the mesh surrounding the scatterer as well as the scatterer itself. The quantification of this effect and methods to mitigate against it (if it is found to be significant) are clearly avenues for further investigation.

ACKNOWLEDGMENTS

The authors thank Bruce Drinkwater for his many helpful discussions on the physical interpretation of data in *S*-matrices. P.D.W. is supported by an EPSRC Advanced Research Fellowship (Grant No. EP/E054951/1).

[1]Y. H. Pao and C. C. Mow, *Diffraction of Elastic Waves and Dynamic Stress Concentrations* (Crane Russak, New York, 1973).
[2]J. D. Achenbach, L. Adler, D. K. Lewis, and H. McMaken, "Diffraction of ultrasonic waves by penny-shaped cracks in metals: Theory and experiment," J. Acoust. Soc. Am. **66**, 1848–1856 (1979).
[3]L. W. Schmerr, *Fundamentals of Ultrasonic NDE—A Modeling Approach* (Plenum, New York, 1998).
[4]A. L. Lopez-Sanchez, H. J. Kim, L. W. Schmerr, and A. Sedov, "Measurement models and scattering models for predicting the ultrasonic pulse-echo response from side-drilled holes," J. Nondestruct. Eval. **24**, 83–96 (2005).
[5]J. E. Gubernatis, E. Domany, J. A. Krumhansl, and M. Huberman, "The Born approximation in the theory of the scattering of elastic waves by flaws," J. Appl. Phys. **48**, 2812–2819 (1977).
[6]L. L. Thompson, "A review of finite element methods for time-harmonic acoustics," J. Acoust. Soc. Am. **119**, 1315–1330 (2006).
[7]J. H. Su, V. V. Varadan, and V. K. Varadan, "Finite element eigenfunction method (FEEM) for elastic (SH) wave scattering," J. Acoust. Soc. Am. **73**, 1499–1504 (1983).
[8]J. A. G. Temple, "Modelling the propagation and scattering of elastic waves in inhomogeneous and anisotropic media," J. Phys. D: Appl. Phys. **21**, 859–874 (1988).
[9]M. Koshiba, S. Karakida, and M. Suzuki, "Finite-element analysis of Lamb waves scattering in an elastic plate waveguide," IEEE Trans. Sonics Ultrason. **31**, 18–25 (1984).
[10]N. N. Kishore, I. Sridhar, and N. G. R. Iyengar, "Finite element modeling of the scattering of ultrasonic waves by isolated flaws," NDT & E Int. **33**, 297–305 (2000).
[11]M. Frehner, S. M. Schamholz, E. H. Saenger, and H. Steeb, "Comparison of finite difference and finite element methods for simulating two-dimensional scattering of elastic waves," Phys. Earth Planet. Inter. **171**, 112–121 (2008).
[12]M. J. S. Lowe and O. Diligent, "Low-frequency reflection characteristics of the s_0 Lamb wave from a rectangular notch in a plate," J. Acoust. Soc. Am. **111**, 64–74 (2002).
[13]M. J. S. Lowe, P. Cawley, J. Y. Kao, and O. Diligent, "The low frequency reflection characteristics of the fundamental antisymmetric Lamb wave a_0 from a rectangular notch in a plate," J. Acoust. Soc. Am. **112**, 2612–2622 (2002).
[14]X. Jian, Y. Fan, R. S. Edwards, and S. Dixon, "Surface-breaking crack gauging with the use of laser-generated Rayleigh waves," J. Appl. Phys. **100**, 064907 (2006).
[15]O. Diligent, T. Grahn, A. Bostrom, P. Cawley, and M. J. S. Lowe, "The low-frequency reflection and scattering of the S_0 Lamb mode from a circular through-thickness hole in a plate: Finite element, analytical and experimental studies," J. Acoust. Soc. Am. **112**, 2589–2601 (2002).
[16]P. Rajagopal and M. J. S. Lowe, "Short range scattering of the fundamental shear horizontal guided wave mode normally incident at a through-thickness crack in an isotropic plate," J. Acoust. Soc. Am. **122**, 1527–1538 (2007).
[17]P. Rajagopal and M. J. S. Lowe, "Scattering of the fundamental shear horizontal guided wave by a part-thickness crack in an isotropic plate," J. Acoust. Soc. Am. **124**, 2895–2904 (2008).
[18]L. Moreau and M. Castaings, "The use of an orthogonality relation for reducing the size of finite element models for 3D guided waves scattering problems," Ultrasonics **48**, 357–366 (2008).
[19]B. W. Drinkwater and P. D. Wilcox, "Ultrasonic arrays for non-destructive evaluation: A review," NDT & E Int. **39**, 525–541 (2006).
[20]J. Zhang, B. Drinkwater, and P. Wilcox, "Defect characterization using an ultrasonic array to measure the scattering coefficient matrix," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **55**, 2254–2265 (2008).
[21]V. Varatharajulu and Y.-H. Pao, "Scattering matrix for elastic waves. I. Theory," J. Acoust. Soc. Am. **60**, 556–566 (1976).
[22]A. Boström and H. Wirdelius, "Ultrasonic probe modeling and nondestructive crack detection," J. Acoust. Soc. Am. **97**, 2836–2848 (1995).
[23]D. Colton, J. Coyle, and P. Monk, "Recent developments in inverse acoustic scattering theory," SIAM Rev. **42**, 369–414 (2000).
[24]F. Simonetti, "Multiple scattering: The key to unravel the subwavelength world from the far-field pattern of a scattered wave," Phys. Rev. E **73**, 036619 (2006).
[25]E. A. Skelton, R. V. Craster, and A. V. Shanin, "Embedding formulae for diffraction by non-parallel slits," Q. J. Mech. Appl. Math. **61**, 93–116 (2007).
[26]H. S. Cabayan, R. C. Murphy, and T. J. Pavlasek, "Numerical stability and near-field reconstruction," IEEE Trans. Antennas Propag. **21**, 346–351 (1973).
[27]E. M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions*, 9th ed. (Dover, New York, 1970), p. 361.
[28]J. D. Achenbach, *Reciprocity in Elastodynamics* (Cambridge University Press, Cambridge, 2003), pp. 51–54.
[29]E. A. Marengo, F. K. Gruber, and F. Simonetti, "Time-reversal music imaging of extended targets," IEEE Trans. Image Process. **16**, 1967–1984 (2007).
[30]M. Drodz, L. Moreau, M. Castaings, M. J. S. Lowe, and P. Cawley, "*Efficient numerical modeling of absorbing regions for boundaries of guided wave problems*," in Ann. Rev. Prog. QNDE, Vol. **25**, edited by D. O. Thompson and D. E. Chimenti (American Institute of Physics, Melville, NY, 2006), pp. 126–133.
[31]R. M. White, "Elastic wave scattering at a cylindrical discontinuity in a solid," J. Acoust. Soc. Am. **30**, 771–785 (1958).

Simplified expressions of the subtracted Kramers–Kronig relations using the expanded forms applied to ultrasonic power-law systems

Joel Mobley[a)]
Department of Physics and Astronomy and National Center for Physical Acoustics, University of Mississippi, University, Mississippi 38677

(Received 11 July 2009; revised 19 October 2009; accepted 22 October 2009)

The Kramers–Kronig (KK) relations are a large class of integral transformations that exploit the broad principle of simple causality in order to link the physical properties of matter and materials. In applications to the complex-valued wavenumber for acoustic propagation, the method of subtractions is used to form convergent integral relations between the phase velocity and the attenuation coefficient. When the method of subtractions is applied in the usual manner, the integrands in the relations become unnecessarily complicated. In this work, an expanded form of the subtracted relations is presented, which is essentially a truncated Taylor series expansion of the Hilbert transforms. The implementation of the relations only requires the explicit evaluation of two simply expressed integrals involving the Hilbert transform kernel. These two integrals determine the values of the other terms in the subtracted relations, demonstrating the computational efficiency of the technique. The method is illustrated analytically through its application to power-law attenuation coefficients and its associated dispersion, which are observed in a wide variety of materials. This approach explicitly shows the central role of the Hilbert transform kernel in the KK relations, which can become obscured in other formulations.
 [DOI: 10.1121/1.3268512]

PACS number(s): 43.35.Cg, 43.20.Hq, 43.35.Bf [RR] Pages: 166–173

I. INTRODUCTION

Kramers–Kronig (KK) relations refer to a large class of integral transformations that emerge from the principle of simple causality and connect the physical properties of matter and materials. The underlying structure of KK relations are the Hilbert transforms (also known as Plemelj formulas).[1,2] In applications to acoustic propagation, particularly at ultrasonic frequencies, the parameter of interest is the complex-valued wavenumber which includes the phase velocity and attenuation coefficient. In forming valid convergent integral relations between the phase velocity and the attenuation coefficient, the method of subtractions[2] is required, and theoretical and experimental evidence suggests that two subtractions are sufficient.[3–5] However, if the method of subtractions is applied in the usual manner, the integrands of these relations can grow unnecessarily complicated. In this work, an expanded form of the subtracted relations is used,[4–6] which subtracts off a truncated Taylor series expansion in terms of the Hilbert integral appropriate to each relation. This approach only requires the explicit evaluation of two relatively simple integrals, from which all of the other terms can be determined. Writing the relations in this expanded form demonstrates the central role of the Hilbert transform in relating the quantities of interest and how the subtraction terms can be determined from these integrals. This method is computationally efficient as it avoids the needlessly complicated integrands of the usual approach. The use of the expanded forms is illustrated through their application to power-law attenuation coefficients (with the exponent y in the range $1 \leq y \leq 2$) and the associated dispersion, which represent a class of behavior observed in a variety of materials (biological tissues and fluids, various plastics, composites, and oils) in the low-megahertz frequency band of the ultrasonic spectrum.[7–11]

The first appearance of Kramers–Kronig relations for acoustic waves was in a paper by Ginzburg[12] printed in English in 1955 (and two years earlier in the USSR). More recent derivations of acoustic KK relations have benefited from the subsequent advances in the understanding of dispersion relations and their foundations. The approach used in this paper evolved from earlier work by O'Donnell *et al.*,[13,14] Weaver and Pao,[3] and Aki and Richards[15] while drawing on concepts described by Nussenzveig.[2] A brief review of the past uses of KK in acoustics is given in Ref. 16. More recently, KK relations (and the power-law model) have been important in determining the physical nature of propagation in bone[17–19] and have been integrated into techniques for measuring shear wave properties in soft tissues.[20] They are also used in the photoacoustic determination of attenuation and dispersion.[21] Other applications include the time-domain predictions of propagation in layered materials[22] and measurement of complex moduli in the vibrational domain.[23] The power-law results have also been extended into the sonic spectrum for the study of porous materials.[24] Subtracted KK relations have been applied to determine acoustic dispersion relations for air.[25] The most extensive use of KK relations

[a)]Author to whom correspondence should be addressed. Electronic mail: jmobley@olemiss.edu

(including their subtracted forms) is in optical spectroscopy where they have been applied in wide variety of contexts, including extensions to non-linear regimes.[26] Following the example of optics, the wide ranging and frequent use of KK in acoustics will no doubt drive new formulations and applications of the relations.

In previous work, the twice subtracted forms of the acoustic KK relations between the attenuation coefficient $\alpha(\omega)$ and phase velocity $c_p(\omega)$ have been successfully applied to finite bandwidth data from suspensions of microspheres[16] and encapsulated microbubbles[4,5] which exhibit resonant structures in their spectra. KK relations have also been derived that can predict the group velocity directly from the attenuation coefficient for these same types of suspensions.[6,27,28] These relations were all implemented using expanded forms of the KK relations, which are straightforward to apply to finite bandwidths and do not require parameters or extrapolations outside the experimental spectrum. However, the finite bandwidth approach is not well suited when the attenuation and dispersion are monotonic in frequency[29] as there are no resonant structures to anchor the integrals within a localized band. In that case, these data are fitted to mathematical models of dispersion and attenuation that are verified to be causally consistent with KK. Within limited experimental windows in the ultrasonic spectrum, many materials have been found to exhibit attenuation which varies with frequency as a power-law, $\alpha(\omega)=\alpha_0\omega^y$, with exponents ranging from 1 to 2. In these materials the dispersion in the phase velocity has been found to follow a causally consistent form prescribed by acoustic KK relations (although the dispersion result was first derived with an alternate technique, known as the time-causal method[30]). This paper focuses on the power-law attenuation coefficient and its causally prescribed phase velocity since both can be expressed analytically. With these analytic expressions, the use of the expanded forms of the subtracted KK relations is demonstrated. The expanded forms can greatly simplify the integrals in these relations, providing computational efficiency and keeping the central role of the Hilbert transform kernel from being obscured. A key point is that only two integrals (or just a single integral if one chooses to combine them), simple in form, are required to evaluate the four or six terms in the expanded forms of the once subtracted or twice subtracted relations, respectively. In fact, this should be true for even higher-order subtractions. Another important idea is that even divergent integrals contain finite, physically relevant results when the divergent quantities are segregated. (This is similar in spirit to a regularization method used in a renormalization procedure in quantum electrodynamics,[31] which extracts finite results from divergent integrals.) In the work to follow it is shown that the true physical content of the KK transforms can be isolated from the divergent terms since the desired outcomes are independent of terms involving the limits of integration.

An alternate KK method has been used that is based on the direct relationship between the complex speed of sound and the dynamic compressibility.[13,32,33] However, there is no clean separation between attenuation and velocity in that approach, although the small attenuation approximation (which is usually justified) does simplify them to a degree. The other formulation may have its advantages in certain applications; however, the relations described here provide for an unambiguous separation of $c_p(\omega)$ and $\alpha(\omega)$ without any approximations and are more straightforward to apply to dispersion and attenuation data. For the power-law problem, the subtracted relations are a more appropriate choice.

II. THEORY

A. Ultrasonic attenuation and dispersion in power-law systems

In the frequency domain, the linear transport of ultrasonic waves along a straight path in an isotropic medium over a distance D can be accounted for in general by a transfer function of the form

$$H(\omega,D)=\exp[iK(\omega)D], \tag{1}$$

where

$$K(\omega)=\omega/c_p(\omega)+i\alpha(\omega) \tag{2}$$

is the complex wavenumber, $\alpha(\omega)$ is the attenuation coefficient, and $c(\omega)$ is the phase velocity. Experimentally, a power-law form of the attenuation coefficient

$$\alpha(\omega)=\alpha_0\omega^y \quad \text{where } 1\leqslant y\leqslant 2 \tag{3}$$

has been found to provide an accurate fit to broadband data from a variety of materials within the respective measurement bandwidths. When Eq. (3) covers the entire frequency spectrum, the phase velocity takes the causally consistent form[8,10]

$$\frac{1}{c_p(\omega)}-\frac{1}{c_p(\omega_0)}=\alpha_0\tan\left(y\frac{\pi}{2}\right)(\omega^{y-1}-\omega_0^{y-1}), \quad 1<y\leqslant 2 \tag{4a}$$

$$=-\alpha_0\frac{2}{\pi}\ln\frac{\omega}{\omega_0}, \quad y=1. \tag{4b}$$

[Equation (4b) can be obtained by taking the $y\to 1$ limit of Eq. (4a).[34]] Note that both $\alpha(\omega)$ and $c(\omega)$ are even functions, a property derived from the fact that $H(\omega,D)$ is the Fourier transform of a real function. [Since negative frequencies are not explicitly considered, no absolute value signs are used in Eqs. (3) and (4)].

B. Subtracted KK relations for the complex wavenumber

The transfer function $H(\omega,D)$ is the Fourier transform of a causal, square-integrable function [i.e., the impulse response $h_D(t)$], which implies via Titchmarsh's theorem[2] that its real and imaginary parts form a Hilbert transform pair.

$$\mathrm{Im}[H(\omega)]=-\frac{1}{\pi}P\int_{-\infty}^{\infty}\frac{\mathrm{Re}[H(x)]}{x-\omega}dx, \tag{5a}$$

$$\mathrm{Re}[H(\omega)] = \frac{1}{\pi} P \int_{-\infty}^{\infty} \frac{\mathrm{Im}[H(x)]}{x-\omega} dx. \tag{5b}$$

Since $h_D(t)$ is real, the components of $H(\omega,D)$ exhibit definite parity, which in turn permit the mapping of the negative frequency components of the Hilbert integrals to positive frequencies. The resulting transforms are labeled as Kramers–Kronig relations. For example, Eq. (5a) can be rewritten as

$$\mathrm{Im}[H(\omega)] = \lim_{\Omega\to\infty}\left\{ -\frac{1}{\pi} P \int_{0}^{\Omega} \frac{\mathrm{Re}[H(x)]}{x-\omega} dx - \frac{1}{\pi} P \int_{-\Omega}^{0} \frac{\mathrm{Re}[H(x)]}{x-\omega} dx \right\} \tag{6a}$$

$$= \lim_{\Omega\to\infty}\left\{ -\frac{1}{\pi} P \int_{0}^{\Omega} \frac{\mathrm{Re}[H(x)]}{x-\omega} dx + \frac{1}{\pi} \int_{0}^{\Omega} \frac{\mathrm{Re}[H(x)]}{x+\omega} dx \right\} \tag{6b}$$

$$= -\frac{2\omega}{\pi} P \int_{0}^{\infty} \frac{\mathrm{Re}[H(x)]}{x^2-\omega^2} dx. \tag{6c}$$

In a similar fashion,

$$\mathrm{Re}[H(\omega)] = \lim_{\Omega\to\infty}\left\{ \frac{1}{\pi} P \int_{0}^{\Omega} \frac{\mathrm{Im}[H(x)]}{x-\omega} dx + \frac{1}{\pi} \int_{0}^{\Omega} \frac{\mathrm{Im}[H(x)]}{x+\omega} dx \right\} \tag{7a}$$

$$= \frac{2}{\pi} P \int_{0}^{\infty} \frac{x\,\mathrm{Im}[H(x)]}{x^2-\omega^2} dx. \tag{7b}$$

The most compact forms are those of Eqs. (6c) and (7b) but the least complicated integrands are those of Eqs. (6b) and (7a). In the expanded form discussed below, it is these forms with the less complicated integrands that are used.

In ultrasonics, the parameters of interest are typically those directly related to propagation, namely, the attenuation coefficient and phase velocity. To form relations between these quantities requires taking the logarithm of the transfer function since $iK(\omega)=\ln H(\omega,D)/D$. However, $K(\tilde{\omega})$ does not satisfy the square integrability condition of Titchmarsh. The method of subtractions provides a way through this difficulty, offering a prescription for forming an analytic and square integrable function from $K(\tilde{\omega})$ and its derivatives at a single frequency ω_0 known as the subtraction frequency. The general procedure used in this work (as well as an alternate form of subtraction method) is described by Nussenzveig.[2] Based on both empirical and analytic evidence,[3–6,27,28] two subtractions appear to be sufficient for establishing a Hilbert transform pair from the acoustic complex wavenumber, although one subtraction is sufficient for the $\alpha(\omega)$ determination in the power-law case.[29] The twice subtracted form for the wavenumber is

$$\frac{K(\omega) - K(\omega_0) - (\omega-\omega_0)\frac{d}{d\omega}K(\omega)\Big|_{\omega=\omega_0}}{(\omega-\omega_0)^2}. \tag{8}$$

This function obeys Titchmarsh's theorem, and thus its real and imaginary parts [which are expressed in terms of $\alpha(\omega)$ and $c_p(\omega)$] form a Hilbert transform pair. The central difficulty of the subtracted approach is that for the general case, the resulting expressions can grow quite involved. For example, the relation for the phase velocity takes the following form:

$$\frac{\omega}{c_p(\omega)} - \frac{\omega_0}{c_p(\omega_0)} - (\omega-\omega_0)\frac{d}{d\omega}\left.\frac{\omega}{c_p(\omega)}\right|_{\omega=\omega_0} = \frac{(\omega-\omega_0)^2}{\pi} P \int_{-\infty}^{\infty} \frac{\alpha(x)-\alpha(\omega_0)-(x-\omega_0)\alpha'(\omega_0)}{(x-\omega_0)^2(x-\omega)} dx. \tag{9}$$

The next step would be to use the parity properties of the attenuation coefficient to map the integration interval from the negative to the positive frequency axis. In general, the integral term would become unnecessarily complicated, as will be seen below. The aim of the method of subtractions is to ensure convergence, but applied in the brute force manner it obscures the underlying simplicity of the Hilbert transforms. Setting the subtraction frequency to 0 can greatly simplify the integrand in some cases, although in practice, the parameter values at zero frequency must usually be extrapolated from higher frequency data. If one is using these relations with experimental data, then it might be preferable to stick with a subtraction frequency from within the measurement bandwidth, whereas often it is set to 0 or ∞ in other treatments. In our work, we keep the subtraction frequency general as an expression of this viewpoint. In order to simplify the evaluation of the KK integrals, the expanded form was developed. If one does a partial fraction expansion of the integrand in Eq. (9), the following form results:

$$\lim_{\Omega\to\infty}\left[I_\alpha(\omega,\Omega) - I_\alpha(\omega_0,\Omega) - (\omega-\omega_0)\left.\frac{d}{d\omega} I_\alpha(\omega,\Omega)\right|_{\omega=\omega_0}\right], \tag{10}$$

where

$$I_\alpha(\omega,\Omega) = \frac{1}{\pi}\int_0^{\Omega} \frac{\alpha(x)-\alpha(\omega)}{x-\omega} dx - \frac{1}{\pi}\int_0^{\Omega} \frac{\alpha(x)-\alpha(\omega)}{x+\omega} dx. \tag{11}$$

The most important issue to be addressed is that of convergence—the individual integrals are no longer convergent when the upper integration variable is allowed to go to infinity. The strategy to maintain convergence is implicit in the form of Eq. (10)—each of the individual terms is evaluated up to some arbitrary high frequency Ω, and then after all of the integrals have been evaluated, the limit of $\Omega\to\infty$ is taken. The key concept here is implicit in the partial fraction expansion: that an expression with a polynomial of order N

in the denominator can be written as the sum of up to N terms that are each of lower order in their respective denominators. Thus subtracting terms can assure convergence in the same way as increasing the order of the integration variable in the denominator of a single term.

The main advantage of the expanded form is that only two of the six integrals in Eq. (10) need be explicitly evaluated; the second integral pair $I_\alpha(\omega_0,\Omega)$ is merely the first pair evaluated at a specific point (the subtraction frequency ω_0), while the final integral pair, $(d/d\omega)I_\alpha(\omega,\Omega)|_{\omega=\omega_0}$, is simply the first pair differentiated with respect to frequency and evaluated at ω_0. Thus all the work required to evaluate the more complicated form of Eq. (9) comes down to the evaluation of $I_\alpha(\omega,\Omega)$ with its two relatively simple integrands.

After rearranging the terms on the left hand side by performing the differentiation of the term $\omega/c_p(\omega)$, the twice subtracted KK relation for the phase velocity in the expended form is given below,

$$\omega\left(\frac{1}{c_p(\omega)}-\frac{1}{c_p(\omega_0)}\right)-(\omega-\omega_0)\omega_0\left.\frac{d}{d\omega}\frac{1}{c_p(\omega)}\right|_{\omega=\omega_0}$$
$$=\lim_{\Omega\to\infty}\left[I_\alpha(\omega,\Omega)-I_\alpha(\omega_0,\Omega)-(\omega-\omega_0)\left.\frac{d}{d\omega}I_\alpha(\omega,\Omega)\right|_{\omega=\omega_0}\right], \tag{12}$$

where I_α is given in Eq. (11).

For the conjugate relation determining the attenuation coefficient, only one subtraction turns out to be necessary for a power-law with exponent $y<2$. The once-subtracted relation is

$$\alpha(\omega)=\alpha(\omega_0)+\lim_{\Omega\to\infty}[I_c(\omega,\Omega)-I_c(\omega_0,\Omega)], \tag{13}$$

where

$$I_c(\omega,\Omega)=-\frac{1}{\pi}\int_0^\Omega\frac{x/c_p(x)-\omega/c_p(\omega)}{x-\omega}dx-\frac{1}{\pi}\int_0^\Omega\frac{x/c_p(x)+\omega/c_p(\omega)}{x+\omega}dx. \tag{14}$$

Note that the attenuation coefficient in the power-law model given in Eq. (3) does not depend on the subtraction frequency. Looking at Eq. (13), the only term on the right hand side (RHS) that is independent of ω_0 is the first term in the brackets. So all of the variation in $\alpha(\omega)$ is determined from $I_c(\omega,\Omega)$; the $I_c(\omega_0,\Omega)$ term will have double duty—it will cancel out the divergence in $I_c(\omega,\Omega)$ as $\Omega\to\infty$ and it will also cancel out the constant term $\alpha(\omega_0)$. So in that respect, the important information about $\alpha(\omega)$ is determined as part of the integral pair $I_c(\omega,\Omega)$, which is essentially the Hilbert transform of $\omega/c_p(\omega)$.

If we look again at the form of the dispersion in the model [Eqs. (4a) and (4b)], one can see that it depends on both ω and the subtraction frequency ω_0. To isolate the dispersion on the left hand side of the relation, Eq. (12) can be rewritten as

$$\frac{1}{c_p(\omega)}-\frac{1}{c_p(\omega_0)}=\frac{1}{\omega}\lim_{\Omega\to\infty}\left[I_\alpha(\omega,\Omega)-I_\alpha(\omega_0,\Omega)+(\omega-\omega_0)\left\{\omega_0\left.\frac{d}{d\omega}\left(\frac{1}{c_p(\omega)}-I_\alpha(\omega,\Omega)\right)\right|_{\omega=\omega_0}\right\}\right]. \tag{15}$$

In contrast to the attenuation relation, all three of the I_α terms are required to get the desired form of the dispersion. This will be shown explicitly later in the paper.

III. EVALUATION OF THE INTEGRALS

The central part of the Kramers–Kronig calculation is the evaluation of the integral terms. In Secs. III A and III B the integrals in $I_c(\omega,\Omega)$ and $I_\alpha(\omega,\Omega)$ are evaluated for the power-law model.

A. Evaluation of $I_c(\omega,\Omega)$

In the relation for the attenuation coefficient, the integral pair that must be evaluated is defined in Eq. (14). With the phase velocity of the power-law system, this becomes

$$I_c(\omega,\Omega)=-\frac{1}{\pi}P\int_0^\Omega\frac{\frac{x-\omega}{c_p(\omega)}+\alpha_0\tan\left(y\frac{\pi}{2}\right)(x^y-x\omega^{y-1})}{x-\omega}dx-\frac{1}{\pi}\int_0^\Omega\frac{\frac{x+\omega}{c_p(\omega)}+\alpha_0\tan\left(y\frac{\pi}{2}\right)(x^y-x\omega^{y-1})}{x+\omega}dx \tag{16a}$$

$$=-\frac{2}{\pi}\frac{\Omega}{c_p(\omega)}+\frac{\alpha_0}{\pi}\tan\left(y\frac{\pi}{2}\right)\left[\omega^{y-1}P\int_0^\Omega\frac{x}{x-\omega}dx+\omega^{y-1}\int_0^\Omega\frac{x}{x+\omega}dx-\int_0^\Omega\frac{x^y}{x-\omega}dx-\int_0^\Omega\frac{x^y}{x+\omega}dx\right]. \tag{16b}$$

After evaluating the first two integrals, $\int_0^\Omega(x/x\pm\omega)dx=\Omega\mp\omega\ln(\Omega\pm\omega)$, the results are

$$I_c(\omega,\Omega)=-\frac{2}{\pi}\frac{\Omega}{c(\omega)}+\frac{\alpha_0}{\pi}\tan\left(y\frac{\pi}{2}\right)\times\omega^{y-1}\left(2\Omega+\omega\ln\frac{1-\omega/\Omega}{1+\omega/\Omega}\right)-\frac{\alpha_0}{\pi}\tan\left(y\frac{\pi}{2}\right)\times\left(\int_0^\Omega\frac{x^y}{x-\omega}dx+\int_0^\Omega\frac{x^y}{x+\omega}dx\right). \tag{17}$$

The integrals left to be evaluated are

$$P\int_0^{\Omega}\frac{x^y}{x-\omega}dx+\int_0^{\Omega}\frac{x^y}{x+\omega}dx. \tag{18}$$

At this point the two integrals could be combined to one, but to keep with the spirit of the approach they will be evaluated separately. Although only the first term explicitly requires a principal value type of integration for $\omega>0$, we will work both integrals using principal value type of limiting procedure. The first step is to replace the denominators with the geometric series $(1\pm r)^{-1}=\Sigma_{n=0}^{\infty}(\mp 1)^n r^n$. Since these series are only convergent for $r<1$, the integration interval is broken up into two regimes, $x<\omega$ and $x>\omega$. The denominator is rewritten in each case so that a geometric series in the appropriate variable can be applied. Since the series are not convergent when $x=\omega$, the integrals are only worked to within $\pm\delta$ of the frequency ω. The limit of $\delta\to 0$ will be taken later when the series convergence is no longer of concern. This limiting process is the definition of a principal value (PV) integral, although it is only explicitly necessary for the first integral which diverges at $x=\omega$. However, using this series approach, it is necessary to use the PV method for evaluating both. The steps are shown below,

$$\begin{aligned}&\lim_{\delta\to 0}\left[\int_0^{\omega-\delta}\frac{x^y}{x-\omega}dx+\int_{\omega+\delta}^{\Omega}\frac{x^y}{x-\omega}dx+\int_0^{\omega-\delta}\frac{x^y}{x+\omega}dx+\int_{\omega+\delta}^{\Omega}\frac{x^y}{x+\omega}dx\right]\\&=\lim_{\delta\to\infty}\left[-\frac{1}{\omega}\int_0^{\omega-\delta}\frac{x^y}{1-x/\omega}dx+\int_{\omega+\delta}^{\Omega}\frac{x^{y-1}}{1-\omega/x}dx+\frac{1}{\omega}\int_0^{\omega-\delta}\frac{x^y}{1+x/\omega}dx+\int_{\omega+\delta}^{\Omega}\frac{x^{y-1}}{1+\omega/x}dx\right].\end{aligned} \tag{19}$$

Using the geometric series substitutions appropriate to each of the four integrals results in the following:

$$\begin{aligned}&\lim_{\delta\to 0}\left[-\frac{1}{\omega}\int_0^{\omega-\delta}\sum_{n=0}^{\infty}\frac{x^{y+n}}{\omega^n}dx+\int_{\omega+\delta}^{\Omega}\sum_{n=0}^{\infty}x^{y-n-1}\omega^n dx\right.\\&\left.+\frac{1}{\omega}\int_0^{\omega-\delta}\sum_{n=0}^{\infty}(-1)^n\frac{x^{y+n}}{\omega^n}dx+\int_{\omega+\delta}^{\Omega}\sum_{n=0}^{\infty}(-1)^n x^{y-n-1}\omega^n dx\right].\end{aligned} \tag{20}$$

The integrals are now condensed from 4 to 2,

$$\begin{aligned}&=\lim_{\delta\to 0}\left[-\int_0^{\omega-\delta}\sum_{n=0}^{\infty}[1-(-1)^n]\frac{x^{y+n}}{\omega^{n+1}}dx\right.\\&\left.+\int_{\omega+\delta}^{\Omega}\sum_{n=0}^{\infty}[1+(-1)^n]x^{y-n-1}\omega^n dx\right].\end{aligned} \tag{21}$$

Since the first series will have only terms for odd n, and the second for even n, they can be expressed as

$$\lim_{\delta\to 0}\left[-2\int_0^{\omega-\delta}\sum_{n=0}^{\infty}\frac{x^{y+2n+1}}{\omega^{2n+2}}dx+2\int_{\omega+\delta}^{\Omega}\sum_{n=0}^{\infty}x^{y-2n-1}\omega^{2n}dx\right]. \tag{22}$$

Upon evaluation, this becomes

$$\begin{aligned}&\lim_{\delta\to 0}\left[-2\sum_{n=0}^{\infty}\frac{1}{2n+2+y}\frac{(\omega-\delta)^{y+2n+2}}{\omega^{2n+2}}\right.\\&\left.+2\Omega^y\sum_{n=0}^{\infty}\frac{1}{y-2n}\frac{\omega^{2n}}{\Omega^{2n}}-2\sum_{n=0}^{\infty}\frac{1}{y-2n}(\omega+\delta)^{y-2n}\omega^{2n}\right].\end{aligned} \tag{23}$$

In the limit of $\delta\to 0$, the two series with δ are not convergent on their own, so they are combined into a single series before the limit is taken.

$$\begin{aligned}&\lim_{\delta\to 0}\left[-2\omega^y\sum_{n=0}^{\infty}\left(\frac{1}{2n+2+y}(1-\delta/\omega)^{y+2n+2}\right.\right.\\&\left.\left.-\frac{1}{2n-y}(1+\delta/\omega)^{y-2n}\right)-2\Omega^y\sum_{n=0}^{\infty}\frac{1}{2n-y}\frac{\omega^{2n}}{\Omega^{2n}}\right],\end{aligned} \tag{24}$$

which yields the following result:

$$=-2\omega^y\sum_{n=0}^{\infty}\left(\frac{1}{2n+2+y}-\frac{1}{2n-y}\right)-2\Omega^y\sum_{n=0}^{\infty}\frac{(\omega/\Omega)^{2n}}{2n-y}. \tag{25}$$

The first series can be rewritten as

$$2\sum_{n=0}^{\infty}\left(\frac{1}{2n+2+y}-\frac{1}{2n-y}\right)=\frac{2}{y}+y\sum_{n=1}^{\infty}\left(\frac{1}{(y/2)^2-n^2}\right). \tag{26}$$

Using the identity[35]

$$\cot(a\pi)=\frac{1}{a\pi}+\frac{2a}{\pi}\sum_{n=1}^{\infty}\left(\frac{1}{a^2-n^2}\right), \tag{27}$$

the result becomes

$$\begin{aligned}&P\int_0^{\Omega}\frac{x^y}{x-\omega}dx+\int_0^{\Omega}\frac{x^y}{x+\omega}dx\\&=-\pi\omega^y\cot\left(y\frac{\pi}{2}\right)-2\Omega^y\sum_{n=0}^{\infty}\frac{1}{2n-y}\frac{\omega^{2n}}{\Omega^{2n}}.\end{aligned} \tag{28}$$

Plugging this back into Eq. (17) yields

$$\begin{aligned}I_c(\omega,\Omega)=\alpha_0\omega^y-\frac{2}{\pi}\frac{\Omega}{c(\omega)}+\frac{\alpha_0}{\pi}\tan\left(y\frac{\pi}{2}\right)\\\times\left(2\Omega\omega^{y-1}+\omega^y\ln\frac{1-\omega/\Omega}{1+\omega/\Omega}\right.\\\left.+2\Omega^y\sum_{n=0}^{\infty}\frac{(\omega/\Omega)^{2n}}{2n-y}\right).\end{aligned} \tag{29}$$

Note that the first term of the result is independent of the

variable Ω and has the functional form of the power-law in frequency we are after. We can expect the Ω dependent terms to cancel out when this result is put back into Eq. (13), and the limit is taken. Thus the physical content of the computation is independent of Ω, which is a parameter that can grow arbitrarily large.

Before proceeding, we can simplify our expressions by eliminating the terms that will become arbitrarily small on their own when $\Omega \gg \omega$,

$$\ln \frac{1-\omega/\Omega}{1+\omega/\Omega} \to \ln 1 = 0$$

and

$$2\Omega^y \sum_{n=0}^{\infty} \frac{(\omega/\Omega)^{2n}}{2n-y} \to -\frac{2\Omega^y}{y}. \tag{30}$$

Thus to order y in Ω,

$$I_c(\omega,\Omega) \to \alpha_0\omega^y - \frac{2}{\pi}\frac{\Omega}{c_p(\omega)} + \frac{\alpha_0}{\pi}\tan\left(y\frac{\pi}{2}\right) \times\left(2\Omega\omega^{y-1} - \frac{2\Omega^y}{y}\right) \quad (\Omega \gg \omega). \tag{31}$$

Later these results will be plugged into the KK relation for $\alpha(\omega)$.

B. Evaluation of $I_\alpha(\omega,\Omega)$

The integral pair for the KK relation for dispersion,

$$I_\alpha(\omega,\Omega) = \frac{1}{\pi}\int_0^\Omega \frac{\alpha(x)-\alpha(\omega)}{x-\omega}dx - \frac{1}{\pi}\int_0^\Omega \frac{\alpha(x)-\alpha(\omega)}{x+\omega}dx, \tag{32}$$

results in the following with the power-law attenuation inserted:

$$= \frac{\alpha_0}{\pi}P\int_0^\Omega \frac{x^y}{x-\omega}dx - \frac{\alpha_0}{\pi}\int_0^\Omega \frac{x^y}{x+\omega}dx + \frac{\alpha_0\omega^y}{\pi}\ln\frac{\Omega+\omega}{\Omega-\omega}. \tag{33}$$

The two integrals left to be evaluated can be performed using the geometric series and limiting procedure in the same manner as before with the following results:

$$P\int_0^\Omega \frac{x^y}{x-\omega}dx - \int_0^\Omega \frac{x^y}{x+\omega}dx = 4y\omega^y\sum_{n=0}^{\infty}\frac{1}{(2n+1)^2-y^2} - 2\Omega^y\sum_{n=0}^{\infty}\frac{(\omega/\Omega)^{2n+1}}{2n+1-y}. \tag{34}$$

Using the identity[35]

$$\tan\left(a\frac{\pi}{2}\right) = \frac{4a}{\pi}\sum_{n=0}^{\infty}\frac{1}{(2n+1)^2-a^2}, \tag{35}$$

the result can be written as

$$P\int_0^\Omega \frac{x^y}{x-\omega}dx - \int_0^\Omega \frac{x^y}{x+\omega}dx = \pi\omega^y\tan\left(y\frac{\pi}{2}\right) - 2\Omega^y\sum_{n=0}^{\infty}\frac{(\omega/\Omega)^{2n+1}}{2n+1-y}. \tag{36}$$

The result for I_α is then

$$I_\alpha(\omega,\Omega) = \alpha_0\omega^y\tan\left(y\frac{\pi}{2}\right) - \frac{2\alpha_0\Omega^y}{\pi}\sum_{n=0}^{\infty}\frac{(\omega/\Omega)^{2n+1}}{2n+1-y} + \frac{\alpha_0\omega^y}{\pi}\ln\frac{1+\omega/\Omega}{1-\omega/\Omega}. \tag{37}$$

The KK relation for dispersion will also require a derivative term

$$\left.\frac{d}{d\omega}I_\alpha(\omega,\Omega)\right|_{\omega=\omega_0} = \alpha_0 y\omega_0^{y-1}\tan\left(y\frac{\pi}{2}\right) - \frac{2\alpha_0\Omega^{y-1}}{\pi}\sum_{n=0}^{\infty}\frac{2n+1}{2n+1-y}(\omega_0/\Omega)^{2n} + \frac{\alpha_0 y\omega_0^{y-1}}{\pi}\ln\frac{1+\omega_0/\Omega}{1-\omega_0/\Omega} + \frac{\alpha_0\omega_0^y}{\pi}\left(\frac{1}{\Omega+\omega_0}-\frac{1}{\Omega-\omega_0}\right). \tag{38}$$

As before, the physical quantity of interest comes from the first term which is independent of the upper limit of integration. These expressions can be simplified by eliminating terms that become arbitrarily small when $\Omega \gg \omega$,

$$I_\alpha(\omega,\Omega) \to \alpha_0\omega^y\tan\left(y\frac{\pi}{2}\right) - \frac{2\alpha_0\omega}{\pi}\frac{\Omega^{y-1}}{1-y}, \tag{39}$$

$$\left.\frac{d}{d\omega}I_\alpha(\omega,\Omega)\right|_{\omega=\omega_0} \to \alpha_0 y\omega_0^{y-1}\tan\left(y\frac{\pi}{2}\right) - \frac{2\alpha_0\Omega^{y-1}}{\pi}\frac{1}{1-y}. \tag{40}$$

IV. EVALUATIONS OF THE KRAMERS–KRONIG RELATIONS

A. KK determination of the attenuation coefficient

Using the results from Sec. III A, the KK prediction for $\alpha(\omega)$ can be evaluated as

$$\begin{aligned}\alpha(\omega) &= \alpha(\omega_0) + \lim_{\Omega\to\infty}[I_c(\omega,\Omega) - I_c(\omega_0,\Omega)] \\ &= \alpha_0\omega_0^y + \alpha_0\omega^y - \alpha_0\omega_0^y \\ &\quad + \lim_{\Omega\to\infty}\left[-\frac{2\Omega}{\pi}\left\{\left(\frac{1}{c(\omega)}-\frac{1}{c(\omega_0)}\right) - \alpha_0\tan\left(y\frac{\pi}{2}\right)(\omega^{y-1}-\omega_0^{y-1})\right\} + \frac{2\alpha_0}{\pi}\tan\left(y\frac{\pi}{2}\right)\left(\frac{\Omega^y}{y}-\frac{\Omega^y}{y}\right)\right].\end{aligned} \tag{41}$$

The first thing to notice is that the subtracted term kills off the Ω^y divergence, as shown in the last term in the above expression. The terms grouped in the curly bracket are together identically zero by Eq. (4a), and the desired result is obtained,

$$\alpha(\omega) = \alpha_0 \omega^y. \tag{42}$$

B. KK determination of the dispersion

The determination of the dispersion begins with Eq. (15′)

$$\frac{1}{c_p(\omega)} - \frac{1}{c_p(\omega_0)} = \frac{1}{\omega} \lim_{\Omega\to\infty} \left[I_\alpha(\omega,\Omega) - I_\alpha(\omega_0,\Omega) - (\omega-\omega_0) \times \left\{ \left. \frac{d}{d\omega} I_\alpha(\omega,\Omega) \right|_{\omega=\omega_0} - \omega_0 \left. \frac{d}{d\omega} \frac{1}{c_p(\omega)} \right|_{\omega=\omega_0} \right\} \right] \tag{15′}$$

To begin, the first two terms in the square brackets are examined in the limit for $\Omega \gg \omega, \omega_0$ using Eq. (39),

$$I_\alpha(\omega,\Omega) - I_\alpha(\omega_0,\Omega) \to \alpha_0 \tan\left(y\frac{\pi}{2}\right)(\omega^y - \omega_0^y) - \frac{2\alpha_0}{\pi}(\omega-\omega_0)\frac{\Omega^{y-1}}{1-y}. \tag{43}$$

The derivative terms in the curly brackets are evaluated in the same limit,

$$\lim_{\Omega\gg\omega} \left. \frac{d}{d\omega} I_\alpha(\omega,\Omega) \right|_{\omega=\omega_0} - \omega_0 \left. \frac{d}{d\omega} \frac{1}{c_p(\omega)} \right|_{\omega=\omega_0} = -\frac{2\alpha_0 \Omega^{y-1}}{\pi}\frac{1}{1-y} + \alpha_0 \tan\left(y\frac{\pi}{2}\right)\omega_0^{y-1}. \tag{44}$$

Putting it all together for the bracketed terms on the RHS results in

$$\alpha_0 \tan\left(y\frac{\pi}{2}\right)(\omega^y - \omega_0^y) - \frac{2\alpha_0}{\pi}(\omega-\omega_0)\frac{\Omega^{y-1}}{1-y} + (\omega-\omega_0)\frac{2\alpha_0\Omega^{y-1}}{\pi}\frac{1}{1-y} - (\omega-\omega_0)\alpha_0 \tan\left(y\frac{\pi}{2}\right)\omega_0^{y-1}. \tag{45}$$

The two terms involving Ω cancel out, leaving

$$\alpha_0 \tan\left(y\frac{\pi}{2}\right)(\omega^y - \omega_0^y - \omega\omega_0^{y-1} + \omega_0^y) = \omega\alpha_0 \tan\left(y\frac{\pi}{2}\right)(\omega^{y-1} - \omega_0^{y-1}). \tag{46}$$

Upon dividing by ω, the desired result is obtained,

$$\frac{1}{c_p(\omega)} - \frac{1}{c_p(\omega_0)} = +\frac{1}{\omega}\left[\omega\alpha_0 \tan\left(y\frac{\pi}{2}\right)(\omega^{y-1} - \omega_0^{y-1}) \right] = \alpha_0 \tan\left(y\frac{\pi}{2}\right)(\omega^{y-1} - \omega_0^{y-1}). \tag{47}$$

V. DISCUSSION

The main point of this work is to demonstrate an alternate method of formulating and interpreting subtracted Kramers–Kronig relations. The expanded form of KK explicitly separates the terms, employing an arbitrary upper limit of integration in order to stave off the divergence until the integrals have been evaluated. It avoids the evaluation of unnecessarily complicated integrals, as the terms due to subtractions are shown to be easily calculable from the simple Hilbert transform integral (divided into two parts to avoid explicitly integrating over negative frequencies). One could describe this approach as computationally efficient since the integrals that must be evaluated are as simple as they could possibly be in a KK formulation between $\alpha(\omega)$ and $c_p(\omega)$, as seen in Eqs. (11) and (14). The expanded method further reveals the central role of the Hilbert transform in Kramers–Kronig transformations, as most compact expressions of KK relations (i.e., single integral formulations) sacrifice the simplicity and transparency of this approach. The compact forms, which are convergent by construction, give rise to the expanded relations via a partial fraction expansion. Breaking the partial fraction terms into the separate integrals of the expanded form requires the upper limit of $\Omega\to\infty$ not to be taken until all of the individual terms have been evaluated to ensure the overall convergence. The physical quantities of interest are completely isolated from the terms involving Ω, and thus their forms are not dependent on the taking of the infinite limit.

One interesting aspect of these relations is that the determination of dispersion requires two subtractions not just to maintain convergence but to obtain the correct result. Here the subtracted terms do contribute in a non-trivial way which is related to the fact that the dispersion is fundamentally dependent on both ω and ω_0. This is not only seen in the analytical solution of the power-law problem but also in the applications to experimental data from suspensions of microspheres and microbubbles. The suspensions data have resonances in their attenuation and dispersion spectra, and the KK predictions for these localized modes are highly accurate even with integration restricted to the experimental bandwidths. In those systems, in determining the dispersion from attenuation, the twice subtracted KK relation is the lowest-order formulation that provides a valid result. So in that case, the extra terms not only assure convergence but are necessary to get the correct answer for the dispersion. Note that this process is general and could be extended to cover relations with any higher number of derivative type subtractions.

VI. CONCLUSIONS

In this work, an alternate procedure for using derivative-type subtracted Kramers–Kronig relations has been demonstrated. The expanded forms greatly simplify the integrals that must be evaluated and also make evaluating the subtracted terms trivial, providing computational efficiency. These integrals are essentially Hilbert transforms and explicitly show the central role they play in Kramers–Kronig rela-

tions. The merits of this technique were illustrated using the power-law attenuation model relevant to ultrasonic propagation.

[1]R. N. Bracewell, *The Fourier Transform and Its Applications* (McGraw-Hill, New York, 1986).
[2]H. M. Nussenzveig, *Causality and Dispersion Relations* (Academic, New York, 1972).
[3]R. L. Weaver and Y.-H. Pao, "Dispersion relations for linear wave propagation in homogeneous and inhomogeneous media," J. Math. Phys. **22**, 1909–1918 (1981).
[4]J. Mobley, K. R. Waters, M. S. Hughes, C. S. Hall, J. N. Marsh, G. H. Brandenburger, and J. G. Miller, "Kramers–Kronig relations applied to finite bandwidth data from suspensions of encapsulated microbubbles," J. Acoust. Soc. Am. **108**, 2091–2106 (2000).
[5]J. Mobley, K. R. Waters, M. S. Hughes, C. S. Hall, J. N. Marsh, G. H. Brandenburger, and J. G. Miller, "Erratum: 'Kramers–Kronig relations applied to finite bandwidth data from suspensions of encapsulated microbubbles' [J. Acoust. Soc. Am. 108, 2091–2106 (2000)]," J. Acoust. Soc. Am. **112**, 760–761 (2002).
[6]J. Mobley, K. R. Waters, and J. G. Miller, "Causal determination of acoustic group velocity and frequency derivative of attenuation with finite-bandwidth Kramers–Kronig relations," Phys. Rev. E **72**, 016604 (2005).
[7]H. A. H. Jongen, J. M. Thijssen, M. van den Aarssen, and W. A. Verhoef, "A general model for the absorption of ultrasound by biological tissues and experimental verification," J. Acoust. Soc. Am. **79**, 535–540 (1986).
[8]T. L. Szabo, "Causal theories and data for acoustic attenuation obeying a frequency power law," J. Acoust. Soc. Am. **97**, 14–24 (1995).
[9]J. C. Bamber, "Ultrasonic properties of tissues," in *Ultrasound in Medicine*, edited by F. A. Duck, A. C. Baker, and H. C. Starritt (Institute of Physics Publishing, Bristol, UK, 1998), pp. 57–88.
[10]K. R. Waters, M. S. Hughes, J. Mobley, G. H. Brandenburger, and J. G. Miller, "On the applicability of Kramers–Kronig relations for ultrasonic attenuation obeying a frequency power law," J. Acoust. Soc. Am. **108**, 556–563 (2000).
[11]J. Mobley, R. A. Mack, J. R. Gladden, and P. R. Mantena, "Determination of power-law attenuation coefficient and dispersion spectra in multi-wall carbon nanotube composites using Kramers–Kronig relations," J. Acoust. Soc. Am. **126**, 92–97 (2009).
[12]V. L. Ginzberg, "Concerning the general relationship between absorption and dispersion of sound waves," Sov. Phys. Acoust. **1**, 32–41 (1955).
[13]M. O'Donnell, E. T. Jaynes, and J. G. Miller, "General relationships between ultrasonic attenuation and dispersion," J. Acoust. Soc. Am. **63**, 1935–1937 (1978).
[14]M. O'Donnell, E. T. Jaynes, and J. G. Miller, "Kramers–Kronig relationship between ultrasonic attenuation and phase velocity," J. Acoust. Soc. Am. **69**, 696–701 (1981).
[15]K. Aki and P. G. Richards, *Quantitative Seismology: Theory and Methods* (Freeman, San Francisco, CA, 1980).
[16]K. R. Waters, J. Mobley, and J. G. Miller, "Causality-imposed (Kramers–Kronig) relationships between attenuation and dispersion," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **52**, 822–833 (2005).
[17]G. Haiat, A. Lhemery, F. Renaud, F. Padilla, P. Laugier, and S. Naili, "Velocity dispersion in trabecular bone: Influence of multiple scattering and of absorption," J. Acoust. Soc. Am. **124**, 4047–4058 (2008).
[18]G. Haiat, M. Sasso, S. Naili, and M. Matsukawa, "Ultrasonic velocity dispersion in bovine cortical bone: An experimental study," J. Acoust. Soc. Am. **124**, 1811–1821 (2008).
[19]A. Q. Bauer, K. R. Marutyan, M. R. Holland, and J. G. Miller, "Negative dispersion in bone: The role of interference in measurements of the apparent phase velocity of two temporally overlapping signals," J. Acoust. Soc. Am. **123**, 2407–2414 (2008).
[20]M. W. Urban and J. F. Greenleaf, "A Kramers–Kronig-based quality factor for shear wave propagation in soft tissue," Phys. Med. Biol. **54**, 5919–5933 (2009).
[21]B. E. Treeby, B. T. Cox, E. Z. Zhang, S. K. Patch, and P. C. Beard, "Measurement of broadband temperature-dependent ultrasonic attenuation and dispersion using photoacoustics," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **56**, 1666–1676 (2009).
[22]R. Raisutis, R. Kazys, and L. Mazeika, "Application of the ultrasonic pulse-echo technique for quality control of the multi-layered plastic materials," NDT & E Int. **41**, 300–311 (2008).
[23]E. L. Madsen, G. R. Frank, M. A. Hobson, S. Lin-Gibson, T. J. Hall, J. Jiang, and T. A. Stiles, "Instrument for determining the complex shear modulus of soft-tissue-like materials from 10 to 300 Hz," Phys. Med. Biol. **53**, 5313–5342 (2008).
[24]B. Castagnede, A. Moussatov, D. Lafarge, and M. Saeid, "Low frequency in situ metrology of absorption and dispersion of sound absorbing porous materials based on high power ultrasonic non-linearly demodulated waves," Acta Phys. Pol. **69**, 634–648 (2008).
[25]F. J. Alvarez and R. Kuc, "Dispersion relation for air via Kramers–Kronig analysis," J. Acoust. Soc. Am. **124**, EL57–EL61 (2008).
[26]K. E. Peiponen and J. J. Saarinen, "Generalized Kramers–Kronig relations in nonlinear optical- and THz-spectroscopy," Rep. Prog. Phys. **72**, 056401 (2009).
[27]J. Mobley, "Finite-bandwidth Kramers–Kronig relations for acoustic group velocity and attenuation derivative applied to encapsulated microbubble suspensions," J. Acoust. Soc. Am. **121**, 1916–1923 (2007).
[28]J. Mobley and R. E. Heithaus, "Ultrasonic properties of a suspension of microspheres supporting negative group velocities," Phys. Rev. Lett. **99**, 124301 (2007).
[29]J. Mobley, K. R. Waters, and J. G. Miller, "Finite-bandwidth effects on the causal prediction of ultrasonic attenuation of the power-law form," J. Acoust. Soc. Am. **114**, 2782–2790 (2003).
[30]T. L. Szabo, "Time domain wave equations for lossy media obeying a frequency power law," J. Acoust. Soc. Am. **96**, 491–500 (1994).
[31]R. Mills, "Tutorial on infinities in QED," in *Renormalization: From Lorentz to Landau (and Beyond)*, edited by L. M. Brown (Springer-Verlag, New York, 1993), pp. 57–85.
[32]S. Temkin, "Attenuation and dispersion of sound in dilute suspensions of spherical particles," J. Acoust. Soc. Am. **108**, 126–146 (2000).
[33]M. J. Buckingham, "On the transient solutions of three acoustic wave equations: van Wijngaarden's equation, Stokes' equation and the time-dependent diffusion equation," J. Acoust. Soc. Am. **124**, 1909–1920 (2008).
[34]K. R. Waters, M. S. Hughes, J. Mobley, and J. G. Miller, "Differential forms of the Kramers–Kronig dispersion relations," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **50**, 68–76 (2003).
[35]I. S. Gradshteyn, I. M. Ryzhik, A. Jeffrey, and D. Zwillinger, *Table of Integrals, Series, and Products* (Academic, San Diego, CA, 2000).

Sound velocities and thermodynamic properties of water to 700 MPa and −10 to 100 °C

Steve Vance[a)]
Jet Propulsion Laboratory, California Institute of Technology, MS 183-401, 4800 Oak Grove Drive, Pasadena, California 91109

J. Michael Brown
Department of Earth and Space Sciences and Astrobiology Program, University of Washington, 070 Johnson Hall, Seattle, Washington 98195-1310

(Received 2 December 2008; revised 28 September 2009; accepted 29 September 2009)

Sound velocities in liquid water were measured by the method of impulsive stimulated scattering in a sapphire-windowed high-pressure cell from −10 to 100 °C and pressures as high as 700 MPa. Velocity measurements are compared with previous experimental efforts relative to the International Association for the Properties of Water and Steam (IAPWS-95) formulation for the equations of state. At 0 and −10 °C, sound velocities are in agreement with the one previously published study at sub-zero temperatures to 350 MPa. At ambient and elevated temperatures, differences between the present measurements and IAPWS-95 velocities approach 0.5% near 700 MPa. Inversion of velocity data for density yields results within IAPWS-95 uncertainties, except at the highest temperatures, where elevated sound velocity at high pressure corresponds to as much as −0.2% disagreement with IAPWS-95. © *2010 Acoustical Society of America.* [DOI: 10.1121/1.3257223]

PACS number(s): 43.35.Cg, 43.30.Bp, 43.20.Hq, 43.20.Ye [AJZ] Pages: 174–180

I. INTRODUCTION

The study of fluid water has applications in nearly all branches of science and engineering. The substance displays unique phenomenology as a polar molecule with a range of behaviors, crystallographic phases, variation in fluid properties from negative to positive thermal expansion, and variable amounts of hydrogen bonding. Water's structure is influenced by pressure and temperature. The full range of water's phenomenology is, in theory, obtained among the various locales in the solar system[1] and, by extension, elsewhere in the universe. Pressure's influence on water's thermodynamic properties is important in numerous technical applications, biology,[2,3] and in the geosciences.[4]

Sound velocities in water are related to density through the bulk modulus (derivative of density with pressure) and are thereby useful for understanding its thermodynamic properties. Pressure obtained in Earth's deepest waters (the Marianas Trench) is ~100 MPa. In subsurface oceans in large icy satellites[5] (Europa, Callisto, and Ganymede), pressure likely approaches the gigapascal range. Coverage of sound velocities for water at higher pressures has been historically lacking; to our knowledge, Holton *et al.*[6] published the only systematic study covering the range of pressures up to 1000 MPa. Improved coverage of sound velocities at high pressures has direct application for future seismic investigations of icy world interiors.

Impulsive stimulated scattering has been previously applied to the measurement of sound velocities in the diamond-anvil cell in a pressure regime above 1 GPa and at elevated temperatures.[7] However, the precision of the pressure determination (about 20 MPa) makes lower pressure measurements problematic. Here we report sound velocities in water to 700 MPa at temperatures from −10 to 100 °C obtained in a new system where pressure can be more precisely characterized.[8]

The International Association for the Properties of Water and Steam (IAPWS-95) equation of state for water, compiled by Wagner and Pruss,[9] provides a useful standard for comparison of the present sound velocities with those obtained in previous studies. The document provides an analytical description of the properties of water at all pressures and temperatures constructed to best fit a selected set of thermodynamic data. Thermodynamic properties are calculated as the appropriate derivatives of a thermodynamic potential, ensuring internal consistency of the different parameters, with most reliable predictions in the low-pressure regime (<100 MPa) where most measurements have been made. Previous work by Abramson and Brown[7] documents deviation of IAPWS-95 predictions of sound velocity and density at pressures above 1 GPa. The present measurements also show systematic deviation at temperatures of 50 and 100 °C. These measurements provide a bridge between earlier low-pressure measurements and those made in diamond-anvil cells at much higher pressures.

II. EQUIPMENT AND METHODS

Sound velocities were measured by impulsive stimulated scattering (ISS) using instrumentation as previously described.[4] Absorption in the sample of convergent picosecond pulses of near-infrared light (λ_{IR}=1064 nm) sets up a standing acoustic wave. The spatially and temporally varying indices of refraction associated with the acoustic wave act as

[a)]Author to whom correspondence should be addressed. Electronic mail: svance@jpl.nasa.gov

 0001-4966/2010/127(1)/174/7/$25.00

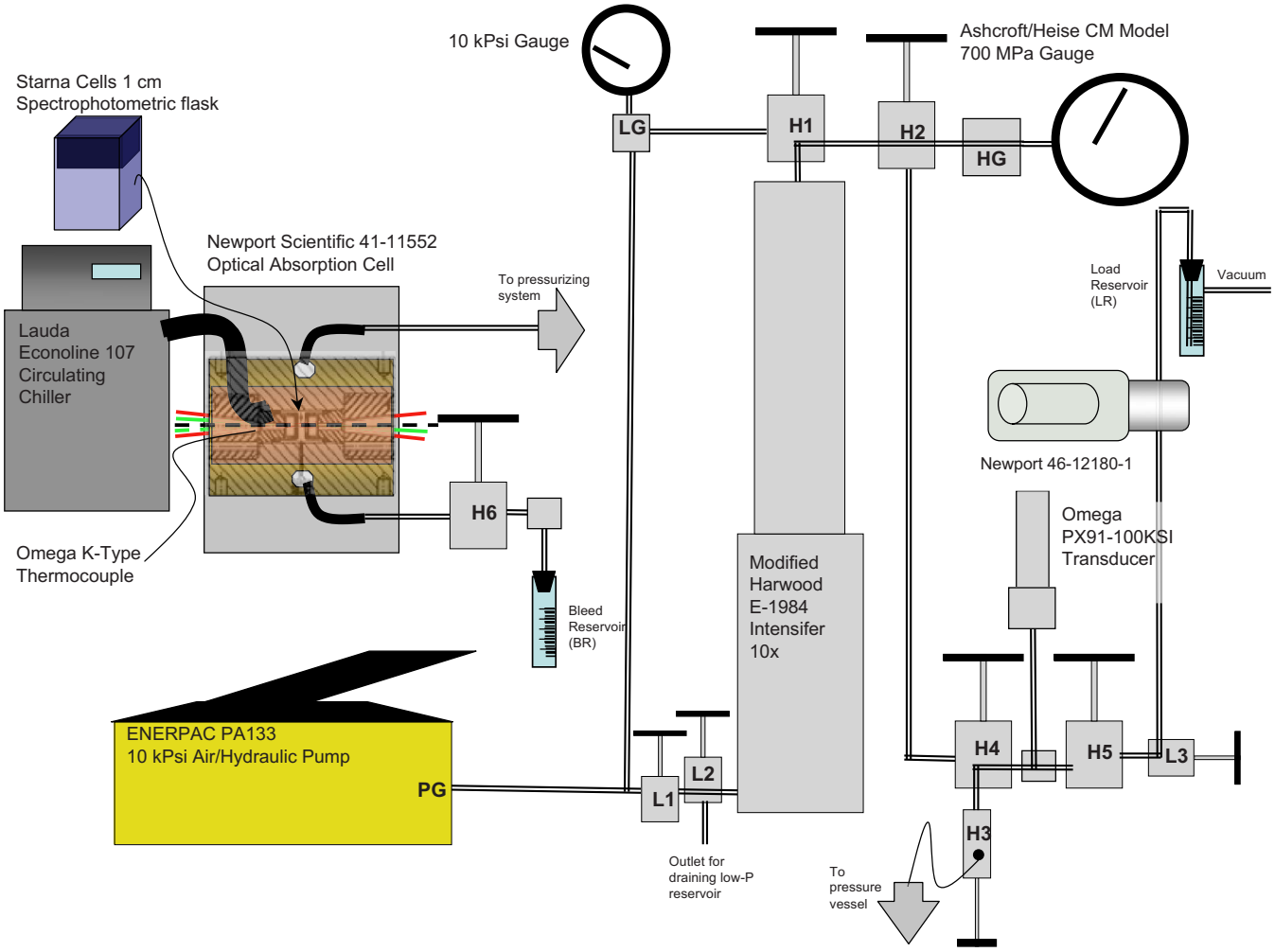


FIG. 1. (Color online) Schematic of the apparatus for obtaining high hydraulic pressure in aqueous solutions (Ref. 8).

a diffraction grating, from which a probe laser (532 nm) can be coherently scattered. The intensity of the scattered probe pulse, modulated at the acoustic frequency, constitutes the signal. The Fourier transform gives the acoustic frequency.

In the present study, the angle between the incident infrared beams was $\theta_{IR}=15°$, reduced from a nominal $\sim 25°$ used in previous studies, in order to accommodate the apparatus described below. The nominal wavelength of the acoustic waves is 4 μm ($\lambda=\lambda_{IR}/2\sin(\theta_{IR}/2)$) with a period of 1.5–2.5 ns, which provides 15–20 cycles in a spot size of 350 μm. The acoustic wavelength is accurately determined from the acoustic frequencies in de-ionized water at ambient pressure and temperature. Depending on pressure, the acoustic excitation persists for 30–50 ns.

A. The simulator for icy world interiors: An apparatus with optical access for *in-situ* measurements to 700 MPa from −10 to 100 °C

The simulator for icy world interiors (SIWIs) were constructed for the investigation of phenomena in simulated ocean world environments.[8] The system (Fig. 1) employs a Newport Scientific optical absorption cell modified for the current geometry.

High purity de-ionized water samples were contained in a Starna 10 mm glass spectrophotometric cell cut to fit in the opening of the pressure vessel and were sealed using a plug of durometer 70 polyurethane. The edges of the plug were coated with silicone vacuum grease to provide additional isolation of the sample from the pressure medium. Dioctyl sebacate was used as the hydraulic fluid. To prevent clouding at low temperatures, 5% kerosene was added by volume. Although the plug served as an imperfect barrier, materials used in the hydraulic system, including dioctyl sebacate, kerosene, and silicone oil, are not significantly miscible in water. Sound velocities were reproducible to experimental limits despite a varying history of pressure and temperature cycling for individual sample loads. Measurements are consistent with values reported in the literature at low pressure for all temperatures.

For the present study in the SIWI, nonplanar distortion of the windows due to pressure was determined to be negligible on the basis of calculation.[10] *C*-axis windows were used to minimize distortion of the incident angle from anisotropy of index of refraction. Any such effects were empirically found to be negligible.

Pressure was generated with a 70 MPa (absolute) pressure source connected to a Harwood 10× intensifier. We

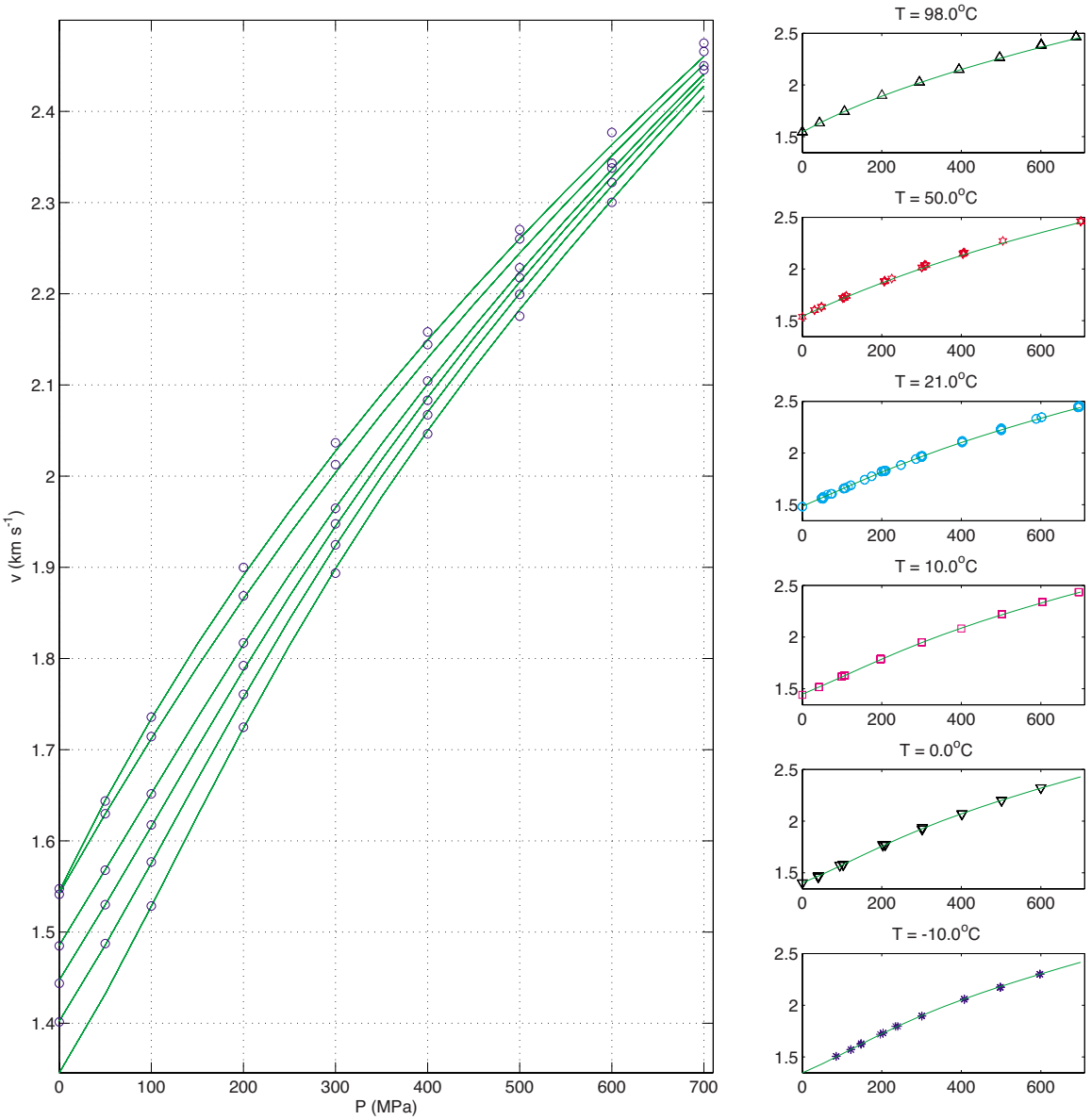


FIG. 2. (Color online) Sound velocities in water to 700 MPa from −10 to 100 °C. Velocities are plotted on the vertical axis (in km s^{-1}) as a function of pressure (in MPa). The panels at the right show the velocity measurements (circles) along the isotherms noted (also see Fig. 3). On both the right and left, temperature increases upward, the lowest curve corresponding to the lowest temperature. Solid lines represent values from the IAPWS-95 equation of state (Ref. 9). In the large panel, open circles represent gridded velocities used in the calculation of density. In the smaller panels on the right, individual data points are given for each temperature

used an Omega 700 MPa pressure transducer cross-checked against a Heise CM model 700 MPa Bourdon-tube gauge. Each is NIST certified by the manufacturer to an accuracy of 0.7 MPa, equivalent to approximately a 0.1% error in sound velocity for water at atmospheric pressure (roughly a factor of 3 less at the highest pressures). The gauges agreed within the stated uncertainty. Pressures were usually stable to within 3 MPa. In the event of pressure drift, an average of pressures during the ISS measurement was used.

Temperatures were measured using calibrated chromel-alumel thermocouples in contact with the external side of the sapphire window. The geometry of the pressure vessel and the high thermal conductivity of sapphire ensured an accurate determination of sample temperature. Thermocouples were calibrated to the melting and boiling temperatures of water at atmospheric pressure. Temperatures were controlled to 0.1 °C, equivalent to a 0.02% error in sound velocity at standard temperature. An observed rapid adiabatic temperature change (roughly 1 °C per 100 MPa) following compression or decompression confirmed correct placement of the thermocouple.

III. RESULTS

A. Velocities

Sound velocities were measured in de-ionized water. Measurements were made in multiple samples at each isotherm, except at 100 °C. Velocities were measured along six isotherms, as displayed in Fig. 2, and at intermediate temperatures during heating and cooling. Measurements were made on increasing and decreasing pressures to assess reproducibility. Successive measurements of frequency at a given

TABLE I. Sound velocities for water as a function of temperature and pressure. The value in the lower line for each entry is calculated using the IAPWS-95 representation. Table entries are averages of measurements made within 2 °C and 50 MPa and corrected to the nominal P and T using derivatives calculated from IAPWS-95. The expanded error (2σ) is 0.2%, except where (†) denotes a slightly larger error of 0.3%. Entries with asterisks are spline interpolations of surrounding data. Bolded velocities at 1 and 2 GPa are derived from measurements by Abramson and Brown (Ref. 7) (Fig. 4 of that reference).

P (MPa)	T (°C)					
	−10.0	0.0	10.0	21.0	50.0	98.0
0.10	1.345*	1.401†	1.444	1.485†	1.541	1.548
	1.345	1.402	1.447	1.485	1.543	1.545
50.00	1.432*	1.487	1.530	1.568	1.630	1.644
	1.432	1.486	1.530	1.568	1.630	1.645
100.00	1.529	1.577	1.617	1.652	1.714	1.736
	1.528	1.575	1.616	1.652	1.712	1.734
200.00	1.725	1.761	1.792	1.817	1.869	1.900
	1.723	1.757	1.787	1.815	1.865	1.891
300.00	1.894	1.925†	1.948	1.965	2.013	2.037
	1.899	1.924	1.945	1.965	2.004	2.028
400.00	2.046	2.067	2.083	2.104	2.144	2.158
	2.050	2.070	2.086	2.101	2.129	2.149
500.00	2.175	2.199	2.218	2.228	2.260	2.270
	2.182	2.200	2.213	2.224	2.245	2.260
600.00	2.300	2.322	2.338	2.343	2.367*	2.377
	2.303	2.318	2.328	2.336	2.351	2.363
700.00	2.421*	2.435*	2.446	2.450	2.466	2.475
	2.416	2.428	2.435	2.441	2.452	2.460
1000.00	**2.739**	**2.731**	**2.729**	**2.729**	**2.728**	**2.728**
	2.733	2.726	2.724	2.724	2.723	2.722
2000.00	**3.588**	**3.513**	**3.472**	**3.448**	**3.415**	**3.391**
	3.609	3.533	3.492	3.468	3.436	3.410

pressure and temperature agreed to be better than 0.1%, as did wavelength standards obtained in the ambient pressure standard before and after an isothermal measurement. However, velocities obtained by returning to the same pressure and temperature showed a standard deviation of up to 0.3%. This suggests either temperature uncertainties of about 1 °C or pressure uncertainties of about 2 MPa (or a combination of both pressure and temperature uncertainties).

Data were organized into representative pressure and temperature bins (Fig. 2, Table I, and also in Fig. 3). The binned values were corrected to the nominal pressures and temperatures using derivatives calculated from IAPWS-95. Averages and standard errors were determined from these results. The expanded error (2σ) is typically smaller than 0.2%. We adopt 0.2% as a representative value for all binned data, except in the few cases where the expanded error is 0.3%, as shown in Fig. 3 and noted in Table I.

B. Density

Density was calculated by recursive integration of

$$\left(\frac{\partial \rho}{\partial P}\right)_T = \frac{1}{v^2} + \frac{T\alpha^2}{C_P} \quad \text{and} \quad \left(\frac{\partial C_P}{\partial P}\right)_T = -T\frac{\partial^2 V_{\text{sp}}}{\partial T^2}. \tag{1}$$

A grid of velocities (Table I) was constructed from the matrix of velocity data.[11] Interpolation and extrapolation in pressure were employed at temperatures and pressures where no binned data were available. Additional points were added at 1 and 2 GPa using a modification of IAPWS-95 velocities based on measurements by Abramson and Brown,[7] and at the lowest pressure and temperature using the IAPWS-95 value.

The constants of integration in Eq. (1) (density and specific heat at 1 bar and all temperatures) were taken from IAPWS-95. Densities, the coefficient of thermal expansion, and specific heats were then calculated.

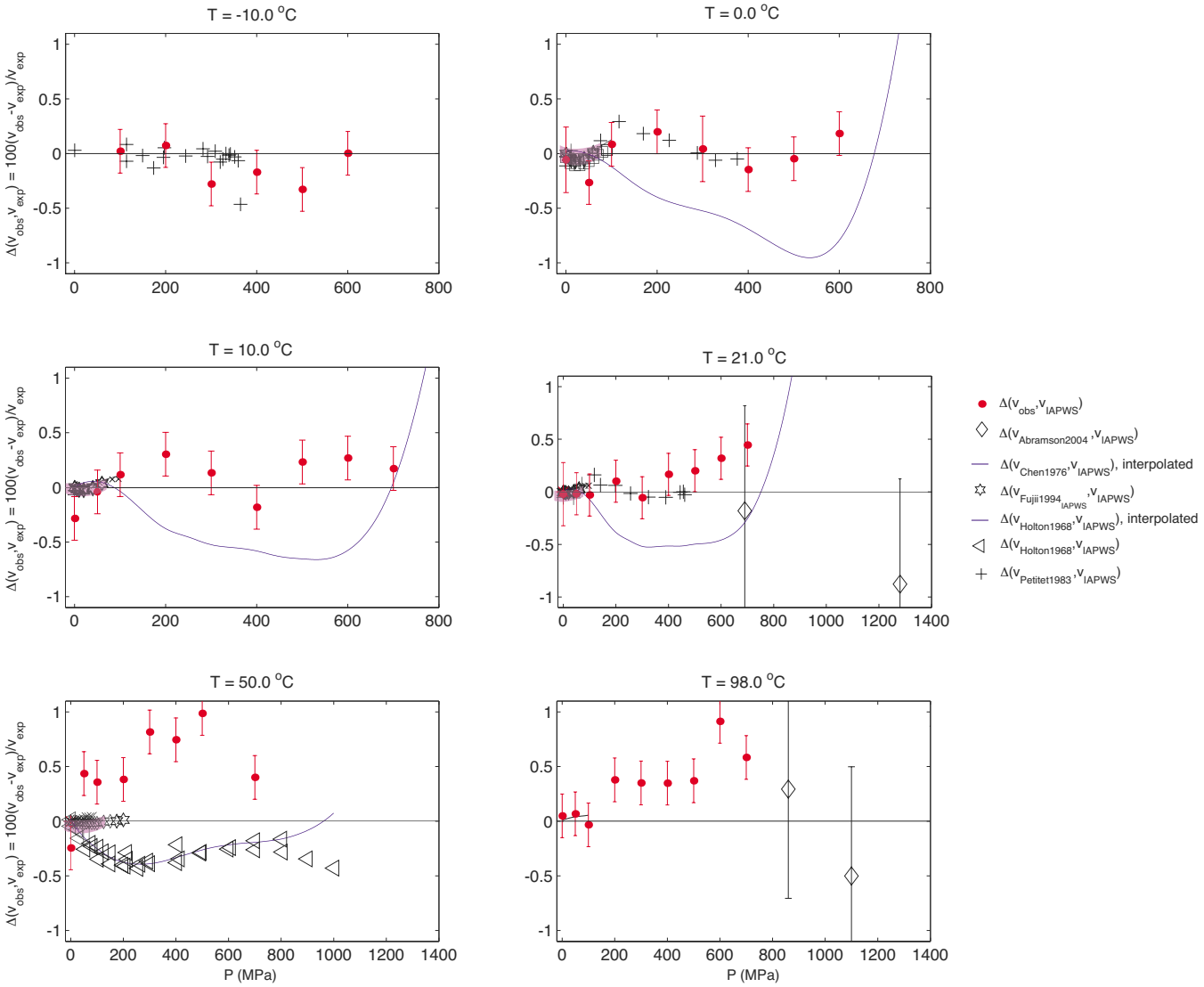


FIG. 3. (Color online) Comparison of sound velocities from this study with previously published values (Refs. 7, 12–17, 6, 18, and 19). The range of residuals for published measurements is shown schematically for overlapping sources measured below 100 MPa. Values for the present study were formed using the data in Table I. Residuals are referenced to the equation of state from the IAPWS-95 (Ref. 9). Error in measurements by Abramson and Brown (Ref. 7) is 1% for the measurements shown. Those errors are larger than the ones reported here due to lower precision in measurement of pressure in the diamond-anvil cell. Reported errors and confidence intervals for the other data are less than the expanded error (2σ) of 0.2% reported for our binned data within the sizes of the markers and lines.

The calculation of density is relatively insensitive to the initial specific heat; substituting standard values for water changes the calculated densities by a maximum of three parts per million (0.003 kg m^{-3}). Variations in the initial densities and specific heats within reasonable estimates of their own uncertainties, and use of other starting pressures and interpolation schemes, change the inferred density by a similarly small amount. The uncertainty in density corresponding to a systematic 0.2% error in sound velocity was found to be less than 0.1% for the pressure regime of these data.

IV. DISCUSSION

Figure 2 compares current sound velocities in water with predictions from IAPWS-95 (lines). The panels on the right side show all velocities measured along the six isotherms. On the left the bin-averaged velocities given in Table I are plotted. The data and predictions of IAPWS-95 are in general accord. However, increasing systematic deviations at pressures above 500 MPa are evident; this is a regime where experimental constraints on IAPWS-95 become sparse.

In order to evaluate the current data set in relation to prior data and IAPWS-95, residuals between measurements and IAPWS-95 are shown in Fig. 3. Temperature derivatives based on IAPWS-95 are used to correct all data to the reported isotherms. A number of studies in the low-pressure regime (often extending to less than 100 MPa)[12–17,6,18,19] are in mutual agreement to better than 0.1%, show small residuals relative to IAPWS-95 (since they were used to construct the equation of state), and are schematically represented by a thick line in the region of overlap. Additional ultrasonic measurements from Belogol'skii *et al.*[20] (0–40 °C to 60 MPa) and Benedetto *et al.*[13] (1–120 °C to 90 MPa) (not used to determine IAPWS-95) lie within 0.02% of IAPWS-95 velocities. Two earlier data sets[6,18] extend beyond 200 MPa.

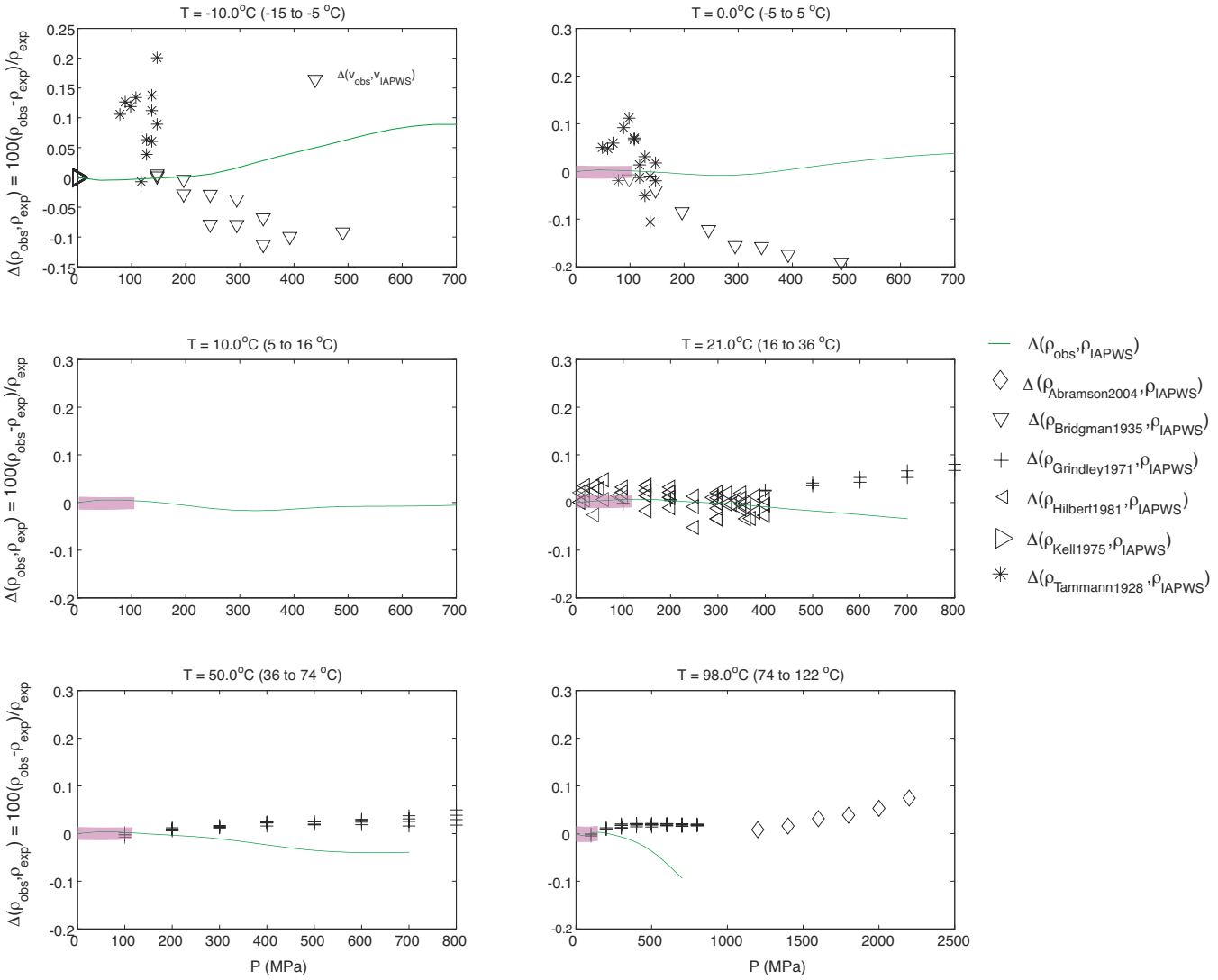


FIG. 4. (Color online) Comparison of densities from this study with previously published values (Refs. 7 and 21–27), most of which were selected for construction of the IAPWS-95 (Ref. 10). Residuals are shown relative to IAPWS-95. The range of residuals for published measurements is shown schematically for overlapping sources measured below 100 MPa for comparison with the velocities presented here (Fig. 3); densities are given at isotherms on which the majority of measurements were made.

Holton *et al.*[6] reported data at 50 °C; the measurements and their fit are plotted. At other temperatures only their fits were reported and are shown as lines.

The present measurements show agreement with IAPWS-95 and most of the other measurements at the lower pressure. In general, the current data and all other data sets do not reproduce trends given in the data of Holton *et al.*[6] The large deviations between their values and IAPWS-95 above 600 MPa may be associated with instabilities in their polynomial fits to the measured velocities. Velocity values from Petitet *et al.*[18] at −10, 0, and 21 °C are in excellent agreement with the current work (note that even small departures from IAPWS-95 are consistent). Above 500 MPa for 21, 50, and 98 °C, the current velocities trend higher than IAPWS-95 by about 0.5%.

The lowest pressure diamond-anvil cell measurements of Abramson and Brown[9] are plotted for the 21 and 98 °C isotherms. These higher pressure measurements (extending to 6 GPa) suggest increasing negative residuals at higher pressure. However, the large velocity uncertainty of 1% associated with the uncertainty in the pressure measurement limits the resolution of comparisons in the current range of pressure.

Densities are recursively calculated from the velocities given in Table I using Eq. (1). Results shown in Fig. 4 as residuals from IAPWS-95 are compared with several direct measurements of density. The current equation of state deviates from IAPWS-95 by less than 0.1% except at 98 °C where the deviations become as large as 0.2% near 700 MPa. The other experimental data are in reasonable accord both with the current equation of state and IAPWS-95.

V. CONCLUSION

Impulsive stimulated scattering measurements of sound velocity for water extend coverage of previous ultrasonic measurements by Petitet *et al.*[18] at subzero temperatures to 700 MPa. These measurements provide improved coverage up to the melting point of high-pressure ice phases and good agreement with previously published measurements below

50 °C. Comparison of our measured sound velocities and previously published values relative to IAPWS-95 (Ref. 9) shows a positive bias in our data at elevated temperature and pressure in the opposite direction to the only other source above 200 MPa.[6]

Density measurements reported here provide precision comparable to previous measurements at elevated pressure. The discrepancy between IAPWS-95 reference data at elevated pressures highlights uncertainty in the equation of state for water at sub-gigapascal pressures, particularly at the lowest and highest temperatures.

ACKNOWLEDGMENTS

The majority of the work presented here was performed while S.V. was a student at the University of Washington, supported by National Science Foundation's IGERT program (Grant No. DGE-9870713), "IGERT: Astrobiology: Life in and beyond Earth's Solar System," NSF Award No. EAR-0337796, and the NASA Astrobiology Institute and NASA Outer Planets Research Program Award Nos. NNX08AQ51G and NNG06GF81G. Later measurements and analysis were supported by the Icy Worlds node of NASA's Astrobiology Institute, the NASA Postdoctoral Program, and the Jet Propulsion Laboratory at the California Institute of Technology.

[1]V. R. Baker, J. M. Dohm, A. G. Fairen, T. P. A. Ferre, J. C. Ferris, H. Miyamoto, and D. Schulze-Makuch, "Extraterrestrial hydrogeology," Hydrogeol. J. **13**, 51–68 (2005).
[2]T. Makita, "Application of high-pressure and thermophysical properties of water to biotechnology," Fluid Phase Equilib. **76**, 87–95 (1992).
[3]M. Willis, T. Ahrens, L. Bertani, and C. Nash, "Bugbuster—Survivability of living bacteria upon shock compression," Earth Planet. Sci. Lett. **247**, 185–196 (2006).
[4]E. H. Abramson, J. M. Brown, and L. J. Slutsky, "Applications of impulsive stimulated scattering in the earth and planetary sciences," Annu. Rev. Phys. Chem. **50**, 279–313 (1999).
[5]R. D. Lorenz, B. W. Stiles, R. L. Kirk, M. D. Allison, P. P. del Marmo, L. Iess, J. I. Lunine, S. J. Ostro, and S. Hensley, "Titan's rotation reveals an internal ocean and changing zonal winds," Science **319**, 1649–1651 (2008).
[6]G. Holton, M. P. Hagelberg, S. Kao, H. Walter, and J. Johnson, "Ultrasonic-velocity measurements in water at pressures to 10000 kg/cm^2," J. Acoust. Soc. Am. **43**, 102–107 (1968).
[7]E. H. Abramson and J. M. Brown, "Equation of state of water based on speeds of sound measured in the diamond-anvil cell," Geochim. Cosmochim. Acta **68**, 1827–1835 (2004).
[8]S. Vance and J. Brown, "The simulator for icy world interiors, an apparatus for optical measurements in aqueous systems in the range −20 to 100 °C and 700 MPa," Rev. Sci. Instrum. **79**, 105105 (2008).
[9]W. Wagner and A. Pruss, "The IAPWS formulation 1995 for the thermodynamic properties of ordinary water substance for general and scientific use," J. Phys. Chem. Ref. Data **31**, 387–535 (2002).
[10]D. Shelton, "Lens induced by stress in optical windows for high-pressure cells," Rev. Sci. Instrum. **63**, 3978–3982 (1992).
[11]C. De Boor, *A Practical Guide to Splines* (Springer, Berlin, 2001).
[12]A. Barlow and E. Yazgan, "Pressure dependence of the velocity of sound in water as a function of temperature," Br. J. Appl. Phys. **18**, 645–651 (1967).
[13]G. Benedetto, R. M. Gavioso, P. A. G. Albo, S. Lago, D. M. Ripa, and R. Spagnolo, "Speed of sound in pure water at temperatures between 274 and 394 K and at pressures up to 90 MPa," Int. J. Thermophys. **26**, 1667–1680 (2005).
[14]C. T. Chen and F. J. Millero, "Re-evaluation of Wilson's sound-speed measurements for pure water," J. Acoust. Soc. Am. **60**, 1270–1273 (1976).
[15]V. DelGrosso and C. Mader, "Speed of sound in pure water," J. Acoust. Soc. Am. **52**, 1442 (1972).
[16]K. Fujii, "Accurate measurements of the sound velocity in pure water under high pressure," in 12th Symposium on Thermophysical Properties, Boulder, CO (1994).
[17]K. Fujii and R. Masui, "Accurate measurements of the sound-velocity in pure water by combining a coherent phase-detection technique and a variable path-length interferometer," J. Acoust. Soc. Am. **93**, 276–282 (1993).
[18]J. P. Petitet, R. Tufeu, and B. Le Neindre, "Determination of the thermodynamic properties of water from measurements of the speed of sound in the temperature range 251.15–293.15 K and the pressure range 0.1–350 MPa," Int. J. Thermophys. **4**, 35–50 (1983).
[19]W. D. Wilson, "Speed of sound in distilled water as a function of temperature and pressure," J. Acoust. Soc. Am. **31**, 1067–1072 (1959).
[20]V. Belogol'skii, S. S. Sekoyan, L. M. Samorukova, S. R. Stefanov, and V. I. Levtsov, "Pressure dependence of the sound velocity in distilled water," Meas. Tech. **42**, 406–413 (1999).
[21]P. Bridgman, "The pressure-volume-temperature relations of the liquid, and the phase diagram of heavy water," J. Chem. Phys. **3**, 597 (1935).
[22]T. Grindley and J. Lind, Jr., "PVT properties of water and mercury," J. Chem. Phys. **54**, 3983–3989 (1971).
[23]R. Hilbert, K. Toedheide, and E. U. Franck, "PVT data for water in the ranges 20 to 600 degrees C and 100 to 4000 bar," Ber. Bunsenges. Phys. Chem **85**, 636–643 (1981).
[24]G. Kell, "Density, thermal expansivity, and compressibility of liquid water from 0. deg. to 150. deg. Correlations and tables for atmospheric pressure and saturation reviewed and expressed on 1968 temperature scale," J. Chem. Eng. Data **20**, 97–105 (1975).
[25]G. Kell and E. Whalley, "Reanalysis of the density of liquid water in the range 0–150 °C and 0–1 kbar," J. Chem. Phys. **62**, 3496–3509 (1975).
[26]M. Takenaka and R. Masui, "Measurement of the thermal expansion of pure water in the temperature range 0 °C–85 °C ," Metrologia **27**, 165–171 (1990).
[27]G. Tammann and W. Jellinghaus, "Die Volumenisobaren des Wassers bis zur Schmelzkurve (The volume isobars of water to the melting curve)," Z. Anorg. Allg. Chem. **174**, 225 (1928).

Experimental and theoretical study of acoustic waves generated by a laser line pulse in an optically absorptive isotropic cylinder

D. Ségur, A. L. Shuvalov, and B. Audoin
Laboratoire de Mécanique Physique, UMR CNRS 5469, Université Bordeaux 1, 33405 Talence, France

Y. D. Pan
Institute of Acoustics, Tongji University, Shanghai, China

(Received 8 April 2009; revised 14 October 2009; accepted 14 October 2009)

The generation of acoustic waves by a line-focused laser pulse in an optically absorptive cylinder is studied experimentally and theoretically. Experiments are performed on a 5 mm diameter NG5 colored glass rod using Nd:yttrium aluminum garnet laser, which delivers 5 ns pulses. The numerical simulation is based on the semi-analytical model of a radially distributed thermoelastic source, which takes into account penetration of laser energy into the bulk of the sample. Good agreement between the experimental and calculated wave forms is observed. Comparison of these wave forms with an auxiliary simulation, which assumes the model of a dipole source located at the cylinder surface, reveals the effect of optical penetration on the shape of the wave form and also on the relative amplitude of bulk and surface waves.
 [DOI: 10.1121/1.3263612]

PACS number(s): 43.35.Cg, 43.35.Sx, 43.35.Ud [YHB] Pages: 181–185

I. INTRODUCTION

Having emerged in the 1980s, the laser ultrasonics technique with its non-contact generation and detection process allows performing experiments for any curved geometry. Royer *et al.*[1] were the first to report laser generation and detection of surface acoustic waves on metallic spheres. Similar method has then been applied to cylinder targets with the aim to study surface acoustic waves and to employ them for defect detection and non-destructive evaluation.[2–4] Pan and co-workers[5,6] analyzed both theoretically and experimentally laser-generated acoustic response of a transversely isotropic cylinder and also solved the inverse problem for finding elastic constants of the cylinder material. These and other studies have assumed that the cylinder material was opaque at a given laser wavelength, and hence the source was located at the surface. In such a case, two different energetic regimes of laser ultrasonics generation are described by the corresponding theoretical models of a surface source, which have earlier been developed for a plate. The ablation regime implies that the deposited energy is locally high enough to vaporize a small area of the cylinder surface, thus creating in reaction a normal loading of the sample (the ablation model).[7,8] The thermoelastic regime corresponds to relatively low incoming energy, when the dilatation of an infinitely small volume adjacent to the surface produces a source equivalent to a dipole of forces (the dipole model of Scruby).[8,9] Using different descriptions of this small excited volume (termed surface center of expansion[10]), Royer[11] and Arias and Achenbach[12] derived an expression for the dipole strength.

Thermoelastic regime of laser ultrasonics generation in optically absorptive materials, where the optical penetration rules out the model of a surface source, has been studied for the case of a planar target (plate).[13] It has not, however, been explored for another basic case, which is when the target has a cylindrical shape. This is important, for instance, for studying fibers that are broadly used in composite materials. In the context of applications, the particular interest is in microfibers studied by a picosecond ultrasonic technique,[14] where taking into account the optical penetration becomes essential.[15] The present paper reports first experimental results for the case of laser acoustic-wave generation in an optically absorptive millimetric sized cylindrical target excited by nanosecond laser pulses.

The outline of this work is as follows. Normally incident line-focused pulse of a Nd:yttrium aluminum garnet (Nd:YAG) laser penetrates into the bulk of the isotropic NG5 colored glass rod, where it falls off exponentially along the radial direction. Absorption of this radiation causes heating, the heated region undergoes thermal expansion, and the thermoelastic stresses generate elastic waves such that propagate into the cylinder and along its surface. Theoretical calculation is based on the two-dimensional (2D) semi-analytical model, which provides closed-form solution to the ordinary differential problem with a distributed thermoelastic source in the Fourier domain and uses two inverse transforms to obtain the acoustic response in the time-space domain. The observed agreement between the calculated and experimental wave forms confirms the consistency of modeling. The effect of optical penetration is visualized through comparison with the simulation based on the model of a surface dipole source. It is seen that the impact of optical penetration visibly broadens the shape of wave arrivals. In particular, the first longitudinal-wave arrival has nearly exponential profile that is directly imposed by the exponential law of optical penetration into the bulk of the cylinder. Another important effect concerns the relative magnitudes of the amplitudes of surface

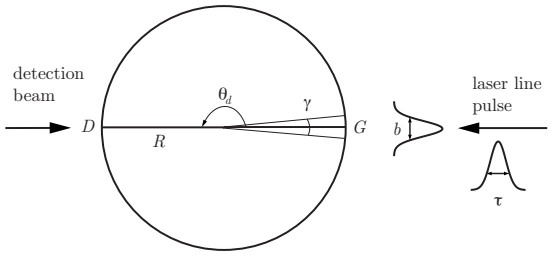


FIG. 1. The cylindrical problem geometry.

and bulk waves. While the former is notably greater than the latter in the case of a surface generation, both amplitudes are evidently comparable in the present case of an optically absorptive cylinder.

II. GOVERNING THERMOELASTIC EQUATIONS

Consider an infinite homogenous isotropic cylinder of radius R irradiated by laser. A line-focused laser pulse is assumed to hit the cylinder surface at the generation point G, and the probe is located at the diametrically opposite detection point D (see Fig. 1). A simplified model of thermoelastic generation is adopted, which disregards the thermal conduction in view of its negligible effect for the glass material under study. We also discard the contribution of deformation to the entropy change (this latter, standard assumption is mainly for brevity of exposition). Acoustic absorption, which is small in the given material, can be neglected as well. In consequence, the system of thermoelasticity equations[16] reduces to the following form:

$$\nabla \cdot \boldsymbol{\sigma} = \rho \ddot{\mathbf{u}}, \tag{1a}$$

$$\boldsymbol{\sigma} = \lambda \, \mathrm{tr}(\boldsymbol{\varepsilon})\mathbf{I} + 2\mu\boldsymbol{\varepsilon} - \beta T\mathbf{I}, \tag{1b}$$

$$\rho c_p \dot{T} = q, \tag{1c}$$

where $\boldsymbol{\sigma}=\boldsymbol{\sigma}^{(\mathbf{m})}+\sigma^{(T)}\mathbf{I}$ ($\sigma^{(T)}=-\beta T$) is the stress tensor including mechanical and thermal stresses, $\boldsymbol{\varepsilon}$ is the strain tensor, $\mathbf{u}$ is the displacement, ρ is the density, λ and μ are the Lamé coefficients, $\beta=(3\lambda+2\mu)\alpha_T$ is the thermal modulus (α_T is the dilatation coefficient), T is the temperature rise, c_p is the specific heat, q is the power density of the line laser source, $\mathbf{I}$ is the identity matrix, and tr denotes the trace.

The line-focused heat input $q(r,\theta,t)$ creates a volume distribution of thermal sources in the cylinder due to the optical absorption. Normal incidence of the laser beam implies that the absorption occurs essentially in the radial direction. At the same time, the laser beam has a Gaussian profile,[17] with a certain width at the surface. A normalized Gauss function is introduced to represent the angular extent of the laser beam in the orthoradial direction,

$$g(\theta) = \frac{1}{\gamma}\sqrt{\frac{4\ln 2}{\pi}}e^{-(4\ln 2)(\theta^2/\gamma^2)}, \tag{2}$$

where γ is the so-called full width at half maximum. It is defined as the angle $\gamma=2\arctan(b/2R)$ determined through the beamwidth b and the cylinder radius R (see Fig. 1). The energy absorption in the radial direction is described by the exponential law governed by the extinction coefficient α. Hence, the heat input pulse is written as follows:

$$q(r,\theta,t) = \alpha E g(\theta)\delta(t)e^{-\alpha(R-r)}, \tag{3}$$

where the intensity per unit length $E=E_0(1-\mathcal{R})$ takes into account reflection of the incident laser energy E_0 from the cylinder surface, which is described by the reflection coefficient $\mathcal{R}=((1-n)/(1+n))^2$ (n is the refractive index). Angular dependence of the reflection coefficient is neglected owing to the small ratio $b/R=0.02$ of the beamwidth b to the cylinder radius R. From Eqs. (1c) and (3), the temperature field $T(r,\theta,t)$ is obtained in the form

$$T(r,\theta,t) = \frac{\alpha E}{\rho c_p} g(\theta)H(t)e^{-\alpha(R-r)}, \tag{4}$$

where H is the Heaviside function.

Inserting Eq. (4) into Eq. (1b) specifies the thermal stress. The problem now amounts to solving the equation of motion (1a) in $\mathbf{u}(r,\theta,t)$ under the traction-free boundary condition (BC) at the cylinder surface $r=R$. Thanks to the line-focused source, this is a 2D plane strain problem in the cylinder cross-section $(\mathbf{e}_r,\mathbf{e}_\theta)$.

III. 2D SEMI-ANALYTICAL MODEL

A. Solution in the Fourier domain

Consider the Fourier domain $\{r,\nu,\omega\}$, where ν is the circumferential wave number and ω is the angular frequency. For brevity, the factor $e^{j(\nu\theta-\omega t)}$ will be omitted and dependence on ν,ω kept tacit. The equation of motion, written with respect to the longitudinal and shear elastic potentials $\varphi(r)$ and $\psi(r)$, reduces in the Fourier domain to the uncoupled ordinary differential equations

$$\nabla^2\varphi(r) + k_L^2\varphi(r) = \frac{\beta T_0}{\lambda+2\mu}e^{-\alpha(R-r)},$$

$$\nabla^2\psi(r) + k_T^2\psi(r) = 0, \tag{5}$$

where $k_L^2=\rho\omega^2/(\lambda+2\mu)$, $k_T^2=\rho\omega^2/\mu$, and T_0 involves double transform of Eq. (4) in θ and t. The potentials $\varphi(r)$ and $\psi(r)$ are coupled through the traction-free BC,

$$\begin{pmatrix}\sigma_{rr}^{(m)}(R)\\ \sigma_{r\theta}^{(m)}(R)\end{pmatrix} + \begin{pmatrix}-\beta T_0\\ 0\end{pmatrix} = \mathbf{0}, \tag{6}$$

in which the mechanical-traction components $\sigma_{rr}^{(m)}(R)$ and $\sigma_{r\theta}^{(m)}(R)$ are linearly expressed through $\varphi(R)$ and $\psi(R)$.

The solution to Eq. (5) can be obtained in two steps. First, the Green's function satisfying the homogeneous BC on the mechanical traction is found.[18] Then its convolution with the radially distributed source term is taken such that satisfies BC (6). Applying Helmholtz-decomposition formula to the obtained potential solutions $\varphi(r)$ and $\psi(r)$ yields the displacement $\mathbf{u}(r)$. The laser ultrasonics technique implicates the surface displacement $\mathbf{u}(R)$. It is found in the following closed form:[19]

$$\mathbf{u}(R) = \begin{pmatrix} u_r \\ u_\theta \end{pmatrix} = IU_0 \frac{K_L^2}{J_\nu(K_L)} \begin{pmatrix} \nu^2 - B_T - \dfrac{K_T^2}{2} \\ j\nu(1-B_T) \end{pmatrix}, \tag{7}$$

where

$$I = \int_0^1 J_\nu(K_L x) e^{-\alpha R(1-x)} x \mathrm{d}x \quad (x = r/R), \quad U_0 = \frac{\beta T_0 R}{2\mu D},$$

$$D = -\nu^4 + \nu^2(B_L B_T + K_T^2 + 1) - \left(\frac{K_T^2}{2} + B_L\right)\left(\frac{K_T^2}{2} + B_T\right), \tag{8}$$

and the auxiliary notations are

$$K_{L,T} \equiv k_{L,T} R, \quad B_{T,L} \equiv \frac{K_{T,L} J'_\nu(K_{T,L})}{J_\nu(K_{T,L})},$$

$$' \equiv \left[\frac{\mathrm{d}J_\nu(x)}{\mathrm{d}x}\right]_{x=K_{L,T}}. \tag{9}$$

The integral I takes into account the radial penetration of the laser input into the bulk of the cylinder, and $D=0$ is the dispersion equation for the eigenmodes of the traction-free cylinder.

B. Acoustic response in the time-space domain

The radial component of surface response, which is a quantity measured in the later described experiment, has been calculated by means of taking 2D Fourier transform of $u_r(R) \equiv \hat{u}_r(R, \nu, \omega)$ given by Eq. (7) (a hat is added to distinguish functions in different domains). Fourier transform in ω has been computed by means of the inverse fast Fourier transform algorithm. The poles associated with the dispersion equation $D=0$ were avoided by providing frequency with a small negative imaginary part, so that $\omega^* = \omega - J\delta$, with $\delta = 0.035$ used in the present calculation. Fourier transform in ν is simplified due to $\hat{u}_r(R, \nu, \omega)$ being an even function of ν [see Eq. (7)]. Thus,

$$u_r(R, \theta, t) = \frac{e^{-\delta t}}{\pi} \int_{-\infty}^{\infty} \left\{ \sum_{\nu=0}^{\infty} \varepsilon_\nu \hat{u}_r(R, \nu, \omega) \cos(\nu\theta) \right\} e^{-J\omega t} d\omega, \tag{10}$$

where ε_ν is the Neumann factor equal to 1 if $\nu=0$ and to 2 otherwise. Note that expressing Eq. (7) in terms of $B_{T,L}$ is helpful for eliminating the impact of exponential growth of the Bessel functions of the order $\nu \gg K_{L,T}$ (see Ref. 20).

The series in Eq. (10) has been truncated basing on the criterion

$$\frac{|\hat{u}_r(R, \nu_m, \omega)|}{\sum_{\nu=0}^{\nu_m} |\hat{u}_r(R, \nu, \omega)|} < \varepsilon, \tag{11}$$

where ε is the error bound and ν_m is the truncation order, which, strictly speaking, depends on ω. It was, however, verified that taking a fixed truncation order ν_m, which keeps about 200 terms of the series, ensures the error bound $\varepsilon \lesssim 10^{-5}$.

IV. EXPERIMENTAL RESULTS AND ANALYSIS

A Nd:YAG laser has been used for the acoustic-wave generation. The pulse duration is about 5 ns. The blue light wavelength $\lambda_g = 355$ nm was used for the generation. The collimated laser beam was line focused by means of the cylindrical lens with a focal length of 100 mm. The line was about 4 cm in length and 0.15 mm in width. With this arrangement, an optical heterodyne interferometer measures the normal displacement $u_r(t)$ at the detection point on the cylinder surface (see Fig. 1). It uses a doubled Nd:YAG laser to generate a probe beam at the wavelength $\lambda_d = 532$ nm. The output power was about 150 mW, and the sensitivity was 2.10^{-7} nm$(\mathrm{W/Hz})^{1/2}$ for a bandwidth between 30 kHz and 120 MHz. The experimental signal was averaged by 4000 shots.

A NG5 colored glass rod with diameter of 5 mm has been taken as the optically absorptive sample. Its physical parameters are $\rho = 2.31$ g cm^{-3}, $\lambda = 17.4$ GPa, $\mu = 22.9$ GPa, $\alpha_T = 6.5 \times 10^{-6}$ K^{-1}, and $c_p = 700$ J kg^{-1} K^{-1} (see the notations in Sec. II). The reflection coefficient at $\lambda_g = 355$ nm is $\mathcal{R} = 0.06$. The optical penetration length α^{-1} is 0.8 mm at the generation wavelength $\lambda_g = 355$ nm and 1.8 mm for $\lambda_d = 532$ nm. The choice of the generation wavelength providing a relatively lower value of α^{-1} was made to avoid overloading the detector. A thin metallic film was deposited on the cylinder surface at the detection point in order to prevent the probe-laser radiation from entering the sample.

Experimental and numerical results are presented in Fig. 2. Identification of wave arrivals, marked by dashed lines, is based on the ray trajectory analysis.[21] Each arrival implies a certain complex wave path from the generation to the detection point. It may involve the same or different modes. This is indicated through the following nomenclature proposed in Ref. 21. The notations nL and nT correspond to the arrivals of longitudinal and transverse bulk modes, traveling along the straight diameter path and undergoing n normal reflections from the cylinder surface at the edge points. The notations nP and nS also correspond to, respectively, the longitudinal and transverse bulk modes, but these ones arrive at the detection point after a broken path resulting from n oblique reflections without modal conversion. The wave arrivals, involving modal conversion at oblique reflections, are denoted by $mPnS_r$, where the subscript r is the number of round trips about the center made within the given path. The Rayleigh-wave arrival denoted by R_r is helpful for finding the epicentrial position for the detection point. This experimental fitting is achieved by superposing the asymmetric arrivals of the two Rayleigh waves traveling from the generation point and making the clockwise and anti-clockwise round trips around the center of the cylinder. The notation H_r indicates the wave arrival due to the head wave, i.e., due to the transverse mode radiated by the skimming longitudinal wave in the direction of the critical angle $\varphi_c = \arcsin\sqrt{2\mu/(\lambda+2\mu)}$. Bold characters are used in the cases when overlapping of different waves with the same arriving time causes ambiguity in their identification.

The top curve in Fig. 2 shows the experimentally observed acoustic-response signal. The middle curve displays

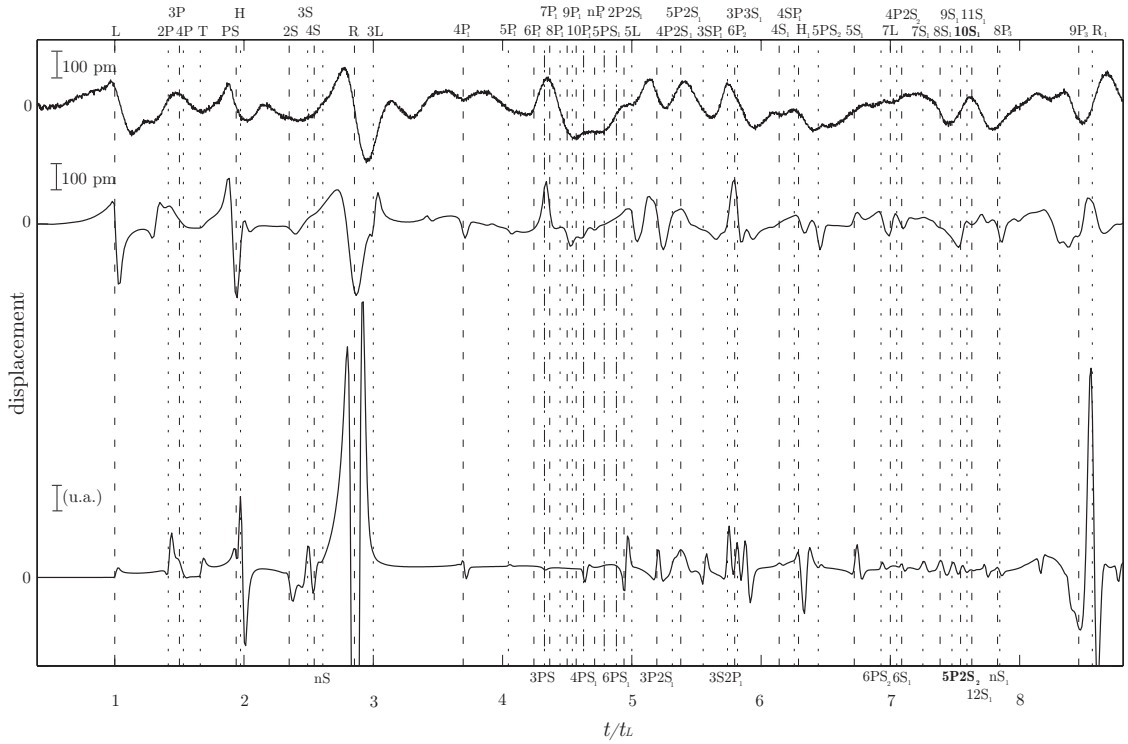


FIG. 2. Acoustic response $u_r(t)$ at the surface of the NG5 glass rod: the experimentally recorded signal (top curve), the calculation taking into account the effect of optical penetration of laser energy (middle curve), and the calculation based on the surface-source model (bottom curve). The wave-arrival notations are explained in the text.

the result of numerical simulation, which is based on the model of a radially distributed thermoelastic source outlined in Sec. III [see Eqs. (7)–(10)]. The overall amplitude scale of the numerical curve has been adjusted to the experimental scale by taking the intensity factor E_0 in Eq. (3) equal to 57 μJ mm^{-1}, which corresponds to the emission energy range of the laser. It is seen that the arrival times, shapes, and relative amplitudes of the experimental and theoretical wave forms are in reasonable agreement. This confirms consistency of the modeling. However, there yet exist some noticeable differences between the predicted and measured wave forms. The frequency content of the measured data is lower than predicted, and the narrowest arrivals are not observed. The reasons for these discrepancies could be due to using the simplified model that disregards acoustic attenuation and does not precisely account for certain subtle details of the experimental procedure such as controlling the linewidth of the laser source and its incidence angle, as well as adjusting the epicentrial position for the detection. For highlighting the effect of laser-energy penetration into the irradiated optically absorptive cylinder, the bottom curve in Fig. 2 presents the result of calculation performed, with the same parameters, on the basis of the surface dipole source model, which assumes vanishingly small optical penetration. While the arrival times are obviously in good accordance with the upper curves, there is an evident discrepancy in the shape and amplitude of main wave arrivals. For instance, there is no prominent difference between the amplitudes of bulk-wave and surface-wave arrivals on the upper curves, whereas the first Rayleigh-wave arrivals R and R_1 are clearly predominant on the bottom curve. Another significant dissimilarity is related to the first longitudinal-wave arrival L. This peak on the upper curves has a bipolar shape and an exponential decay corresponding to the laser-energy absorption in the bulk of

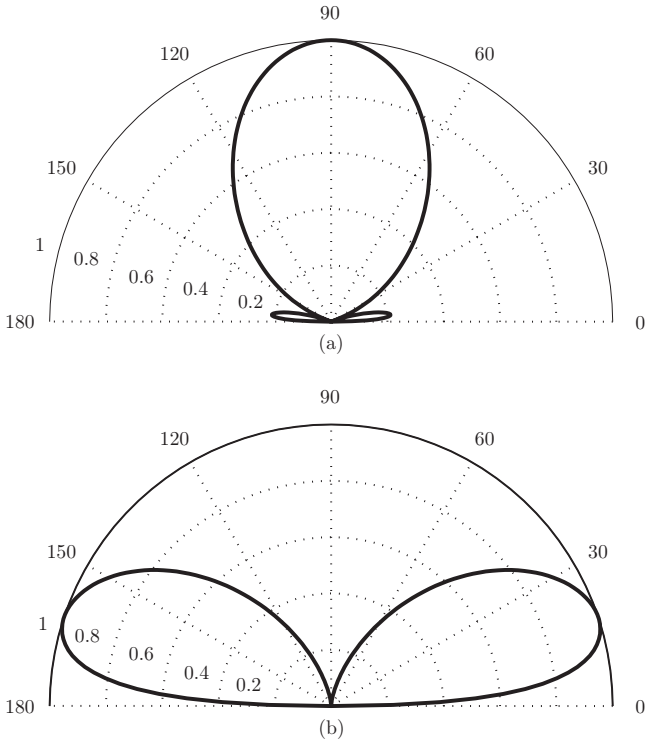


FIG. 3. Directivity patterns of radiation of longitudinal wave by (a) the volume-distributed source and (b) the surface dipole source into the half-space consisting of NG5 glass material.

the optically absorptive cylinder. By contrast, the same peak on the bottom curve has a unipolar shape, which is typical of the surface-source model.[22] A relatively larger amplitude of the longitudinal-wave arrival L due to the optical penetration can be explained through the approximate directivity patterns (Fig. 3), which assume radiation into a half-space in view of the small beamwidth-to-radius ratio $b/R=0.15/2.5$ in hand. The patterns are calculated for the wavelength $\lambda_L=2\alpha^{-1}$, which is characteristic of a distributed source with the penetration length α^{-1}.[15] It is seen that the actual, volume-distributed, source radiates the longitudinal wave essentially in the normal direction [Fig. 3(a)], whereas the surface dipole source radiates this wave into a pair of symmetric lobes with maxima close to the surface [Fig. 3(b)]. That is why the surface-source model substantially underestimates the amplitude of the longitudinal-wave arrival L.

V. CONCLUSIONS

Laser ultrasonics experiment has been performed on an optically absorptive NG5 colored glass rod. Experimentally recorded acoustic-response signal has been compared with the calculation, which relies on the model of radially distributed thermoelastic source owing to laser-energy penetration into the irradiated cylinder. This calculation is in good agreement with the experiment. Comparison with another simulation, based on the model of surface dipole source appropriate for opaque materials, reveals a marked effect of optical penetration on the shape and amplitude of main wave arrivals.

This work has dealt with a line-focused laser pulse. The future outlooks are related to a point laser source, which would open up interesting perspectives for picosecond ultrasonics generation in optically absorptive micrometric fibers and spheres.

[1]D. Royer, E. Dieulesaint, X. Jia, and Y. Shui, "Optical generation and detection of surface acoustic waves on a sphere," Appl. Phys. Lett. **52**, 706–708 (1988).

[2]D. Clorennec, D. Royer, and S. Catheline, "Saw propagation on cylindrical parts using laser-ultrasonics: Application to crack detection," Proc.-IEEE Ultrason. Symp. **1**, 207–210 (2002).

[3]D. Clorennec, D. Royer, and H. Walaszek, "Nondestructive evaluation of cylindrical parts using laser ultrasonics," Ultrasonics **40**, 783–789 (2002).

[4]D. Clorennec and D. Royer, "Analysis of surface acoustic wave propagation on a cylinder using laser ultrasonics," Appl. Phys. Lett. **82**, 4608–4610 (2003).

[5]Y. Pan, C. Rossignol, and B. Audoin, "Acoustic waves generated by a laser line pulse in a transversely isotropic cylinder," Appl. Phys. Lett. **82**, 4379–4381 (2003).

[6]M. Perton, B. Audoin, Y. Pan, and C. Rossignol, "Numerical analysis of bulk conical waves in anisotropic cylinders; application to stiffness tensor measurement," Ultrasonics **44**, 859–862 (2006).

[7]J. F. Ready, *Effects of High Power Laser Radiation* (Academic, New York, 1971).

[8]C. B. Scruby and L. E. Drain, *Laser Ultrasonics Technique and Applications* (Hilger, Bristol, 1990).

[9]C. B. Scruby, R. J. Dewhurst, D. A. Hutchins, and S. B. Palmer, "Quantitative studies of thermally generated elastic waves in laser-irradiated metals," J. Appl. Phys. **51**, 6210–6216 (1980).

[10]L. R. F. Rose, "Point-source representation for laser-generated ultrasound," J. Acoust. Soc. Am. **75**, 723–732 (1984).

[11]D. Royer, "Mixed matrix formulation for the analysis of laser-generated acoustic waves by a thermoelastic line source," Ultrasonics **39**, 345–354 (2001).

[12]I. Arias and J. D. Achenbach, "Thermoelastic generation of ultrasound by line-focused laser irradiation," Int. J. Solids Struct. **40**, 6917–6935 (2003), special issue in honor of George J. Dvorak.

[13]M. Dubois, F. Enguehard, and L. Bertrand, "Analytical one-dimensional model to study the ultrasonic precursor generated by a laser," Phys. Rev. E **50**, 1548–1551 (1994).

[14]D. Ségur, A. Shuvalov, Y. D. Pan, N. Chigarev, C. Rossignol, and B. Audoin, "Acoustic waves generated by a laser point pulse in a micrometric fiber," J. Acoust. Soc. Am. **123**, 3157 (2008).

[15]C. Thomsen, H. T. Grahn, H. J. Maris, and J. Tauc, "Surface generation and detection of phonons by picosecond light pulses," Phys. Rev. B **34**, 4129–4138 (1986).

[16]W. Nowacki, *Dynamics Problems of Thermoelasticity* (Noordhoff, Leiden, 1975).

[17]E. Hecht, *Optics* (Addison-Wesley, San Francisco, 2002).

[18]E. Kamke, *Differential Equations: Solutions and Solution Methods* (Chelsea, New York, 1971).

[19]D. Ségur, A. L. Shuvalov, Y. Pan, and B. Audoin, "Semi-analytical model of acoustic-wave generation by a laser line pulse in a transparent isotropic cylinder," J. Phys.: Conf. Ser. **195**, 012007 (2009).

[20]J. L. Rousselot, in *Acoustic Scattering by Elastic Targets With Simple Geometric Forms: Theory and Experiences*, edited by N. Gespa (CEDOCAR, Paris, 1987), Chap. 21.

[21]Y. Pan, C. Rossignol, and B. Audoin, "Acoustic waves generated by a laser line pulse in cylinders; application to the elastic constants measurement," J. Acoust. Soc. Am. **115**, 1537–1545 (2004).

[22]K. L. Telschow and R. J. Conant, "Optical and thermal parameter effects on laser-generated ultrasound," J. Acoust. Soc. Am. **88**, 1494–1502 (1990).

Analytical study of the acoustic field in a spherical resonator for single bubble sonoluminescence

Damián Dellavale, Raúl Urteaga, and Fabián J. Bonetto
Laboratorio de Cavitación y Biotecnología, Instituto Balseiro-CAB-CONICET, R8402AGP Av. Bustillo Km. 9,5 San Carlos de Bariloche-Río Negro, Argentina

(Received 28 March 2009; revised 24 September 2009; accepted 28 September 2009)

The acoustic field in the liquid within a spherical solid shell is calculated. The proposed model takes into account Stoke's wave equation in the viscous fluid, the membrane theory to describe the solid shell motion and the energy loss through the external couplings of the system. A point source at the resonator center is included to reproduce the acoustic emission of a sonoluminescence bubble. Particular calculations of the resulting acoustic field are performed for viscous liquids of interest in single bubble sonoluminescence. The model reveals that in case of radially symmetric modes of low frequency, the quality factor is mainly determined by the acoustic energy flowing through the mechanical coupling of the resonator. Alternatively, for high frequency modes the quality factor is mainly determined by the viscous dissipation in the liquid. Furthermore, the interaction between the bubble acoustic emission and the resonator modes is analyzed. It was found that the bubble acoustic emission produces local maxima in the resonator response. The calculated amplitudes and relative phases of the harmonics constituting the bubble acoustic environment can be used to improve multi-frequency driving in sonoluminescence.
 [DOI: 10.1121/1.3257208]

PACS number(s): 43.35.Hl [CCC] Pages: 186–197

I. INTRODUCTION

Liquid-filled resonant systems with simple symmetrical shapes are extensively used in basic and applied investigation areas. The eigenfrequencies of the spherical shells and fluid-filled spherical resonators have been studied in detail.[1–3] In particular, the radially symmetric modes of spherical acoustic resonators have been traditionally applied in the study of thermophysical properties of fluids[3] and more recently in single bubble sonoluminescence (SBSL).[4] This is due to the significant practical advantages of the radially symmetric modes: Resonances have high quality factor (Q), their resonance frequencies are first-order insensitive to geometrical imperfections of the solid shell,[3] and resonance widths of these non-degenerate modes are determined by the energy balance (i.e., the Q of the mode). In contrast, the degeneracy of the non-radially symmetric modes is lifted by boundary shape perturbations.[3] Thus, the resulting resonance widths depend of the geometrical imperfections and transducer properties as well as the Q. With regard to SBSL, the radially symmetric modes allow to drive a SL bubble at high acoustic pressures at the resonator center.

Recently researches in SBSL have been carried out using high viscosity liquids,[5–8] with values in the range 25–200 times the water viscosity. These studies have shown that liquid viscosity produces relevant effects on SBSL stability: shift of the Rayleigh–Taylor boundary stability[9] and generation of quasiperiodic bubble trajectories.[10,11] In addition, the bubble acoustic emission is constituted by harmonics of the driving frequency. The resulting harmonics from the interaction with the resonant system produces strong effects on the bubble stability and SL intensity.[12–14] In the scientific literature the non-linear bubble dynamic have been numerically modeled.[9,15,16] The excitation of high frequency modes of the resonator due to the outgoing shock wave produced by the bubble collapses have been experimentally observed.[13,14] The effect of the observed higher harmonics over the bubble stability has been numerically calculated.[13] However, there are missing calculations about the amplitudes and relative phases of the harmonics constituting the bubble acoustic environment nor how are they related to the liquid and solid shell properties. We believe that this kind of analysis is essential to understand and improve the multi-frequency driving in SBSL.[8,12,17] Moreover, understanding of the acoustic field far from the bubble within the resonator is important in SL researches because it allows performing non-invasive measurements of the resonant system. Such measurements are relevant for SBSL experiments that use chemically aggressive liquids such as sulfuric acid (SA) or phosphoric acid (PA) aqueous solutions. In this work we propose a model to describe the acoustic field, far from the bubble. The approach developed here is focused on four major points of interest: (1) The effect of the liquid viscosity, (2) the acoustic energy balance in the resonant system, (3) the effect of elastic properties of the spherical shell, and (4) the interaction between the bubble acoustic emission and the radially symmetric modes of the spherical resonator.

II. SYSTEM DESCRIPTION

Figure 1 shows the cross section of the resonant system. The viscous liquid fills the spherical solid thin shell which is surrounded by the gaseous medium of infinite extension. The two cylindrical ducts are joined to the spherical shell in diametrically opposed positions. These ducts are the filling ports

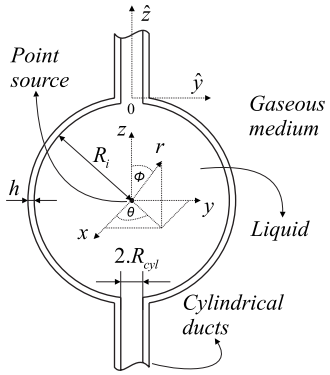


FIG. 1. Cross section of the acoustic resonator.

and the mechanical support of the actual resonator. Thus, the resonator is coupled to the external world through the gaseous medium as well as the cylindrical ducts.

III. ANALYTICAL MODEL FORMULATION

In this section we describe the equations for the resonant system. We considered displacements in the liquid and deformations in the solid shell that are small enough to be well described by linear equations. Thus, the acoustic waves in the viscous fluid are characterized by Stoke's wave equation:[18–20]

$$\frac{\partial^2 p}{\partial t^2} - \nabla^2\left[d \cdot \frac{\partial p}{\partial t} + c_0^2 \cdot p \right] = 0. \tag{1}$$

In Eq. (1), t is the time, ∇^2 is the Laplacian operator, p is the excess pressure, c_0 is the speed of sound in the fluid in absence of viscous loss, and d is the dissipative coefficient. The latter is related to the dynamic viscosity μ, coefficient of bulk viscosity η, and fluid density ρ_0 as follows:[20,21] $d=\eta+(4/3)\cdot(\mu/\rho_0)$. The bulk viscosity becomes relevant in case of big deformations like non-linear propagation of shock waves.[21] Therefore, for the linear acoustic model we have taken Stoke's assumption:[22] $\eta=0$.

A. Linear boundary value problem for the axisymmetric vibration of the spherical resonator

Taking into account the spherical coordinates of Fig. 1, the solution for the axisymmetric case (all derivatives with respect to θ vanish) can be written in the factorized form

$$p(r,\phi,t) = F(r)\cdot E(\phi)\cdot D(t). \tag{2}$$

By imposing the boundary condition that $E(\phi)$ and $\partial E(\phi)/\partial\phi$ are finite at the extremes ($\phi=0$; $\phi=\pi$), the solutions for the permanent time-harmonic evolution are well known:

$$D(t) = e^{i\cdot\omega\cdot t},$$

$$E_m(\phi) = \frac{1}{2^m\cdot m!}\cdot\frac{d^m}{dx^m}(x^2-1)^m, \quad x=\cos(\phi),$$

$$F_m(r) = C_1\cdot j_m(k\cdot r) + C_2\cdot y_m(k\cdot r), \tag{3}$$

where $D(t)$ is the temporal evolution, i is imaginary unity, ω is the angular frequency, $E_m(\phi)$ are the Legendre polynomials (Rodrigues formula) with $m=0,1,2,\ldots$, and $F_m(r)$ is a linear combination of the mth-order spherical Bessel functions (j_m and y_m). The separation constant k is the dispersion relation:

$$k = \frac{\omega\cdot\sqrt{c_0^2 - i\cdot\omega\cdot d}}{\sqrt{c_0^4 + d^2\cdot\omega^2}} = \frac{\omega}{c} + i\cdot\alpha. \tag{4}$$

From Eq. (4) we obtain the attenuation coefficient α, the phase speed c, and the propagation coefficient $\beta=\omega/c$. It is possible to identify a characteristic frequency of the viscous fluid:[19] $\omega_v=c_0^2/d$. In defining the normalized frequency $\omega_N=\omega/\omega_v$, the normalized coefficients are as follows:[19,20] $\alpha_N=\alpha\cdot d/c_0$, $\beta_N=\beta\cdot d/c_0$, and $c_N=\omega_N/\beta_N=c/c_0$. In case of $\omega_N\ll 1$ (wave propagation behavior), we have $|\alpha_N|\propto\omega_N^2$, $\beta_N\propto\omega_N$, and $c_N\approx 1$, whereas $\omega_N\gg 1$ (diffusive behavior) produces $|\alpha_N|\propto\sqrt{\omega_N}$, $\beta_N\propto\sqrt{\omega_N}$, and $c_N\propto\sqrt{\omega_N}$ (see Fig. 1 in Ref. 20).

The boundary condition in the fluid-solid interface is given by the dynamic of the solid shell. For small enough thickness to radius ratios, the thin shell approximation can be assumed and the membrane theory is applicable.[23,24] In considering the axisymmetric motion of the spherical shell without the ducts, the equations for the shell motion are[1]

$$\frac{\partial N_\phi}{\partial\phi} + N_\phi\cot(\phi) - N_\theta\cot(\phi) + r_0\cdot Y_T = r_0\cdot\rho_{\text{shell}}\cdot h\cdot\frac{\partial^2\varepsilon}{\partial t^2},$$

$$N_\phi + N_\theta + r_0\cdot Y_R = r_0\cdot\rho_{\text{shell}}\cdot h\cdot\frac{\partial^2\xi}{\partial t^2}, \tag{5}$$

where h is the shell thickness, r_0 is the mean radius of the spherical shell, ρ_{shell} is the density of the solid, $\varepsilon(\phi,t)$ and $\xi(\phi,t)$ are the tangential and radial displacements, respectively. The external loads (forces per unit area) are Y_R and Y_T for radial and tangential directions, respectively. The magnitudes of normal forces per unit length are

$$N_\phi = \frac{E\cdot h}{r_0\cdot(1-\nu^2)}\cdot\left(\frac{\partial\varepsilon}{\partial\phi} - \xi + \nu\cdot(\varepsilon\cot(\phi) - \xi)\right),$$

$$N_\theta = \frac{E\cdot h}{r_0\cdot(1-\nu^2)}\cdot\left(\varepsilon\cot(\phi) - \xi + \nu\cdot\left(\frac{\partial\varepsilon}{\partial\phi} - \xi\right)\right), \tag{6}$$

where E is the Young modulus and ν is the Poisson ratio. Substituting Eq. (5) into Eq. (6) we obtain a system of two differential equations in term of the tangential and radial displacements. Equations (1)–(6) describe the axisymmetric modes of the fluid-filled spherical shell. The proper boundary conditions are displacement equalization in the liquid-shell interface and finite pressure amplitude at the resonator center, i.e., $C_2=0$ [Eq. (3)].

B. Linear boundary value problem for the radially symmetric modes of the spherical resonator

In Sec. III E, we model a spatially stable SL bubble by considering a point source at the resonator center. Thus, the system has radial symmetry. So, henceforth we focus our study on the radially symmetric modes.

In an irrotational field, the excess pressure p and the radial velocity $\dot{\xi}=\partial\xi/\partial t$ are related to the velocity potential φ by the following equations:[19]

$$\dot{\xi}=-\bar{\nabla}\varphi,$$

$$\rho_0\cdot\frac{\partial\varphi}{\partial t}=p+\frac{d}{c_0^2}\cdot\frac{\partial p}{\partial t}. \tag{7}$$

In case of radial symmetry (all derivatives with respect to θ and ϕ vanish) and assuming permanent time-harmonic dependence, the solution of Eq. (1) can be obtained applying variable separation:

$$p_{\mathrm{liq}}=F(r)\cdot e^{i\cdot\omega\cdot t}, \tag{8}$$

where the agreement is that the excess pressure corresponds to the real part of the right-hand side of Eq. (8). Replacing Eq. (8) into Eq. (1) produces the *Helmholtz equation* in case of radial symmetry:

$$F+\frac{(c_0^2+i\cdot\omega\cdot d)}{\omega^2}\cdot\left(\frac{2}{r}\cdot F'+F''\right)=0. \tag{9}$$

In previous equation, $F'(r)$ and $F''(r)$ are the first and second spatial derivatives of the function $F(r)$. The two independent solutions of Eq. (9) are $e^{i\cdot k\cdot r}/r$, $e^{-i\cdot k\cdot r}/r$. The dispersion relation k is given by Eq. (4). The solution of Eq. (9) is obtained by the superposition of its independent solutions:

$$p_{\mathrm{liq}}=\left(A^+\cdot\frac{e^{-i\cdot k_{\mathrm{liq}}\cdot r}}{r}+A^-\cdot\frac{e^{i\cdot k_{\mathrm{liq}}\cdot r}}{r}\right)\cdot e^{i\cdot\omega\cdot t}=p_{\mathrm{liq}}^+ + p_{\mathrm{liq}}^-, \tag{10}$$

where p_{liq}^+ is the pressure wave flowing from the resonator center and p_{liq}^- is the reflected pressure wave flowing from the outer liquid-shell interface. To obtain the standing wave solution in the liquid within the resonator, we introduce the finite condition in the origin, accordingly the solution of Eq. (9) is

$$p_{\mathrm{liq}}=A_{\mathrm{liq}}\cdot j_0(k_{\mathrm{liq}}\cdot r)\cdot e^{i\cdot\omega\cdot t}, \tag{11}$$

where $j_0(k_{\mathrm{liq}}\cdot r)$ is the zero-order spherical Bessel function.

The solution of Eq. (1) in case of a traveling longitudinal wave in the gaseous medium surrounding the resonator is

$$p_g=A_g\cdot\frac{e^{-i\cdot k_g\cdot r}}{r}\cdot e^{i\cdot\omega\cdot t}=A_g\cdot F_g(r)\cdot e^{i\cdot\omega\cdot t}. \tag{12}$$

In previous equations A_{liq} and A_g are the complex magnitudes independent of the radial and time coordinates.

The displacement, velocity, and acceleration of the liquid and the gaseous medium can be obtained by replacing Eqs. (11) and (12) into Eq. (7). We define the external excitation as $p_e(t)=P_e\cdot e^{i\cdot\omega\cdot t}$ which is radially symmetric and time-harmonic with frequency $f=\omega/(2\cdot\pi)$ and amplitude P_e. Thus, considering no tangential deformations and purely radial external load for the boundary condition at the fluid-solid interface, we have $\varepsilon(\phi,t)=0$, $Y_T=0$, $Y_R=p_{\mathrm{liq}}(R_i,t)-p_g(R_e,t)-p_e(t)$. Thus, Eqs. (5) and (6) produce

$$\ddot{\xi}_{\mathrm{shell}}\cdot\rho_{\mathrm{shell}}\cdot h+\frac{2\cdot E\cdot h}{(1-\nu)\cdot r_0^2}\cdot\xi_{\mathrm{shell}}-p_{\mathrm{liq}}(R_i,t)+p_g(R_e,t)=-p_e(t). \tag{13}$$

Equation (13) is the balance of radial forces acting on a surface element of the spherical shell, where $\xi_{\mathrm{shell}}(t)=\xi_{\mathrm{shell}}\cdot e^{i\cdot\omega\cdot t}$ and $\ddot{\xi}_{\mathrm{shell}}(t)$ are the radial displacement and acceleration of the shell, respectively, $p_{\mathrm{liq}}(R_i,t)$ is the excess liquid pressure on the spherical wall, and $p_g(R_e,t)$ is the excess pressure of the external gaseous medium evaluated at the external radius R_e. Finally, according to the thin shell approximation, we introduce the equalization of displacements at the fluid-solid interfaces:

$$\xi_{\mathrm{shell}}(t)=\xi_{\mathrm{liq}}(R_i,t)=\xi_g(R_e,t). \tag{14}$$

In case of inviscid fluids ($d_{\mathrm{liq}}=d_g=0$) and $p_e(t)=0$, Eq. (13) reduces to a *standard eigenvalue problem*. As a result, the solution is an infinite set of non-harmonic eigenfrequencies. Alternatively, if $p_{\mathrm{liq}}(R_i,t)=p_g(R_e,t)=p_e(t)=0$ Eq. (13) reduces to the dynamic equation of the empty spherical thin shell in vacuum. Under these conditions, the fundamental frequency of the shell (i.e., the eigenfrequency of the breathing mode) is of the form

$$f_{\mathrm{shell}}=\frac{1}{2\cdot\pi}\cdot\frac{1}{r_0}\cdot\sqrt{\frac{2\cdot E}{\rho_{\mathrm{shell}}\cdot(1-\nu)}}. \tag{15}$$

Substituting Eqs. (11) and (12) into Eq. (13) with A_g satisfying Eq. (14) produces an expression for the coefficient A_{liq} as a function of the excitation amplitude P_e and frequency, ω, as well as the resonator parameters. Thus, the resulting complex coefficient A_{liq} properly fits displacements and pressures to satisfy the required boundary condition in the fluid-solid interface. We have performed a particular calculation for the typical resonator parameters described in Sec. IV. In considering Eq. (13) for the boundary condition of the resonant system, i.e., the liquid filled spherical shell without ducts, the model predicts $Q\approx10^4$ for the first radially symmetric mode ($j_{0,1}$). On the other hand, the measured Q in the corresponding actual resonator was about 300 in case of the first mode (sixth column of Table II). This discrepancy motivated us to consider the effect of the ducts joined to the spherical shell. What follows is the analysis of the acoustic energy balance of the resonant system. We eliminate the time-harmonic dependence $e^{i\cdot\omega\cdot t}$ and multiply Eq. (13) by $\dot{\xi}_{\mathrm{shell}}^*/2$, where $\dot{\xi}_{\mathrm{shell}}^*$ is the complex conjugate of the shell velocity. As a result, we obtain an expression of second order quantities all of which are time averaged powers. Integrating the resulting equation over the surface of the shell and taking into account Eq. (14), we obtain the following balance equation for the complex acoustic power:

$$
\begin{aligned}
& i \cdot 2 \cdot \omega \cdot \left(\frac{1}{4} \cdot \rho_{\text{shell}} \cdot |\dot{\xi}_{\text{shell}}|^2 \right) \cdot h \cdot S_{\text{shell}} \\
& \quad - i \cdot 2 \cdot \omega \cdot \left(\frac{1}{4} \cdot \frac{2 \cdot E}{(1-\nu) \cdot r_0^2} \cdot |\xi_{\text{shell}}|^2 \right) \cdot h \cdot S_{\text{shell}} \\
& \quad - \frac{p_{\text{liq}}(R_i) \cdot \dot{\xi}^*_{\text{liq}}(R_i)}{2} \cdot S_{\text{shell}} + \frac{p_g(R_e) \cdot \dot{\xi}^*_g(R_e)}{2} \cdot S_{\text{shell}} \\
& = - \frac{P_e \cdot \dot{\xi}^*_{\text{shell}}}{2} \cdot S_{\text{shell}}. \qquad (16)
\end{aligned}
$$

In Eq. (16), S_{shell} is the surface of the spherical shell. The real (imaginary) parts of terms in Eq. (16) are time averaged active (reactive) acoustic powers. On the left hand side, the first and second terms are the kinetic and potential energy rates of the shell respectively, where $2 \cdot \omega$ is the frequency of the second order quantities and i implies reactive power. The real part of the third term is the active acoustic power flowing inward the liquid, namely, the viscous dissipation rate in the liquid. The real part of the fourth term is the active acoustic power transmitted to the external gaseous medium. The real part of the term on the right-hand side is the active acoustic power supplied by the external excitation and it is flowing inward the liquid. Now, we shall consider the effect of the energy loss through the liquid within the two cylindrical ducts. The solution of Eq. (1) in case of cylindrical coordinates, and neglecting variations in pressure along the radial coordinate of the duct is

$$
p^+_{\text{cyl}} = A_{\text{cyl}} \cdot e^{-i \cdot k_{\text{liq}} \cdot \hat{z}} \cdot e^{i \cdot \omega \cdot t} = p^+_{\text{cyl}}(\hat{z}) \cdot e^{i \cdot \omega \cdot t}. \qquad (17)
$$

Previous equation is the proper solution for a traveling wave along $\hat{z}$ direction in the liquid within the cylindrical ducts (see Fig. 1). Particle displacements and velocities of the liquid within the ducts are obtained by substituting Eq. (17) in Eq. (7). In the shell-duct interface a multi-dimensional analysis is required. Results of this analysis are often implicitly incorporated into the classic lumped parameter Helmholtz resonator model (i.e., low frequency model in case of $\lambda \gg 2 \cdot R_{\text{cyl}}$), resulting in an "effective duct length." In the present work we are interested in modes of high frequency. Therefore, the resulting wavelengths are not large as compared to the diameter of the ducts. The resonant system must be considered as one having distributed constants.[25] Hence, the particle velocity and acoustic pressure into the ducts can be written as the transmitted waves through the shell-duct interface:

$$
\begin{aligned}
& p^+_{\text{cyl}}(0,t) = p^+_{\text{liq}}(R_i,t) - \Gamma_r \cdot p^-_{\text{liq}}(R_i,t), \\
& \dot{\xi}^+_{\text{cyl}}(0,t) = \dot{\xi}^+_{\text{liq}}(R_i,t) - \Gamma_r \cdot \dot{\xi}^-_{\text{liq}}(R_i,t). \qquad (18)
\end{aligned}
$$

In previous equation $\Gamma_r = A^-/A^+$ is the reflection coefficient at the shell-duct interface. The distributed acoustic impedances of the spherical and plane waves are defined as follows[25]

$$
Z_{\text{spherical}}(r) = \frac{p^+_{\text{liq}} + p^-_{\text{liq}}}{4 \cdot \pi \cdot r^2 \cdot (\dot{\xi}^+_{\text{liq}} + \dot{\xi}^-_{\text{liq}})},
$$

$$
Z_{\text{plane}}(\hat{z}) = \frac{p^+_{\text{cyl}} + p^-_{\text{cyl}}}{S_{\text{cyl}} \cdot (\dot{\xi}^+_{\text{cyl}} + \dot{\xi}^-_{\text{cyl}})}, \qquad (19)
$$

where S_{cyl} is the internal cross sectional area of the cylindrical ducts. The reflection coefficient Γ_r is obtained by equalizing the distributed acoustic impedances at the shell-duct interface: $Z_{\text{spherical}}(R_i) = Z_{\text{plane}}(0)$ and considering no reflected wave at the outer end of ducts ($p_{\text{cyl}} = p^+_{\text{cyl}}, \dot{\xi}_{\text{cyl}} = \dot{\xi}^+_{\text{cyl}}$). As a result, from Eq. (18) we obtain the pressure and radial velocity to calculate the acoustic power flowing into the two liquid cylinders:

$$
\frac{p_{\text{cyl}}(0,t) \cdot \dot{\xi}^*_{\text{cyl}}(0,t)}{2} \cdot 2 \cdot S_{\text{cyl}}. \qquad (20)
$$

The effective stiffness of the mechanical coupling for horizontal and vertical motions will produce different effect on the different resonance modes.[2] In case of the resonator depicted in Fig. 1, the ducts are the mechanical support of the resonant system. These ducts are joined to the spherical shell to form a single piece of quartz. Therefore, we have assumed that vibrations of the spherical shell transmit and propagate through the ducts. We also assume perfect acoustic impedance matching at the outer end of the ducts. Note that the no reflection condition at the outer end of the ducts could cause underestimated Q values from the model. The time averaged acoustic power flowing through the ducts wall can be written as

$$
D_{\text{duct}} = \rho_{\text{shell}} \cdot c_{\text{shell}} \cdot \frac{|\dot{\xi}_{\text{duct}}|^2}{2} \cdot 2 \cdot S_{\text{duct}}, \qquad (21)
$$

where S_{duct} is the cross sectional area of the duct wall, and $\dot{\xi}_{\text{duct}} = A_{\text{duct}} \cdot e^{i \cdot \omega \cdot (t - \hat{z}/c_{\text{shell}})}$ is the velocity of the duct wall in the direction $\hat{y}$. Referring to Fig. 1, $\hat{y}$ and $\hat{z}$ are the coordinates defining the directions of motion and propagation of the waves in the duct wall, respectively. Since the vibration is produced by the shell motion, we assume the equalization of the velocities at the shell-duct junction: $\dot{\xi}_{\text{duct}}(0,t) = \dot{\xi}_{\text{shell}}(t)$. In case of transverse waves in the duct wall, we have $c^2_{\text{shell}} = \mu_{\text{shell}}/\rho_{\text{shell}}$, where the Lamé elastic constant for the isotropic case is $\mu_{\text{shell}} = E/(2 \cdot (1+\nu))$.

We include the effect of the energy loss through the cylindrical ducts by adding Eqs. (20) and (21) into Eq. (16). Thus, eliminating the factor $\dot{\xi}^*_{\text{shell}}/2$ from the resulting equation, we obtain the boundary condition in the fluid-solid interface that satisfies the power balance:

$$
\begin{aligned}
& \ddot{\xi}_{\text{liq}}(R_i,t) \cdot \rho_{\text{shell}} \cdot h + \frac{2 \cdot E \cdot h}{(1-\nu) \cdot r_0^2} \cdot \xi_{\text{liq}}(R_i,t) - p_{\text{liq}}(R_i,t) \\
& \quad + p_g(R_e,t) + p_{\text{cyl}}(0,t) \cdot \frac{\dot{\xi}^*_{\text{cyl}}(0,t)}{\dot{\xi}^*_{\text{shell}}(t)} \cdot \frac{2 \cdot S_{\text{cyl}}}{S_{\text{shell}}} \\
& \quad + \rho_{\text{shell}} \cdot c_{\text{shell}} \cdot \dot{\xi}_{\text{shell}}(t) \cdot \frac{2 \cdot S_{\text{duct}}}{S_{\text{shell}}} = - p_e(t). \qquad (22)
\end{aligned}
$$

From Eqs. (14), (18), and (22), we obtain the expression of

the complex coefficient A_{liq} that satisfies the boundary conditions of the spherical shell including the ducts.

C. Quality factor

In order to determine the Q of the radially symmetric modes as a function of the system parameters, we shall identify the energy components of the system.

1. Energy in the liquid

The time averaged potential and kinetic energy densities in the liquid are[19]

$$U+T=\frac{1}{4\cdot\rho_0\cdot c_0^2}\cdot|p_{\mathrm{liq}}|^2+\frac{\rho_0}{4}\cdot|\dot{\xi}_{\mathrm{liq}}|^2. \tag{23}$$

The total time averaged energy stored in the liquid within the resonator can be calculated as

$$E_{\mathrm{liq}}=\int_0^{R_i}(U+T)\cdot 4\cdot\pi\cdot r^2\cdot dr. \tag{24}$$

In order to obtain E_{liq} we have numerically integrated Eq. (24) using the standard Simpson method $[O(\delta r)^4]$.

2. Energy in the spherical shell

We assume that the energy losses in the solid shell are negligible. Therefore, the time averaged total energy in the solid shell can be written as

$$E_{\mathrm{shell}}=\left[\frac{1}{4}\cdot\frac{2\cdot E}{(1-\nu)\cdot r_0^2}\cdot|\xi_{\mathrm{shell}}|^2+\frac{1}{4}\cdot\rho_{\mathrm{shell}}\cdot|\dot{\xi}_{\mathrm{shell}}|^2\right]\cdot S_{\mathrm{shell}}\cdot h. \tag{25}$$

3. Viscous dissipation in the liquid

The equalization between the viscous dissipation rate into the liquid and the flux of acoustic energy inward the liquid is written as $D_v=|I_{\mathrm{liq}}(R_i)|\cdot S_{\mathrm{shell}}$, where $I_{\mathrm{liq}}(R_i)$ is the time averaged acoustic intensity inward the liquid, and D_v is the time averaged energy loss rate due to viscous effects.

4. Energy flowing to the external medium

The acoustic energy rate transmitted to the gaseous medium can be written as $D_g=|I_g(R_e)|\cdot S_{\mathrm{shell}}$, where $I_g(R_e)$ is the time averaged acoustic intensity in the gaseous medium and D_g is the time averaged energy rate transmitted to the gaseous medium.

5. Energy flowing through the two cylindrical ducts

The acoustic energy rate transmitted into the liquid within the two cylindrical ducts joined to the spherical shell is $D_{\mathrm{cyl}}=2\cdot|I_{\mathrm{cyl}}(0)|\cdot S_{\mathrm{cyl}}$, where $I_{\mathrm{cyl}}(0)$ is the time averaged acoustic intensity in the liquid within the cylindrical ducts and D_{cyl} is the time averaged energy rate transmitted into the cylindrical ducts.

The time averaged acoustic energy rate transmitted by the transversal waves of the ducts wall D_{duct} is given by Eq. (21). In case of time-harmonic dependences and frequency $\omega=2\cdot\pi/T$, we can compute the time averaged acoustic intensities $I_{\mathrm{liq}}(R_i)$, $I_g(R_e)$, and $I_{\mathrm{cyl}}(0)$ as follows: $I(\vec{x})=\mathrm{Re}\{p(\vec{x})\cdot\dot{\xi}^*(\vec{x})/2\}$, where $I(\vec{x})$ is the time averaged acoustic intensity at a point $\vec{x}$, and $p(\vec{x})$ and $\dot{\xi}(\vec{x})$ are the acoustic pressure and velocity, respectively. Finally, the quality factor results to[26]

$$Q=\frac{2\cdot\pi}{T}\cdot\frac{(E_{\mathrm{liq}}+E_{\mathrm{shell}})}{(D_g+D_v+D_{\mathrm{cyl}}+D_{\mathrm{duct}})}. \tag{26}$$

D. Linear value problem for a periodic point source in an unboundary domain

In this section, we define the equations to describe the acoustic emission of the point source. In order to reproduce the outgoing shock wave produced by the bubble collapses, we consider the point source as one emitting periodic in time pressure pulses. If the acoustic emission in the time period T is defined by the function $p_b(t)$, Stoke's wave equation in case of radial symmetry including the proper nonhomogeneous term results to

$$\frac{\partial^2 p}{\partial t^2}-\frac{1}{r^2}\cdot\frac{\partial}{\partial r}\left(r^2\cdot\frac{\partial}{\partial r}\left(d\cdot\frac{\partial p}{\partial t}+c_0^2\cdot p\right)\right)=\frac{\delta(r)}{4\cdot\pi\cdot r^2}\cdot(p_b(t)\otimes\delta_T(t-\tau)),\quad\forall\ r>0. \tag{27}$$

The spatial location of the point source is defined by the Dirac delta function $\delta(r)$. In Eq. (27), the symbol $\otimes$ denotes convolution in time. The periodic nature of the pressure pulses is defined by the periodic impulses $\delta_T(t-\tau)$, where τ is the temporal lag. The time dependent factor of the right-hand side of Eq. (27) can be written in series form by representing the periodic impulses $\delta_T(t-\tau)$ in Fourier series and then applying the Fourier transform:

$$p_b(t)\otimes\delta_T(t-\tau)=\frac{1}{T}\cdot\sum_{n=0}^{\infty}S_n\cdot P_b(\omega_n)\cdot e^{i\cdot\omega_n\cdot(t-\tau)},$$

$$S_n=\begin{cases}1 & \forall n=0\\ 2 & \forall n\neq 0\end{cases}, \tag{28}$$

where $P_b(\omega)=\int_{-\infty}^{\infty}p_b(t)\cdot e^{-i\cdot\omega\cdot t}\cdot dt$ is the Fourier transform of the acoustic emission $p_b(t)$, and $\omega_n=n\cdot 2\cdot\pi/T$ is the angular frequency of the nth harmonic component. It is essential to note that the acoustic emission of the point source is constituted by harmonics of the fundamental frequency $\omega_1=2\cdot\pi/T$, which is defined by the temporal interval (T) between pulse occurrences. Taking into account the free space boundary conditions, the particular solution for a single harmonic component of Eq. (27) can be written in the factorized form $p_n=G_n(r)\cdot e^{i\cdot\omega_n\cdot(t-\tau)}$. Substituting into Eq. (27) produces

$$-\omega_n^2 \cdot G_n(r) - (c_0^2 + i \cdot \omega_n \cdot d) \cdot \left[\frac{2}{r} \cdot G_n'(r) + G_n''(r) \right] = \frac{\delta(r)}{4 \cdot \pi \cdot r^2} \cdot S_n \cdot \frac{P_b(\omega_n)}{T}, \quad \forall \ r > 0. \tag{29}$$

We obtain the solution of Eq. (29) by applying the bilateral Fourier transform[20,27] over the radius r:

$$G_n(r) = S_n \cdot \frac{P_b(\omega_n)}{T} \cdot \left(\frac{k_n}{\omega_n} \right)^2 \cdot \frac{e^{-i \cdot k_n \cdot r}}{4 \cdot \pi \cdot r}, \quad \forall \ r > 0, \tag{30}$$

where k_n is the complex wave number for the nth harmonic component [Eq. (4)]. The resulting acoustic pressure emitted from the point source in the free space is obtained by superposition of the harmonic solutions:

$$p = \sum_{n=0}^{\infty} G_n(r) \cdot e^{i \cdot \omega_n \cdot (t-\tau)}, \quad \forall \ r > 0. \tag{31}$$

E. Linear boundary value problem for the spherical resonator including the point source in its center

The linear model approximation for the resonator including the bubble is valid since the non-linear propagation of the outgoing shock wave, produced by the bubble collapse, vanishes within few microns from the bubble.

The linear model for the spherical resonator with a point source in its center is described by Eqs. (27) and (28) in conjunction with the proper boundary conditions. The most general solution of Eq. (27) is constituted by the sum of the homogeneous [Eq. (11)] and particular [Eq. (30)] solutions.[19] Disregarding the spectral component in $n=0$ the general solution can be written as

$$p_{\text{liq}} = \sum_{n=1}^{\infty} (A_{\text{liq},n} \cdot j_{0,n}(k_{\text{liq},n} \cdot r) + G_{\text{liq},n}(r) \cdot e^{-i \cdot \omega_n \cdot \tau}) \cdot e^{i \cdot \omega_n \cdot t}, \quad \forall \ r > 0. \tag{32}$$

In Eq. (32), the exponent $\omega_n \cdot \tau$ is the phase shift between the acoustic emission of the point source and the external excitation $p_e(t)$.

The boundary conditions for the harmonic components of Eq. (32) are given by Eqs. (13) and (14) in considering the spherical shell without the ducts. Eqs. (14), (18), and (22) are the boundary conditions that include the effect of the ducts joined to the spherical shell. Thus, the complex coefficient $A_{\text{liq},n}$ ensures that Eq. (32) satisfies the required boundary conditions for the harmonic components. We consider a time-harmonic driving pressure of frequency $f_{0,1}$, where $f_{0,1}$ is the eigenfrequency of the first radially symmetric mode $(j_{0,1})$. Since the frequency of the driving defines the acoustic emission period of the point source $(T=1/f_{0,1})$, the harmonic frequencies of Eq. (28) results to $\omega_n = n \cdot 2 \cdot \pi \cdot f_{0,1}$. Then, the excitation term of Eqs. (13) and (22) is defined as $P_{e,n}(t) = \delta_{1n} \cdot P_e \cdot e^{i \cdot \omega_n \cdot t}$, where δ_{1n} is the Kronecker delta.

TABLE I. Liquid properties.

Fluid	Density ρ_0 (kg/m^3)	Viscosity μ (Pa×s)	Sound speed c_0 (m/s)
Air	1.2	1.877×10^{-5}	340
Water	1000	1.002×10^{-3}	1482
SA 85% wt.	1778.6	0.015	1473
SA 98% wt.	1831	0.0254	1470
PA 100%	1836	0.14	1500

IV. PARTICULAR CALCULATIONS

We shall focus the calculations in a set of common values for system parameters in sonoluminescence. We consider a quartz made spherical shell whose physical properties are density $\rho_{\text{shell}}=2200$ kg/m^3. Young's modulus $E=7.306 \times 10^{10}$ Pa, and Poisson's ratio $\nu=0.171$. The geometrical parameters are the external radius $R_e=R_i+h=29.2$ mm, shell and wall ducts thickness $h=0.9$ mm and the internal radius of the cylindrical ducts $R_{\text{cyl}}=3$ mm. Table I summarizes the physical properties of viscous fluids considered in the resonant system.

In Table I, the fourth and fifth rows correspond to the sulfuric acid aqueous solution in cases of 85% and 98% in weight concentrations, respectively. The last row corresponds to phosphoric acid 100%. We consider the spherical shell surrounded by air as the infinite external medium.

For thickness-to-radius ratios up to approximately 0.01 the results from the membrane approximation are superimposed with those of the full theory.[23] This approximation still provide a quantitatively accurate representation for the geometrical values listed above, which produce $h/r_0 \approx 0.03$.

V. RESULTS AND DISCUSSION

In the following calculations, we have obtained the excess pressure in the liquid using Eq. (11). The displacements, velocity and acceleration of the fluid are given by Eq. (7). The boundary condition of the system is stated by Eqs. (14), (18), and (22), i.e., the spherical shell including the cylindrical ducts. We have considered the four liquids listed in Table I as filling fluids of the spherical shell. In case of water and SA 85% wt. aq. solution, the eigenfrequencies of the first mode $(j_{0,1})$ are listed in the fourth and fifth rows of Table II, respectively, whereas the calculated eigenfrequencies for the 50th radially symmetric mode $(j_{0,50})$ ranging up to 1.5 MHz. In regard to external excitation, the time averaged active acoustic power supplied to the system $W_e = \text{Re}\{P_e \cdot \dot{\xi}^*_{\text{shell}}/2\} \cdot S_{\text{shell}}$ was remained constant in the whole range of frequency.

A. Spectra of the resonator

In case of water and SA 85% wt. aq. solution, Fig. 2 shows the amplitude (relative to the first mode $j_{0,1}$) and phase (relative to the driving) of the shell acceleration as a function of the driving frequency. As expected, we obtain maximum amplitudes and 180° phase shifts at the resonances. Besides, the eigenfrequencies of the system are mainly determined by the mean radius of the shell (r_0) and

TABLE II. Measured and calculated data for spherical resonators.

Spherical shell properties	Liquid	Frequency (calculated) (Hz)	Frequency difference	Q (calculated)	Q (measured)
Pyrex[a] (OD=60 mm; h=0.9 mm)	Water	28 439($j_{0,1}$)	<1%	163	300
Pyrex[b] (OD=89 mm; h=0.25 mm)	SA 85% wt.	33 253($j_{0,2}$)	<1%	655	700
Quartz (OD=58.4 mm; h=0.9 mm)	Water	29 367($j_{0,1}$)	<3%	140	300
Quartz (OD=58.4 mm; h=0.9 mm)	SA 85% wt.	28 116($j_{0,1}$)	<3%	192	

[a]Reference 28.
[b]Reference 8.

the sound speed in the liquid (c_0). Thus, for the first 50 radially symmetric modes, a frequency difference less than 5% have been found among the eigenfrequencies corresponding to the four liquids listed in Table I.

If we take into account no solid shell ($h \to 0$), no gaseous external medium ($p_g=0$), and no external excitation ($p_e=0$), the boundary condition stated in Eq. (13) assumes the form $p_{\text{liq}}(R_i,t)=0$. In considering an inviscid liquid, the latter equation leads obtaining the natural modes for the system which we refer to as the *liquid sphere in vacuum*. In such a system we obtain a set of harmonic eigenfrequencies for the radially symmetric modes $[n \cdot c_0/(2 \cdot R_i)]$. On the other hand, taking into account a solid spherical shell of finite thickness h, the balance of forces at the liquid-shell interface is given by $\ddot{\xi}_{\text{shell}} \cdot \rho_{\text{shell}} \cdot h + \xi_{\text{shell}} \cdot 2 \cdot E \cdot h/((1-\nu) \cdot r_0^2) - p_{\text{liq}}(R_i,t)=0$. As a consequence, a set of non-harmonic eigenfrequencies for the radially symmetric modes is obtained. Therefore, the boundary condition given by Eq. (22) produces a set of non-harmonic eigenfrequencies (Fig. 2) due to the elastic properties of the solid shell.

In Fig. 3 the pressure amplitudes relative to the first mode ($j_{0,1}$) are displayed in decibel units. Figure 3(a) corresponds to the pressure amplitudes at the resonator center. Figure 3(b) shows the pressure amplitudes on the shell. In case of the four liquids, Fig. 3(b) shows a local minimum in the relative amplitude of the acoustic pressure at the second mode ($j_{0,2}$). We found that the eigenfrequency of the second mode ($f_{0,2}$) is the closest to the eigenfrequency of the shell f_{shell}=49.551 kHz [Eq. (15)]. As a consequence, in the second mode the minimum liquid-shell interaction occurs. This behavior is interesting when the resonator is driven at high acoustic pressures and cavitations on the spherical wall must be avoided. In general, the difference between f_{shell} and the eigenfrequencies of the resonator mainly depends of the elastic shell properties as well as the sound speed in the liquid. As we will explain in Sec. V B, the liquid within the cylindrical ducts have poor effect on the resonances of the system. Therefore, the relative amplitudes of the pressure shown in Figs. 2 and 3 are mainly determined by the dissipative coefficient of the liquid (d) and the solid shell elastic properties.

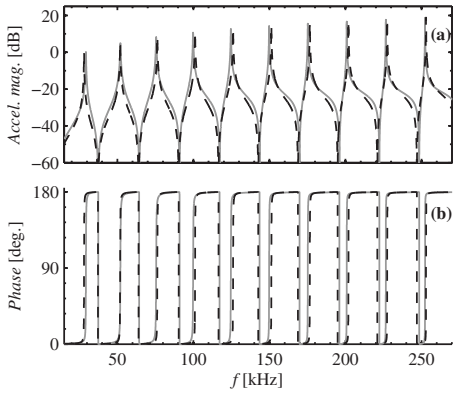


FIG. 2. Relative amplitude and phase of the shell acceleration as a function of the driving frequency. (a) Amplitude relative to the first mode $j_{0,1}$. (b) Phase relative to the driving. Gray filled and black dashed lines correspond to water and SA 85% wt. aq. solution, respectively.

B. Quality factor

Figure 4 shows the quality factor and the energy loss rates as functions of the mode number. The quality factor was calculated by two methods: (1) determining the width of the resonances at $1/\sqrt{2}$ of the maximum pressure amplitude evaluated at the resonator center and (2) performing the energy balance stated in Eq. (26). The model was consistent with these two calculation methods. We found that, in case of low frequencies, the energy balance of the system is mainly determined by the acoustic energy loss through the duct wall. Alternatively, in case of high frequencies, the dominant energy loss is due to the viscous dissipation in the liquid. The

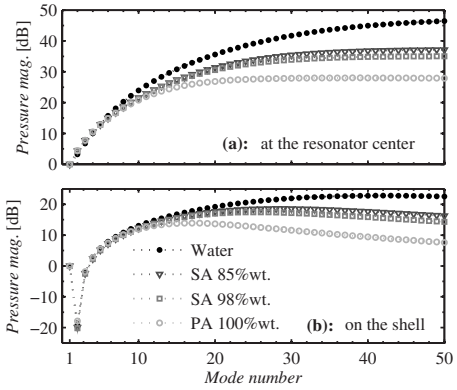


FIG. 3. Relative amplitudes of the acoustic pressure: (a) at the resonator center, (b) on the shell.

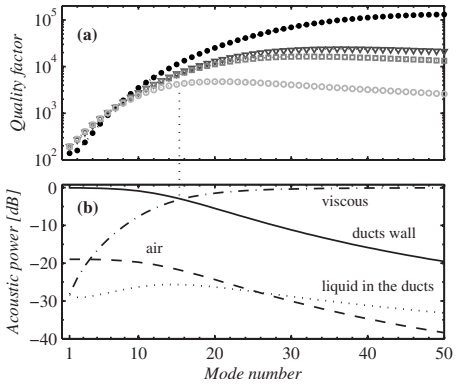


FIG. 4. (a) Quality factor as a function of the mode number: Dots, triangles, squares, and circles correspond to water, SA 85% wt., SA 98% wt., and PA 100%, respectively. (b) Energy loss rates in case of PA 100%: Energy loss rate through the duct wall (filled line), viscous dissipations rate in the liquid (dotted dashed line), energy rate flowing through the air (dashed line), and energy rate flowing through the liquid within the two cylindrical ducts (dotted line). The vertical dotted line highlights that the maximum Q value occurs when the viscous dissipation in the liquid equalizes the energy loss rate through the duct wall.

energy loss through the air and the liquid within the ducts are negligible in almost the whole range of frequency of Fig. 4.

Figure 4(b) shows the energy loss rates in case of PA 100%. The relative values of the acoustic power were obtained as $10\log_{10}(\text{energy loss rate}/W_e)$. Figure 4(a) shows that Q values have a maximum when the viscous dissipation in the liquid becomes dominant (vertical dotted line in Fig. 4) and the total energy stored in the liquid (not shown) reach the maximum. Besides, as the viscosity of the liquid decreases the viscous dissipation into the liquid becomes dominant at higher frequencies. Thus, the maximum Q values shifts to higher frequencies in turn.

Table II summarizes the experimental values of the frequencies and Q's for the radially symmetric modes of quartz and Pyrex made spherical resonators. The eigenfrequencies of the system were determined by trapping a SL bubble at the resonator center. The temperature of the resonant system was controlled during each measurement. Fourth column of Table II shows the difference between the calculated and measured eigenfrequencies. The calculated Q values listed in the fourth and fifth rows of Table II correspond to the first radially symmetric modes shown in Fig. 4(a).

In considering the liquid filled spherical shell without ducts, the model predicts $Q \approx 10^4$ for the first radially symmetric mode ($j_{0,1}$). Then, in modeling the radially symmetric modes we include the effect of the ducts in terms of the energy balance. Equation (22) is the boundary condition that includes the acoustic energy losses through the ducts. In this case, the model predicts more realistic Q values. The calculated and experimental data are listed in fifth and sixth columns of Table II, respectively. Since the acoustic impedance matching at the end of the ducts depends of the particular experimental conditions, it is not currently contained in the analytical model (perfect acoustic impedance matching is assumed). Thus, the quality factor values shown in Fig. 4(a) act as a lower bound for the Q's of the actual resonator. Though these Q values are somewhat underestimated due to model limitations, it is clear that the energy loss rate through the

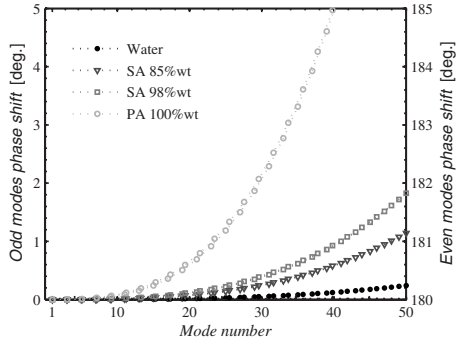


FIG. 5. Phase of the shell acceleration relative to the acoustic pressure at the resonator center.

mechanical coupling is the mechanism that limits the Q of the system at low frequencies. At higher frequencies, the viscous dissipation in the liquid becomes dominant in the energy balance of the system and the duct effect turns not important [Fig. 4(b)].

C. Relative phase

In a typical SBSL experiment, the acoustic pressure at the resonator center is usually inferred from the acoustic acceleration at the shell wall.[7,28] The latter can be measured by non-invasive methods (e.g., PZT transducer glued on the shell wall). We define the phase shift as the relative phase between the liquid acceleration at the liquid-shell interface $\ddot{\xi}_{\text{liq}}(R_i,t)$ (i.e., the shell acceleration) and the acoustic pressure at the resonator center $p_{\text{liq}}(0,t)$. Figure 5 shows the phase shift as a function of the mode number.

In considering the liquid sphere in vacuum (as defined in Sec. V A), the spatial profile of the acoustic pressure for the nth radially symmetric mode ($j_{0,n}$) has n spherical nodal surfaces ($p_{\text{liq}}=0$). Besides, the surface of the liquid sphere is one of those nodal surfaces of the pressure $[p_{\text{liq}}(R_i,t)=0]$. Furthermore, the phase shift for odd and even mode numbers is 0° and 180°, respectively. On the other hand, taking into account a solid spherical shell of finite thickness h, an external gaseous medium, and a radially symmetric and time-harmonic external excitation, the liquid pressure does not vanish at the liquid-shell interface and the balance stated in Eq. (13) is satisfied instead. Moreover, the phase shift remains the same as the previous case. Then, in considering a viscous liquid, the phase shift for odd and even mode numbers increases from 0° and 180°, respectively, with increasing viscosity and increasing frequency, as shown in Fig. 5.

D. Resonator including the point source

Regardless the resonator shape, the SL bubble is affected by a complex pressure field. This acoustic field results from the superposition of the driving pressure and its harmonics. The former is the standing wave produced by the external excitation; the latter results from the interaction between the bubble acoustic emission and the resonator response. We shall calculate the resulting harmonics of the acoustic field in case of the spherical resonator.

The SL bubble can be considered as a spatially stable acoustic source trapped at the resonator center provided low

TABLE III. Calculated parameters for SL and non-SL bubbles in case of water and SA 85% wt.

Case	Bubble conditions	No. of photons per SL pulse	$r=2.5$ mm Pressure (atm)	$r=2.5$ mm FWHM (ns)	$r=R_i$ THD of the shell acceleration (for a bandwidth of 1.5 MHz)
1	Non-SL in water $R_0=3.5$ μm; $P_a=1.1$ atm.	0	0.0151	63.4	<0.1%
2	SL in water $R_0=3.5$ μm; $P_a=1.19$ atm.	7.2×10^5	1.1	1.88	2.4%
3	Non-SL in SA 85% wt. $R_0=8$ μm; $P_a=1.1$ atm.	0	0.066	131.2	2.2%
4	SL in SA 85% wt. $R_0=8$ μm; $P_a=1.33$ atm.	2.6×10^6	2.25	6.97	59.3%

driving acoustic pressures. The limits for the driving pressure amplitude are the Rayleigh–Taylor boundary stability in case of water[28] and the Bjerknes (positional) boundary stability in case of SA 85% wt.[7]

The pressure in the liquid due to the isotropic volumetric oscillations of the bubble is obtained from the compressible Navier–Stokes equation.[4] Besides, purely radial sound field around the bubble is assumed[4] and viscous stresses are neglected. As a result, the excess pressure in the liquid can be written as[4,26]

$$p_b(r,t)=\rho_{\mathrm{liq}}\cdot\left(\frac{R}{r}\cdot(\ddot{R}\cdot R+2\cdot\dot{R}^2)-\frac{1}{2}\cdot\left(\frac{R}{r}\right)^4\cdot\dot{R}^2\right), \tag{33}$$

where R, $\dot{R}$ and $\ddot{R}$ are the radius, velocity and acceleration of the bubble wall respectively.

The bubble dynamics is determined by the local complex pressure field surrounding the bubble where the liquid may be considered incompressible and non-linear inertia forces, because of the convective accelerations, are essential.[16] In considering the far field of the bubble acoustic emission, we may approximate the bubble dynamics by assuming purely sinusoidal driving pressure. In the field far from the bubble ($r\gg R$), the weak compressibility of the liquid is essential, and the non-linear convective terms in the mass and momentum conservation equations are negligibly small.[16] Thus, the field far from the bubble implies negligible non-linear effects in the resulting acoustic emission from the bubble collapses. In this work, the dynamics of the SL bubble was obtained using a comprehensive numerical model based on the Keller version of the Rayleigh-Plessett equation generalized for non-equilibrium condensation-evaporation mass transfer at the bubble interface (see Refs. 9 and 15). In case of SA 85% wt. aq. solution due to the fact that the vapor pressure of the solution consists of water vapor, we have used the accommodation coefficient of water.[7,9] Moreover, in all cases presented here we have used argon as the dissolved gas in the liquid.

For SBSL in water, the non-linear propagation of the shock wave occurs within of about 5–200 μm from the bubble.[29–32] For larger radii the emitted pressure pulse is broadened by the spherical spreading mainly. In case of SA and PA aq. solutions, it is expected that non-linear propagation vanishes at minor radii due to the higher viscosity of these liquids. Our conservative criterion to avoid the non-linear propagation effects was to calculate the pressure pulse to be included in Eq. (27) using Eq. (33) in case of $r=2.5$ mm (i.e., the far field from the bubble $r\gg R$): $p_b(t)=p_b(r,t)|_{r=2.5\ \mathrm{mm}}$. Table III shows the calculations for SL and non-SL bubbles in case of water and SA 85% wt. The ambient radius (R_0) and acoustic pressure (P_a) values shown in Table III corresponds to typical bubble parameters observed in SBSL experiments for water[28] and SA 85% wt.[7] Fourth and fifth columns of Table III show the amplitude and full width at half maximum (FWHM) of the emitted pressure pulse, respectively. Both parameters have been calculated at $r=2.5$ mm from the bubble center. The amplitude of the emitted pressure pulse in case of a SL bubble in water (Case 2 of Table III) is consistent with the experimental data reported in Ref. 31. On the other hand, the calculated FWHM underestimates the experimental data of Refs. 31 and 33. The discrepancy in the FWHM may be due to limitations in the model used to calculate the pressure pulse [Eq. (33)]. In all calculations presented here, we have included the acoustic emission corresponding to a complete cycle of the bubble dynamics, i.e., the main collapse and subsequent rebounds.

In considering the bubble at the resonator center, the amplitude of the driving $p_e(t)$ was set to produce the P_a values listed in Table III. The frequency of the driving was set to the eigenfrequency of the first radially symmetric mode of the resonator ($f_{0,1}$ listed in fourth and fifth rows of Table II). The harmonics of the acoustic field were calculated using Eq. (32) and taking into account the boundary condition given by Eq. (22).

Figure 6(a) shows the temporal evolution of the bubble wall [$R(t)$] in case of the SL bubble in SA 85% wt. (case 4 of Table III). The acoustic emission from the bubble collapses $p_b(t)$ is also shown in Fig. 6(a) for a complete period of the bubble dynamics. Figure 6(b) shows the resulting acoustic pressure calculated at $r=2.5$ mm from the resonator center. The complex pressure field shown in Fig. 6(b) is constituted by harmonics from the driving wave extending up to $1000f_{0,1}\approx30$ MHz. It is important to note that a bandwidth of about 150 MHz it is required to fully resolve the calculated pulse temporal width in case of the SL bubble in SA 85% wt. (FWHM=6.97 ns). However, Fig. 6(b) allows comparison of the relative amplitudes of the outgoing and refocused pressure pulses for both the main collapse (C1), and the first rebound (C2).

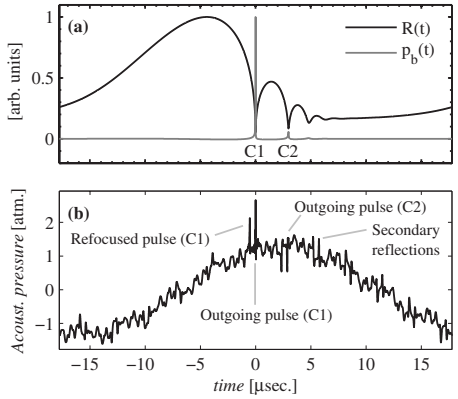


FIG. 6. (a) Normalized temporal evolution of the bubble wall (black line) in case of the SL bubble in SA 85% wt. (case 4 of Table III). The radius was normalized with $R_{max}=35.4$ μm. The emitted pressure from the bubble dynamics was calculated at $r=2.5$ mm from the bubble (gray line). The pressure was normalized with $P_{max}=2.25$ atm. (b) Pressure field at $r=2.5$ mm from the resonator center. The resulting pressure is constituted by the harmonics from the driving wave ($f_{0,1}$ listed in the fifth row of Table II) extending up to $1000f_{0,1}\approx 30$ MHz. Secondary reflections are due to C1.

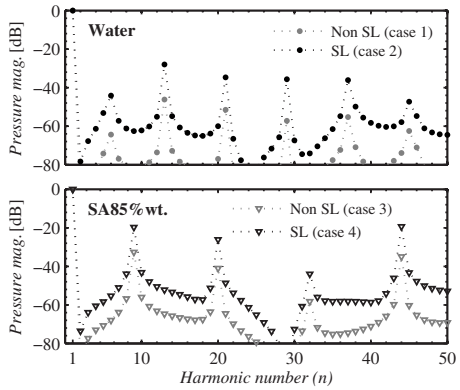


FIG. 7. Relative amplitudes of the pressure harmonics at $r=2.5$ mm from the point source. Cases 1–4 are depicted in Table III. The calculation includes the driving pressure and the acoustic emission of the bubble in one complete cycle (the main collapse and subsequent rebounds).

In our range of frequency, we have $\omega \ll \omega_v$ for the four liquid listed in Table I. As a consequence, the phase speed is $c=\omega/\beta \cong c_0$. So, the reflected pressure pulses (produced by the main collapse C1) do not suffer significant viscous attenuation or spreading in time such that weaker secondary reflection occurs. Moreover, in case of the SL bubble in water (not shown) and SA 85% wt., the pressure pulses due to the bubble rebounds (i.e., C2 in Fig. 6 for SA 85% wt.) have small amplitudes even below the amplitude of weaker secondary reflections due to C1. This is consistent with the experimental observations of Ref. 29 for SBSL in case of water and cylindrical symmetry.

In the region near the bubble the non-linear propagation of the shock wave occurs at high Mach numbers.[29–32] After this short distance (of the order of hundreds of microns), the speed of the pressure pulse is c_0 and the pulse propagates to the outer liquid-shell interface, reflects, and moves inward again. Thus, the refocused pulse reaches the bubble again about $2R_i/c_0$ seconds after leaving it. Besides, due to the boundary condition imposed by the elastic solid shell, we obtain $1/f_{0,1}<2R_i/c_0$ for the first radially symmetric mode of the resonator. As a consequence, the refocused pressure pulse reaches the bubble after its maximum compression stage. Moreover, in considering the spherical shape perturbations of the actual solid shells, the distorted reflected pressure pulse could cause significant disturbances in the bubble spatial stability. This effect could be very significant in case of the SL bubble in SA 85% wt. due to the higher amplitude of the emitted shock wave as compared to that of the SL bubble in water (see fourth column of Table III). It is important to note that in the calculations presented here the temporal evolution of the bubble dynamics do not account for the reflected pulses at the liquid-shell interface.

As described in Sec. V A, in the second radially symmetric mode of the resonator, the minimum liquid-shell interaction occurs. Besides, we have found that the eigenfrequency of the second mode produces $1/f_{0,2}\approx R_i/c_0$. As a consequence, the refocused pressure pulse reaches the bubble just about its maximum compression stage. In such a condition, the refocused pressure pulses could contribute to enhance the intensity of the bubble main collapse. This behavior also occurs for high frequency modes in which eigenfrequencies ($f_{0,n}$) are close to $n\cdot c_0/2\cdot R_i$ asymptotically (i.e., large n values).

Figure 7 shows the amplitude relative to the driving for harmonics of the acoustic pressure at $r=2.5$ mm from the point source (resonator center). In Fig. 7 the harmonic amplitudes are given from the driving wave extending up to $50f_{0,1}\approx 1.5$ MHz. In case of water and SA 85% wt., the harmonics have large amplitudes of 20 dB below that of the driving. This amplitude values are consistent with the experimental results stated in Ref. 13 for a SL bubble in water (cylindrical resonator). In addition, for non-SL bubbles (cases 1 and 3), the harmonics have large amplitudes of about 40 dB below that of the driving.

Figures 8 and 9 show the amplitude relative to the driving for the harmonics of the shell acceleration in case of water and SA 85% wt., respectively. We define Δf as the frequency difference between the harmonic of the driving and the closest eigenfrequency of the resonator. Harmonics that produce minimum $|\Delta f|$ have maximum amplitude because they produce maximum resonator response. This fact

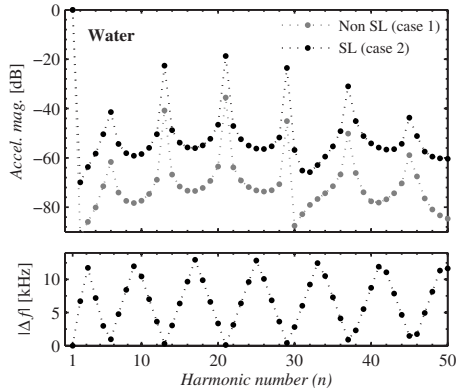


FIG. 8. (Higher graph) Relative amplitude of the shell acceleration as a function of the harmonic number (n) in case of the water filled resonator. Case 1: Non-SL bubble. Case 2: SL bubble (see Table III). (Lower graph) Frequency difference (Δf) as a function of the harmonic number (n). The parameter Δf is defined as the frequency difference between the harmonic of the driving ($n\cdot f_{0,1}$) and the closest eigenfrequency of the resonator.

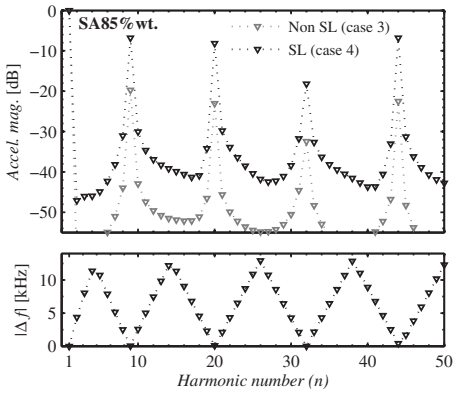


FIG. 9. (Higher graph) Relative amplitude of the shell acceleration as a function of the harmonic number in case of SA 85% wt. Case 3: Non-SL bubble. Case 4: SL bubble (see Table III). (Lower graph) Frequency difference (Δf) as a function of the harmonic number (n). The parameter Δf is defined as the frequency difference between the harmonic of the driving ($n \cdot f_{0,1}$) and the closest eigenfrequency of the resonator.

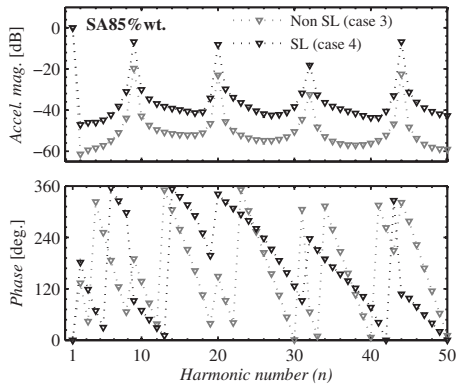


FIG. 10. Amplitude (higher graph) and phase (lower graph) of the shell acceleration relative to the driving as functions of the harmonic number for the SL and non-SL bubbles in case of SA 85% wt. The relative phase is normalized in the range 0–360°.

can be understood taking into account that the acoustic emission of the point source [Eq. (28)] is constituted by harmonics of the driving ($n \cdot f_{0,1}$). Therefore, a minimum $|\Delta f| = |n \cdot f_{0,1} - f_{0,m}|$ implies maximum mode excitation of the corresponding eigenfrequency $f_{0,m}$. We emphasize that, due to the symmetry of the system, only radially symmetric modes have been considered.

Figure 8 (case 2) shows that the acceleration has large amplitudes of 20 dB below the driving, and Fig. 9 (case 4) show that the acceleration has large amplitudes of 7 dB below the driving. Accordingly, the fourth column of Table III shows that the pressure pulse amplitude for the case 4 is about twice the amplitude of the case 2.

The large amplitudes of the harmonics are mainly determined by the intensity of the bubble acoustic emission. On the other hand, the frequency and relative amplitude of the peaks (local maxima of the harmonic amplitudes) shown in Figs. 8 and 9 are determined by the properties of the acoustic resonator (mainly by the liquid and solid shell properties) and are insensitive to the Q of the resonator. This is due to the fact that as the quality factor increases, the eigenfrequency of the modes remains almost unchanged and so the frequency difference Δf. Then, compensation occurs between decreasing full-width half-power bandwidth (Q) and increasing amplitude of the mode responses.

The decay of the acceleration amplitude as a function of $|\Delta f|$ (not shown) is approximately potential. For the harmonics in which $|\Delta f| \lesssim 1.5$ kHz we obtain a maximum resonator response (peaks in Figs. 8 and 9). In case of SA 85% wt., Fig. 10 shows that the phase of the shell acceleration relative to the driving has small phase shift between adjacent harmonics of the driving and phase jumps at the local maxima of the resonator response (peaks in higher graphs of Figs. 9 and 10). Moreover, the phase shifts between adjacent harmonics are larger in case of the SL bubble. This behavior is related with higher intensity of the acoustic emission from the SL bubble.

We have calculated the resulting total harmonic distortion ($\mathrm{THD} = \Sigma_{n=2}^{50} |\ddot{\xi}_{\mathrm{shell},n}|^2 / |\ddot{\xi}_{\mathrm{shell},1}|^2$) of the shell acceleration extending up to the 50th harmonic of the driving ($50 f_{0,1} \approx 1.5$ MHz). The THD values are listed in the sixth column of Table III. In case of water, the non-SL bubble produces a negligible THD. On the other hand, the SL bubble in SA 85% wt. produces a THD about 20 times higher than that of the SL bubble in water. These results are consistent with observations of the signal from the typical PZT transducer (microphone) glued to the resonator wall.[7,8,28]

VI. SUMMARY AND CONCLUSIONS

An analytical model for a spherical resonator has been described. The eigenfrequencies and quality factors of the radially symmetric modes have been calculated from $j_{0,1}$ up to $j_{0,50}$. In case of the first and second radially symmetric modes, we have obtained a good agreement with the experimental data in terms of the eigenfrequencies as well as the Q.

The model reveals that in case of low frequency mode, the quality factor is mainly determined by the acoustic energy flowing through the mechanical coupling of the resonator. Then, the exact prediction of the Q value depends on the details of the external mechanical coupling of the system. On the other hand, for high frequency modes the quality factor is mainly determined by the viscous dissipation in the liquid and the effect of the mechanical couplings turns not important. Besides, we found that the liquid viscosity produces a phase shift between the shell acceleration and the pressure at the resonator center. This phase shift is relevant in case of large viscosity and high frequency modes.

We have used a numerical model to obtain the bubble dynamics and the acoustic emission in case of water and SA 85% wt. aqueous solution. The calculations have shown good agreement with the observed amplitude of the outgoing shock wave from the main collapse of a SL bubble in water. On the other hand, the calculated FWHM of the shock wave underestimates the experimental data. It was found that the pressure pulse amplitude from the main collapse of a typical SL bubble in SA 85% wt. is about twice the amplitude obtained in case of a SL bubble in water. The model predicts a FWHM of the outgoing shock wave about four times greater in case of a SL bubble in SA 85% wt. than of the SL bubble in water.

We found that the reflected pressure pulses produced by the main collapse do not suffer significant viscous attenua-

tion or spreading in time such that weaker secondary reflections occur. Moreover, in case of the SL bubble in water and SA 85% wt., the pressure pulses due to the bubble rebounds have small amplitudes even below the amplitude of weaker secondary reflections due to the main collapse. In considering the SL bubble driving at the eigenfrequency of the first mode, the model predicts that the converging pressure pulses reflected at the spherical shell do not affect the dynamics of the bubble main collapse. This is due to the delayed arrival of the refocused pulse relative to the bubble main collapse. On the other hand, by modifying the properties of either the solid shell or the liquid of the resonator, the arrival time of the refocused pulse relative to the bubble main collapse can be adjusted. This mechanism could be used to enhance the intensity of the bubble main collapse.

We also found that for the radially symmetric mode whose eigenfrequency is very close to the eigenfrequency of the elastic shell, the minimum liquid-shell interaction occurs. Moreover, the eigenfrequency of this mode is very close to the corresponding eigenfrequency of the liquid sphere without the solid shell. As a consequence, if the SL bubble is driven at the eigenfrequency of this mode, the refocused pressure pulse reaches the bubble just about its maximum compression stage. This behavior also occurs for high frequency modes of the resonator.

The complex acoustic field in the liquid of the resonator results from the superposition of the driving pressure and its harmonics. The former is the standing wave produced by the external excitation; the latter results from the interaction between the bubble acoustic emission and the resonator response. We found that the relative amplitude of the harmonics as a function of the harmonic number shows peaks (i.e., local maximum amplitudes). A local maximum of the harmonic amplitude occurs when the harmonic of the driving produces minimum frequency difference with an eigenfrequency of the resonator. The frequency and relative amplitude of the local maximum amplitudes of the harmonics are determined by the properties of the acoustic resonator (mainly by the liquid and solid shell properties). On the other hand, the large amplitudes of the harmonics are mainly determined by the intensity of the bubble acoustic emission.

Under the scope of the energy concentration maximization, a remaining challenge is to develop a model for the calculation of the bubble dynamics accounting for the reflected shock waves.

ACKNOWLEDGMENT

We gratefully acknowledge the contribution of Gabriela F. Puente in providing the numerical model of the bubble dynamiscs. Special thanks to Alejandro Miguel Castro and Ludmila Rechiman who participated in some discussion stages of this work. This work was funded by CONICET and CNEA.

[1]W. E. Baker, "Axisymmetric modes of vibration of thin spherical shell," J. Acoust. Soc. Am. **33**, 1749–1758 (1961).
[2]J. B. Mehl, "Spherical acoustic resonator: Effects of shell motion," J. Acoust. Soc. Am. **78**, 782–788 (1985).
[3]M. R. Moldover, J. B. Mehl, and M. Greenspan, "Gas-filled spherical resonator: Theory and experiment," J. Acoust. Soc. Am. **79**, 253–272 (1986).
[4]M. P. Brenner, S. Hilgenfeldt, and D. Lohse, "Single-bubble sonoluminescence," Rev. Mod. Phys. **74**, 425–484 (2002).
[5]D. J. Flannigan and K. S. Suslick, "Plasma formation and temperature measurement during single-bubble cavitation," Nature (London) **434**, 52–55 (2005).
[6]S. D. Hopkins, S. J. Putterman, B. A. Kappus, K. S. Suslick, and C. G. Camara, "Dynamics of sonoluminescing bubble in sulfuric acid," Phys. Rev. Lett. **95**, 254301 (2005).
[7]R. Urteaga, D. Dellavale, G. F. Puente, and F. J. Bonetto, "Positional stability as the light emission limit in sonoluminescence with sulfuric acid," Phys. Rev. E **76**, 056317 (2007).
[8]R. Urteaga and F. J. Bonetto, "Trapping an intensely bright, stable sonoluminescing bubble," Phys. Rev. Lett. **100**, 074302 (2008).
[9]G. F. Puente, P. García-Martínez, and F. J. Bonetto, "Single-bubble sonoluminescence in sulfuric acid and water: Bubble dynamics, stability and continuous spectra," Phys. Rev. E **75**, 016314 (2007).
[10]A. Moshaii and R. Sadighi-Bonabi, "Role of liquid compressional viscosity in the dynamics of a sonoluminescing bubble," Phys. Rev. E **70**, 016304 (2004).
[11]R. Toegel, S. Luther, and D. Lohse, "Viscosity destabilizes sonoluminescing bubbles," Phys. Rev. Lett. **96**, 114301 (2006).
[12]J. Holzfuss, M. Rüggeberg, and R. Mettin, "Boosting sonoluminescence," Phys. Rev. Lett. **81**, 1961–1964 (1998).
[13]J. Holzfuss, M. Rüggeberg, and R. Glynn Holt, "Acoustical stability of a sonoluminescing bubble," Phys. Rev. E **66**, 046630 (2002).
[14]F. B. Seeley, "Effects of higher-order modes and harmonics in single-bubble sonoluminescence," J. Acoust. Soc. Am. **105**, 2236–2242 (1999).
[15]G. F. Puente, R. Urteaga, and F. J. Bonetto, "Numerical and experimental study of dissociation in an air-water single-bubble sonoluminescence system," Phys. Rev. E **72**, 046305 (2005).
[16]R. I. Nigmatulin, I. S. Akhatov, N. K. Vakhitova, and R. T. Lahey, "On the forced oscillations of a small gas bubble in a spherical liquid-filled flask," J. Fluid Mech. **414**, 47–73 (2000).
[17]F. J. Moraga, R. P. Taleyarkhan, R. T. Lahey, and F. J. Bonetto, "Role of very-high-frequency excitation in single-bubble sonoluminescence," Phys. Rev. E **62**, 2233–2237 (2000).
[18]G. G. Stokes, "On the theories of the internal friction of fluids in motion, and of the equilibrium and motion of elastic solids," Trans. Cambridge Philos. Soc. **8**, 287–319 (1845).
[19]P. M. Morse and K. Uno Ingard, *Theoretical Acoustics* (McGraw-Hill, New York, 1970).
[20]M. J. Buckingham, "Causality, Stokes' wave equation, and acoustic pulse propagation in a viscous fluid," Phys. Rev. E **72**, 026610 (2005).
[21]J. Lighthill, *Waves in Fluids* (Cambridge University Press, New York, 2003).
[22]P. K. Kundu and I. M. Cohen, *Fluid Mechanics* (Academic, San Diego, 2002).
[23]W. Soedel, *Vibrations of Shells and Plates* (Dekker, New York, 2004).
[24]Stephen P. Timoshenko, *Theory of Plates and Shells* (McGraw-Hill, New York, 1940).
[25]Lawrence E. Kinsler and Austin R. Frey, *Fundamentals of Acoustics* (Wiley, New York, 1962).
[26]T. G. Leighton, *The Acoustic Bubble* (Academic, London, 1994).
[27]B. Van Der Pol and H. Bremmer, *Operational Calculus Based on the Two-sided Laplace Integral* (Cambridge University Press, London, 1955).
[28]R. Urteaga, D. Dellavale, G. F. Puente, and F. F. Bonetto, "Experimental study of transient paths to the extinction in sonoluminescence," J. Acoust. Soc. Am. **124**, 1490–1496 (2008).
[29]J. Holzfuss, M. Rüggeberg, and A. Billo, "Shock wave emissions of a sonoluminescing bubble," Phys. Rev. Lett. **81**, 5434–5437 (1998).
[30]T. J. Matula, I. M. Hallaj, O. Cleveland, L. A. Crum, W. C. Moss, and R. A. Roy, "The acoustic emissions from single-bubble sonoluminescence," J. Acoust. Soc. Am. **103**, 1377–1382 (1998).
[31]Z. Q. Wang, R. Pecha, B. Gompf, and W. Eisenmenger, "Single bubble sonoluminescence: Investigations of the emitted pressure wave with a fiber optic probe hydrophone," Phys. Rev. E **59**, 1777–1780 (1999).
[32]R. Pecha and B. Gompf, "Microimplosions: Cavitation collapse and shock wave emission on a nanosecond time scale," Phys. Rev. Lett. **84**, 1328–1330 (2000).
[33]K. R. Weninger, B. P. Barber, and S. J. Putterman, "Pulsed Mie scattering measurements of the collapse of a sonoluminescing bubble," Phys. Rev. Lett. **78**, 1799–1802 (1997).

Excitation and focusing of Lamb waves in a multilayered anisotropic plate

Bastien Chapuis
Department of Composite Materials and Structures, ONERA, 29 Avenue de la Division Leclerc, 92322 Châtillon Cedex, France

Nicolas Terrien
CETIM, 74 Route de la Jonelière, 44326 Nantes Cedex 3, France

Daniel Royer
Laboratoire Ondes et Acoustique, ESPCI, Université Paris 7, CNRS UMR 7587, 10 Rue Vauquelin, 75231 Paris Cedex 05, France

(Received 15 June 2009; revised 13 October 2009; accepted 13 October 2009)

The radiation of Lamb waves by an axisymmetric source on the surface of an anisotropic plate is investigated. An analytical expression of the Green's function, valid in the far field domain, is derived. This approximation shows that the anisotropy of the propagation medium induces a focusing of Lamb modes in some directions, which correspond to minima of the slowness. Numerical simulations and experiments demonstrate that for the fundamental A_0 and S_0 modes, this phenomenon, analog to the phonon focusing effect, can be very strong in materials such as composite fiber-reinforced polymers. This effect due to the plate anisotropy must be correctly taken into account, for example, in order to develop systems for *in situ* structural health monitoring. The choice of the most appropriate Lamb mode, the excitation frequency, and the design of the array of piezoelectric disks used as transmitters and receivers depends on such considerations.
 [DOI: 10.1121/1.3263607]

PACS number(s): 43.35.Zc, 43.20.Mv, 43.35.Cg [YHB] Pages: 198–203

I. INTRODUCTION

Nondestructive testing is a technique to detect and image defects in a material, while structural health monitoring (SHM) can be defined as a continuous process to monitor the integrity of a structure. The study of the propagation of Lamb waves has become increasingly popular over the past three decades due to their potential ability to inspect large structures such as plates or pipes.[1–3] They have also been used to characterize isotropic or anisotropic materials as they can provide the elastic matrix of the material.[4] The growing interest for composites, in particular for aeronautical industry or for wind turbines, has justified the large amount of work on the propagation of Lamb waves in that kind of layered and anisotropic materials.[5] In order to inspect the region of interest, multiple transmitters and receivers are embedded in the structure. Thus, the practical use of Lamb waves for SHM of plate-like structures needs a better understanding of the generation and detection of these waves by piezoelectric transducers (PZTs).[6] For example, to avoid the existence of dead zones, the design of transducer arrays used for SHM requires a more precise description of the radiated guided wave field. The role of the anisotropy on the directivity pattern of bulk acoustic waves[7,8] and Rayleigh waves[9] has already been studied. Within the geometrical acoustics approach, the angular dependence of the acoustic power was found to be proportional to the focusing factor introduced by Maris.[10] Using the same assumption, focusing of elastic waves guided by thin anisotropic plates has also been investigated.[11,12] The contribution of the present work is to show that the Maris factor can explain focusing in anisotropic plates for dispersive Lamb modes at any frequency, including an experimental verification in the case of very strong focusing.

This paper presents theoretical developments, numerical results, and related experiments on the excitation of Lamb waves by a set of sources distributed on the surface of anisotropic plates. In the first part, a far field approximation giving the amplitude of Lamb modes radiated by an axisymmetric source is derived. The analysis is based on the three dimensional (3D) Green's function given in Ref. 13. It is shown that the anisotropy of the propagation medium induces a focusing of Lamb waves in particular directions. In the second part, numerical examples of this focusing effect in a fiber-reinforced composite are provided for A_0 and S_0 Lamb modes. In the last part of the paper, these predictions are experimentally verified.

II. GENERAL DESCRIPTION

A multilayered anisotropic plate with Cartesian coordinates (x_1,x_2,x_3) is considered with the x_3 axis normal to the surface. As shown in Fig. 1, the source (i.e., the PZT) is modeled by a force distribution acting on the upper surface of the plate at $x_3=0$. Using the Fourier transform method to solve the wave equation, the displacement field $\mathbf{u}$ produced by a time-harmonic force $\mathbf{f}(x_1,x_2)e^{-i\omega t}$ is given by

$$\mathbf{u}(x_1,x_2,x_3) = \int\int \mathbf{g}(x_1-\tilde{x}_1,x_2-\tilde{x}_2,x_3) \times \mathbf{f}(x_1-\tilde{x}_1,x_2-\tilde{x}_2)\mathrm{d}\tilde{x}_1\mathrm{d}\tilde{x}_2, \quad (1)$$

where $\mathbf{g}(x_1,x_2,x_3)$ is the 3D Green's function of the system.

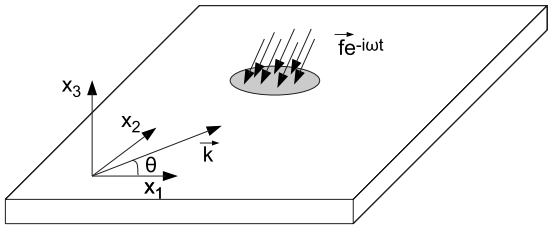


FIG. 1. System geometry.

This function can be expressed from its two dimensional spatial Fourier transform $\mathbf{G}$,[13]

$$\mathbf{g}(x_1,x_2,x_3)=\frac{1}{(2\pi)^2}\int\int \mathbf{G}(k_1,k_2,x_3)e^{i(k_1x_1+k_2x_2)}\mathrm{d}k_1\mathrm{d}k_2, \tag{2}$$

where $\mathbf{G}(k_1,k_2,x_3)$ stands for the Green's function for plane waves propagating in the direction given by the wave vector $\mathbf{k}(k_1=k\cos\theta,k_2=k\sin\theta,0)$, where θ is the angle between the direction of the phase velocity and the x_1 axis (Fig. 2).

For a given source/receiver configuration, defined by the distance $r=\mathrm{OP}$ and the angle φ: $x_1=r\cos\varphi, x_2=r\sin\varphi$, the Green's function can be written as

$$\mathbf{g}(r,\varphi,x_3)=\frac{1}{(2\pi)^2}\int_{\varphi-\pi/2}^{\varphi+3\pi/2}\int_{\Gamma}\mathbf{G}(k\cos\theta,k\sin\theta,x_3)\times e^{ikr\cos\psi}k\mathrm{d}k\mathrm{d}\theta, \tag{3}$$

where Γ is the real positive half axis and $\psi=\varphi-\theta$.

The integral with respect to θ is divided into two parts: from $\varphi-\pi/2$ to $\varphi+\pi/2$ and from $\varphi+\pi/2$ to $\varphi+3\pi/2$; in the second one, θ is transformed in $(\theta+\pi)$ and k in $-k$. Thus, the Green's function (3) can be rewritten as[13]

$$\mathbf{g}(r,\varphi,x_3)=\frac{1}{(2\pi)^2}\int_{\varphi-\pi/2}^{\varphi+\pi/2}\int_{\Gamma\cup-\Gamma}\mathbf{G}(k\cos\theta,k\sin\theta,x_3)\times e^{ikr\cos\psi}k\mathrm{d}k\mathrm{d}\theta. \tag{4}$$

In order to evaluate the far field contribution as $r\to\infty$, the integration path with respect to k is closed by a half-circle C at infinity in the upper half plane so that the factor $e^{ikr\cos\psi}$ vanishes on the semicircle. At a given angular frequency ω, the Green's function presents poles for a discrete

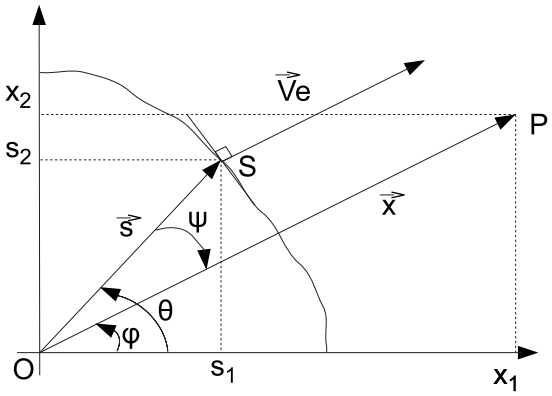


FIG. 2. At the stationary phase point S, the energy velocity $\boldsymbol{V}^e$ is normal to the slowness surface $s(\theta)$.

number of values $k_M(\theta,\omega)$ corresponding to Lamb modes M. The principle of limiting absorption can be used to deal with real poles that are on the integration contour since the dissipation effects have been neglected. For a positive pole (corresponding to a propagation in the x direction), a small positive imaginary part β is associated with the wave number: $k\to k+i\beta$ so that the propagation factor $e^{ikr\cos\psi}$ vanishes when $r\to+\infty$. Equivalently, a small imaginary part $-\beta$ is associated with a negative pole (corresponding to a propagation in the $-x$ direction) by the relation $-k\to-k-i\beta$. Due to the geometrical dispersion of the wave, the amplitude of propagating modes, described by real poles, decreases as $r^{-1/2}$ when $r\to\infty$. In the far field, the contribution of complex poles can be neglected as the corresponding modes have a larger order of decay.[14]

Using the residue theorem, the Green's function for propagative modes M is given by

$$\mathbf{g}(r,\varphi,x_3)=\frac{1}{2\pi}\sum_{\mathbf{M}}\int_{\varphi-\pi/2}^{\varphi+\pi/2}\mathbf{G}_{\mathbf{M}}(\theta,x_3)|k_M(\theta)|e^{ik_Mr\cos\psi_M}d\theta,$$

$$\mathbf{G}_{\mathbf{M}}(\theta,x_3)=i\,\mathrm{res}[\mathbf{G}(k\cos\varphi,k\sin\varphi,x_3)]_{k=k_M(\theta)}. \tag{5}$$

For each Lamb mode and for a given direction of observation φ, the integral with respect to θ can be evaluated by the stationary phase method. Only the contribution of the angle θ for which the phase $\Phi_M(\theta)=k_Mr\cos\psi$ varies slowly is considered. As $(\mathrm{d}\psi/\mathrm{d}\theta)_\varphi=-1$, θ is solution of the equation

$$\left(\frac{\mathrm{d}\Phi_M}{\mathrm{d}\theta}\right)_\varphi=\frac{\mathrm{d}k_M}{\mathrm{d}\theta}r\cos\psi_M+k_Mr\sin\psi_M=0. \tag{6}$$

The first derivative of the phase Φ_M vanishes for a value θ_M such that

$$\frac{\mathrm{d}k_M}{\mathrm{d}\theta}=-k_M\tan\psi_M\quad\text{with}\quad\psi=\varphi-\theta. \tag{7}$$

This result can be interpreted in terms of energy propagation in anisotropic media.[15] As shown in Fig. 2, Eq. (7) indicates that the propagation direction for which the phase is stationary is such that the direction joining the source O to the observation point P is parallel to the energy velocity $\mathbf{V}^e$, i.e., normal to the slowness curve $\mathbf{s}(\theta)=\mathbf{k}/\omega$ plotted at a given frequency.[16] In the following, we assume that there will only be one stationary phase point S. In this case, the slowness curve is convex for all θ, and Eq. (7) defines a one-to-one relation between angles φ and θ that will be used in the next developments.

From Eq. (6) and for a given angle φ, the second order derivative of the phase is given by

$$\left(\frac{\mathrm{d}^2\Phi_M}{\mathrm{d}\theta^2}\right)_\varphi=\left(\frac{\mathrm{d}^2k_M}{\mathrm{d}\theta^2}+2\frac{\mathrm{d}k_M}{\mathrm{d}\theta}\tan\psi_M-k_M\right)r\cos\psi_M. \tag{8}$$

Substituting the second order derivative of k_M, deduced from Eq. (7),

$$\frac{\mathrm{d}^2k_M}{\mathrm{d}\theta^2}=-\frac{\mathrm{d}k_M}{\mathrm{d}\theta}\tan\psi_M-\frac{k_M}{\cos^2\psi_M}\left(\frac{\mathrm{d}\varphi}{\mathrm{d}\theta}-1\right), \tag{9}$$

into Eq. (8) leads to

TABLE I. Parameters of the CFRP T700GC/M21 (Hexcel Composites).

C_{11} (GPa)	$C_{22}=C_{33}$ (GPa)	$C_{12}=C_{13}$ (GPa)	C_{23} (GPa)	C_{44} (GPa)	$C_{55}=C_{66}$ (GPa)	Mass density (kg/m^3)	Ply thickness (mm)
123.44	11.54	5.55	6.35	2.6	4.5	1.6×10^3	0.25

$$\left(\frac{d^2\Phi_M}{d\theta^2}\right)_\varphi = \left(\frac{dk_M}{d\theta}\tan\psi_M + k_M\tan^2\psi_M - \frac{k_M}{\cos^2\psi_M}\frac{d\varphi}{d\theta}\right) \times r\cos\psi_M. \quad (10)$$

Since, from Eq. (6), the first two terms in brackets are opposite, the second order derivative of the phase Φ_M reduces to

$$\Phi''_M = \left(\frac{d^2\Phi_M}{d\theta^2}\right)_\varphi = -\frac{k_M r}{\cos\psi_M}\frac{d\varphi}{d\theta}. \quad (11)$$

Substituting in Eq. (5), $\Phi_M(\theta)$ by its second order Taylor expansion versus θ yields

$$\mathbf{g}(r,\varphi,x_3) = \frac{1}{2\pi}\sum_M \mathbf{G}_M(\theta_M,x_3) \times |k_M| e^{i\Phi_M}\int_{\varphi-\pi/2}^{\varphi+\pi/2} e^{i(\Phi''_M/2)(\theta-\theta_M)^2}d\theta, \quad (12)$$

and finally

$$\mathbf{g}(r,\varphi,x_3) = \frac{1}{\sqrt{2\pi r}}\sum_M \mathbf{G}_M(\theta_M,x_3)e^{i[\Phi_M+\mathrm{sgn}(\Phi''_M)\pi/4]} \times\sqrt{|A_M k_M\cos\psi_M|}, \quad (13)$$

where the factor A_M is defined for each Lamb mode M by the relation

$$A = \frac{d\theta}{d\varphi}. \quad (14)$$

The angular dependence of the Green's function comes from several terms. However, in highly anisotropic materials, the dominant term is the so-called Maris factor A. This factor, which measures the acoustic ray density anisotropy, has been postulated by Maris in order to explain the phonon focusing effect.[10] For a given mode, it can be calculated from the slowness curve $s(\theta)$ in the (x_1,x_2)-plane as[11]

$$A = [s^2 + (ds/d\theta)^2]^{-1/2}|K_s|^{-1}, \quad (15)$$

where K_s is the curvature of the slowness curve.

Equation (13) shows that, in the far field domain, this phenomenon is at the origin of the focusing of the Lamb waves. This result generalizes the description given in Ref. 12 for the SH_0 mode using a low frequency-thickness approximation. It should be noted that if the curvature vanishes, the stationary phase approximation breaks down. It occurs at the transition point between convex and concave regions of the slowness curve. These points correspond to the presence of caustics for which the focusing factor becomes infinite. The existence of a singularity for the Green's function is due to the restricted nature of the ray approach.

III. NUMERICAL SIMULATIONS

In some anisotropic materials, the focusing effect can be very strong. In this section, some numerical computations of slowness curves and Maris factors for a $[0°/90°]_S$-T700GC/M21 cross-ply composite fiber-reinforced polymer (CFRP) plate are presented. The parameters used in these simulations are listed in Table I. Dispersion curves show that in any direction of the plate only three modes can propagate below 500 kHz: the two fundamental Lamb modes A_0 and S_0 and the first shear horizontal mode SH_0. However, it should be noted that, except for 0° and 90° directions, the polarizations of S_0 and SH_0 modes are never purely radial nor ortho-radial. In the following S_0 denotes the quasi-longitudinal mode and SH_0 denotes the quasi-shear mode. In the next part, we will show that S_0 and SH_0 modes are, in fact, coupled in some direction by the axisymmetric source used in the experiments (see also Ref. 13 for the case of an out-of-plane source).

The theoretical presentation previously derived showed that Maris factor is intimately related to the slowness curve. In order to compare with experiments presented in Sec. IV, slowness curves and corresponding focusing factors are plotted in Fig. 3 for A_0 mode at 40 kHz and for S_0 and SH_0 modes at 300 kHz. For the A_0 mode, the energy is essentially focused in the direction 0°, but a second local maximum

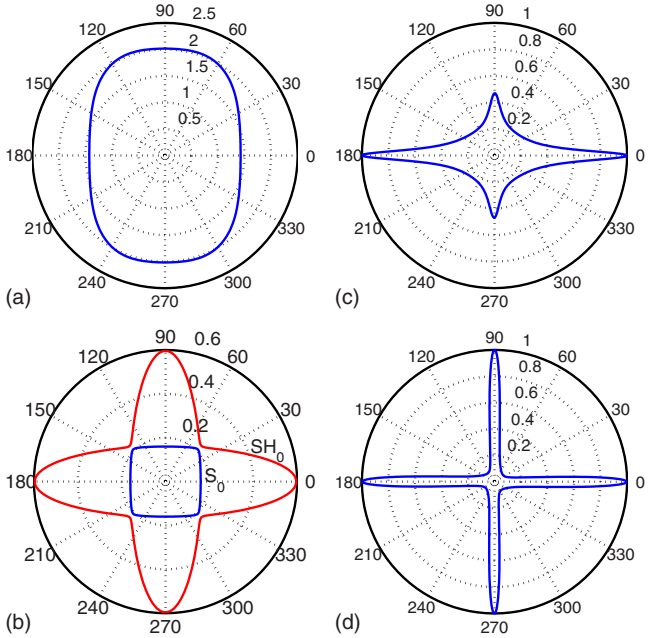


FIG. 3. (Color online) Slowness curves in $s\cdot km^{-1}$ in a cross-ply $[0°/90°]_S$ CFRP (T700GC/M21) for (a) A_0 mode at 40 kHz and (b) S_0 and SH_0 modes at 300 kHz. Corresponding focusing factor $\sqrt{A}$ for (c) A_0 mode at 40 kHz and (d) S_0 mode at 300 kHz.

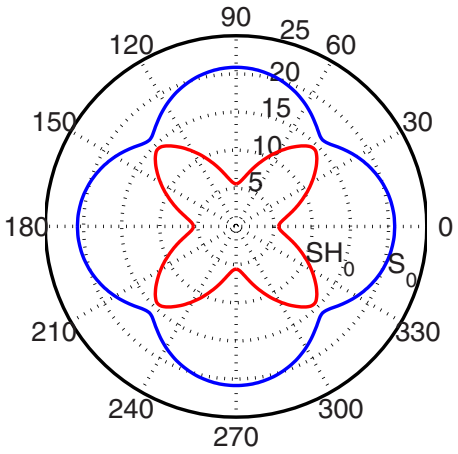


FIG. 4. (Color online) Wavelength of S_0 and SH_0 modes in a cross-ply $[0°/90°]_S$ CFRP (T700GC/M21) at 300 kHz versus propagation direction θ.

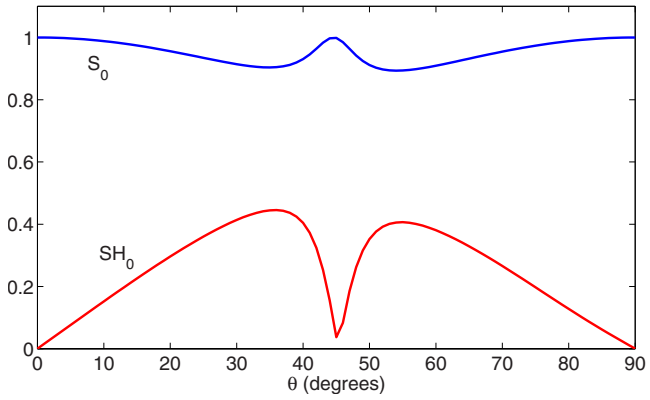


FIG. 5. (Color online) Radial displacement at the surface of the plate for S_0 and SH_0 modes at 300 kHz, normalization by the total in-plane displacement.

of the Maris factor exists in the direction $\varphi=90°$. The amplitude ratio between the direction $\varphi=90°$ and 0° is 47%; the minimum of energy occurs in the direction $\varphi=50°$, where the ratio is only 23.4% compared to the direction $\varphi=0°$. That difference in the focusing between the directions $\varphi=0°$ and 90° for the A_0 mode can be explained by the fact that at this low frequency-thickness product, the A_0 mode is almost a flexural wave. As a consequence, the lay-out sequence strongly influences the flexural behavior of the composite plate, as it is accounted for from plate theory.

For S_0 mode at 300 kHz, focusing occurs at $\varphi=0°$ and 90°, and the minimum is in the direction $\varphi=45°$. The focusing is very acute in fiber direction since the ratio between the minimum and the maximum of the focusing factor $\sqrt{A}$ is 8%. S_0 mode focusing is thus much more important than A_0 mode focusing.

The SH_0 mode slowness curve is not always convex. As a consequence, this mode exhibits caustics at $\varphi=7°$ and $\varphi=83°$. In these directions corresponding to inflexion points on the slowness curve, the focusing is very sharp.[11] However, our model cannot describe this phenomenon since in such directions the second term in the expansion of the phase $\Phi_M(\theta)$ is also equal to zero. The Taylor expansion in Eq. (12) is no more valid.

IV. EXPERIMENTS

In order to illustrate the phenomenon of Lamb wave focusing, some experiments have been performed. The tested plate is a $[0°/90°]_S$-T700GC/M21 cross-ply composite of thickness $d=1$ mm. A thin PZT disk of diameter $D=10$ mm is bonded at the center of the plate and is used to generate Lamb waves. As its acoustical impedance is much larger than that of the composite material, the PZT disk imposes an axisymmetric force at the surface of the plate along its periphery.[17,18] Lamb waves of wavelength $\lambda=1.7D$, i.e., 17 mm, having a large radial displacement are preferentially excited.[19] Green's function represents the response of the plate to a point-like source. In practice, the lateral dimensions of the source must be smaller that the wavelength. With $D\cong\lambda/2$, this condition is nearly fulfilled, and due to the symmetry, the finite dimension of the PZT disk introduces only a frequency filtering. Taking into account the dispersion curves, the wave number–thickness product $kd=0.37$ corresponds to frequencies equal to 40 and 300 kHz for A_0 and S_0 modes, respectively. Figure 4 shows the wavelength of S_0 and SH_0 modes versus the propagation direction at 300 kHz. Conversely to S_0 mode, the SH_0 mode wavelength varies rapidly and approaches the preferentially generated value (17 mm), only between 30° and 60°. Figure 5 shows that the normalized radial displacement at the surface of the plate of SH_0 mode is also very sensitive to the propagation direction, whereas for S_0 mode the radial displacement is almost constant for all angles. In the range of directions [30°,60°], the radial displacement of SH_0 mode is not negligible. Therefore, this mode can be generated by the axisymmetric source for θ between 30° and 40° and between 50° and 60°, approximately. The detection techniques, described later, are both sensitive to the out-of-plane displacement. Figure 6 shows that the normalized out-of-plane displacement of the SH_0 mode, always smaller than that of S_0 mode, is significant in these two ranges of directions. From these considerations, it results that S_0 and SH_0 modes are coupled and that it is not possible to separate them. However, the combination of the wavelength mismatch and of reduced radial and out-of-plane components makes the generation and detection

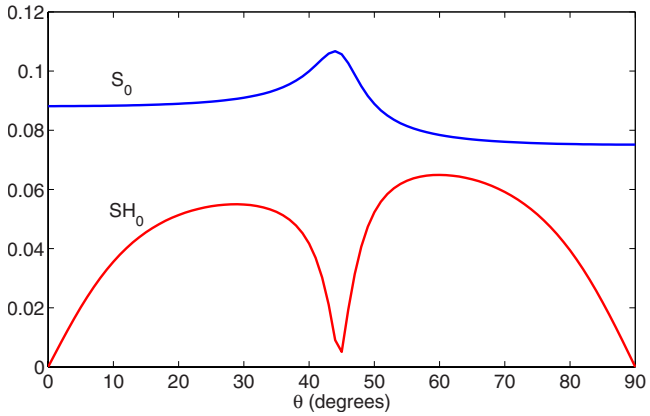


FIG. 6. (Color online) Out-of-plane displacement at the surface of the plate for S_0 and SH_0 modes at 300 kHz, normalization by the total in-plane displacement.

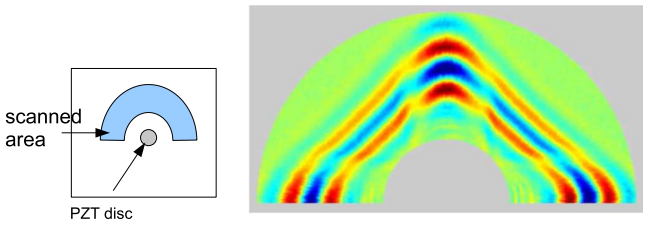


FIG. 7. (Color online) Cartography at $t=228$ μs of the S_0 mode propagating in a cross-ply CFRP plate obtained by scanning the plate with an air-coupled transducer. Lamb waves are generated at 300 kHz with a thin PZT disk bonded at the center of the plate.

much less efficient for SH_0 mode than for S_0 mode. Therefore, due to the shape of the slowness curves presented in Fig. 3(b), the contribution of the shear mode will be significant only in the direction of the two caustics, for $\theta=35°$ and $\theta=55°$, i.e., for observation directions $\varphi=7°$ and $\varphi=83°$.

The PZT disk is excited with a three cycle tone burst of amplitude 10 V_{pp}, with a repetition rate of 10 ms; the detected signal is averaged 200 times at each measurement point. The effect of the anisotropy on the propagation of these guided waves can be observed with an air-coupled transducer mounted in an XY-translational stage, which permits us to observe the wave front during the propagation in the plate. Figure 7 shows a cartography at time $t=228$ μs of the S_0 mode amplitude at 300 kHz. Due to the anisotropy of the material, the S_0 mode travels faster in the directions $\varphi=0°$ and 90°. The focusing in these directions is revealed by the increase in the amplitude of the measured signal.

However, in order to quantify precisely this focusing effect, another experimental setup has been used. The normal displacement at the surface of the plate is measured for the different angles of propagation with an optical heterodyne interferometer at $r=10$ cm from the center of the PZT disk. As $kr\approx 40$, the far field assumption is therefore respected. In order to increase the reflection of the laser on the black surface of the composite plate, a thin reflective film has been bonded on the surface of the plate at the measurement points. The film is sufficiently thin so that wave propagation is not disturbed.

Results for S_0 and SH_0 modes at 300 kHz are presented in Fig. 8. As the measured quantity is the displacement and not the energy, the relevant theoretical curve is the focusing factor $\sqrt{A}$ and not the Maris factor $A=\mathrm{d}\theta/\mathrm{d}\varphi$ [see Eq. (13)]. The strong focusing in the direction 0° for S_0 mode is clearly observed and fits well with the prediction. Amplitude fluctuations can be ascribed to the surface roughness of the plate. The possible contribution of SH_0 mode at $\varphi=\pm 7°$ is blurred in the central peak at 0° and, consequently, cannot be highlighted.

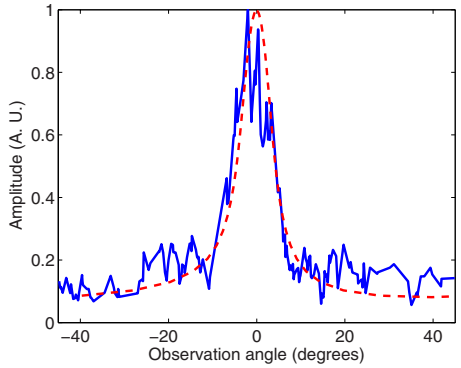


FIG. 8. (Color online) S_0 Lamb mode at 300 kHz. Normalized displacement measured at the surface of the cross-ply composite plate compared with the theoretical focusing factor $\sqrt{A}$ (dashed line).

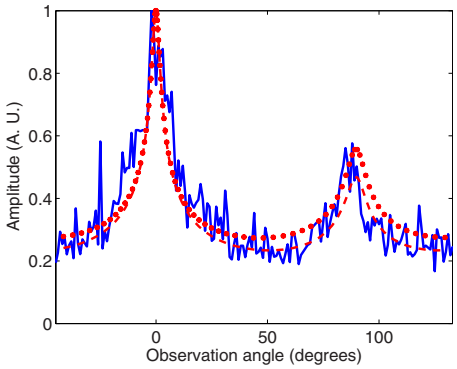


FIG. 9. (Color online) A_0 Lamb mode at 40 kHz. Normalized displacement measured at the surface of the cross-ply composite plate compared with theoretical focusing factors $\sqrt{A}$ (dashed line) and $\sqrt{Ak}\cos\psi$ (dots).

For the A_0 mode at 40 kHz, Fig. 9 clearly shows the focusing that occurs mainly in the direction $\varphi=0°$ and with less intensity in the 90° direction. Variations in factors $\sqrt{A}$ and $\sqrt{Ak}\cos\psi$ versus the observation angle φ indicate the prominent role of the Maris factor in the focusing effect. The agreement between theoretical and experimental curves is good.

Slowness curves of Lamb modes are frequency dependent. Then, an analysis in a large frequency range must be performed for a complete understanding of the focusing effect. However, these preliminary results indicate that the A_0 mode at a relatively low frequency ($kd=0.37$) is a better candidate than S_0 mode for the inspection of a CFRP plate. Moreover, the lay-out of transmitters and receivers should avoid any acoustic path at angles around 45° from the fibers.

V. CONCLUSION

An approximation of the Green's function giving a general description of the generation of Lamb waves in a thin anisotropic plate was established. The investigation presented here has pointed out that Maris factor provides a good approximation of the energy radiated in the far field by an axisymmetric source. Numerical simulations and experiments show that the phenomenon of focusing of Lamb waves can be very strong in some materials such as CFRPs. That is why this phenomenon must be correctly taken into account, for example, in order to develop systems for *in situ* SHM. The choice of the most appropriate Lamb mode, of the excitation frequency, and of the design of the array of PZT disks used as transmitters and receivers depends on such considerations. The next step for the present investigation could consist in an experimental observation of configurations where the slowness curve generates caustics, inducing a stronger focusing of Lamb waves.

ACKNOWLEDGMENTS

The authors wish to thank Dr. Dominique Clorennec (Laboratoire Ondes et Acoustique, Paris) for his help in conducting laser ultrasonics experiments and for useful discussions of the results and Philippe Nuñez (ONERA) for the manufacturing of the composite plate.

[1]D. E. Chimenti, "Guided waves in plates and their use in materials characterization," Appl. Mech. Rev. **50**, 247–284 (1997).
[2]M. Lowe, "Characteristics of the reflection of Lamb waves from defects in plates and pipes," Rev. Prog. Quant. Nondestr. Eval. **17**, 113–120 (1998).
[3]P. Cawley, M. Lowe, D. Alleyne, B. Pavlakovic, and P. Wilcox, "Practical long range guided wave testing: Applications to pipes and rail," Mater. Eval. **61**, 66–74 (2003).
[4]M. Castaings, B. Hosten, and T. Kundu, "Inversion of ultrasonic, plane-wave transmission data in composite plates to infer viscoelastic material properties," NDT & E Int. **33**, 377–392 (2000).
[5]A. H. Nayfeh, *Wave Propagation in Layered Anisotropic Media with Application to Composites* (Elsevier, Amsterdam, 1995).
[6]E. Moulin, S. Grondel, J. Assaad, and L. Duquenne, "Modeling a surface-mounted Lamb wave emission-reception system: Applications to structural health monitoring," J. Acoust. Soc. Am. **124**, 3521–3527 (2008).
[7]A. G. Every, W. Sachse, K. Y. Kim, and M. O. Thompson, "Phonon focusing and mode-conversion effects in silicon at ultrasonic frequencies," Phys. Rev. Lett. **65**, 1446–1449 (1990).
[8]D. Royer, "Mixed matrix formulation for the analysis of laser-generated acoustic waves by a thermoelastic line source," Ultrasonics **39**, 345–354 (2001).
[9]A. A. Maznev, A. M. Lomonosov, P. Hess, and Al. A. Kolomenskii, "Anisotropic effects in surface acoustic wave propagation from a point source in a crystal," Eur. Phys. J. B **35**, 429–439 (2003).
[10]H. J. Maris, "Enhancement of heat pulses in crystals due to elastic anisotropy," J. Acoust. Soc. Am. **50**, 812–818 (1971).
[11]A. A. Maznev and A. G. Every, "Focusing of acoustic modes in thin anisotropic plates," Acta Acust. **3**, 387–391 (1995).
[12]H.-Y. Chen and C.-S. Man, "Radiation and focusing of SH_0 elastic waves in anisotropic plates," Appl. Phys. Lett. **64**, 966–968 (1994).
[13]A. Velichko and P. D. Wilcox, "Modeling the excitation of guided waves in generally anisotropic multilayered media," J. Acoust. Soc. Am. **121**, 60–69 (2007).
[14]S. Banerjee, W. Prosser, and A. Mal, "Calculation of the response of a composite plate to localized dynamic surface loads using a new wave number integral method," J. Appl. Mech. **72**, 18–24 (2005).
[15]D. Royer and E. Dieulesaint, *Elastic Waves in Solids, I* (Springer, Berlin, 1999).
[16]G. R. Liu, J. Tani, T. Ohyoshi, and K. Watanabe, "Transient waves in anisotropic laminated plates, Part 1: Theory," J. Vibr. Acoust. **113**, 230–234 (1991).
[17]V. Giurgiutiu, "Tuned Lamb wave excitation and detection with piezoelectric wafer active sensors for structural health monitoring," J. Intell. Mater. Syst. Struct. **16**, 291–305 (2005).
[18]F. Lanza di Scalea, H. Matt, and I. Bartoli, "The response of rectangular piezoelectric sensors to Rayleigh and Lamb ultrasonic waves," J. Acoust. Soc. Am. **121**, 175–187 (2007).
[19]L. R. F. Rose and C. H. Wang, "Mindlin plate theory for damage detection: Source solutions," J. Acoust. Soc. Am. **116**, 154–171 (2004).

Extracting guided waves from cross-correlations of elastic diffuse fields: Applications to remote structural health monitoring

Adelaide Duroux and Karim G. Sabra[a)]
School of Mechanical Engineering, Georgia Institute of Technology, Atlanta, Georgia 30332-0405

James Ayers and Massimo Ruzzene
School of Aerospace Engineering, Georgia Institute of Technology, Atlanta, Georgia 30332-0150

(Received 20 March 2009; revised 24 September 2009; accepted 9 October 2009)

Due to their apparent complexity, diffuse fields are not commonly used by conventional structural health monitoring methodologies. However, recent theoretical and experimental studies have demonstrated that local Green's functions (GFs) can be estimated from the cross-correlation of elastic diffuse fields (CDFs) recorded between any points of a sensor grid and generated by sources located remotely from the monitoring area. Hence, this CDF technique effectively transforming each measurement point into a virtual elastic source, thus allowing to measure the local structural response (or GF). Theoretical and experimental studies are conducted to assess the broadband performance of the CDF technique for the extraction of low-frequency coherent Lamb waves (4–90 kHz) propagating in a thin aluminum plate of complex geometry using a scanning laser Doppler vibrometer. The coherent signal-to-noise ratio and the accuracy of the phase and group velocity measurements of the coherent A_0 mode obtained from CDF are also measured to quantify the influence of the spatio-temporal characteristics of the recorded diffuse field, the number and spatial configuration of the remote ultrasonic sources, and the influence of incoherent measurement noise (e.g., sensor noise). In addition, a criterion to assess the emergence rate of late coherent coda arrivals from CDF is also formulated.
 [DOI: 10.1121/1.3257602]

PACS number(s): 43.40.Le, 43.40.Sk, 43.40.Hb, 43.35.Zc [RLW] Pages: 204–215

I. INTRODUCTION

Structural health monitoring (SHM) often relies on propagating elastic guided waves through a structure using embedded sensors in order to assess its structural integrity and detect eventual degradations (Giurgitiu, 2007). Most inspected structures are geometrically complex (e.g., due to the presence of rivets holes and stiffeners) which causes scattering, multiple reflections, and mode conversion of the guided waves used for testing. Thus, the superposition and complex interaction of multiply scattered guided waves can rapidly lead to complicated waveforms away from the original elastic source. Indeed, the geometric complexity of the inspected structures enhances the randomization of the elastic energy within these structures and thus actually favors the formation of diffuse fields over long reverberation time. Overall, a structure capable of sustaining a diffuse field must be lightly damped, allowing many reflections of the input wave energy. Fully diffuse wavefields are often defined as ones that are globally equi-partitioned, with all normal modes having uncorrelated amplitudes with equal mean squares (Weaver, 1982, 1984; Evans and Cawley, 1999; Weaver and Lobkis, 2004).

Diffuse fields in structures have an apparent random nature and are thus generally discarded in conventional SHM systems. However, recent theoretical and experimental studies have demonstrated that the local Green's functions (GFs) (i.e., local elastic response) can be estimated from the cross-correlation of elastics diffuse fields (CDF) recorded between a pair of (embedded) sensors (Weaver and Lobkis, 2001; Larose *et al.*, 2007; Sabra *et al.*, 2008). One potential method to generate a diffuse field is to probe the structure at random locations using one or several controlled sources (called "secondary sources" thereafter) which can be remotely located from the pair of sensors used for local monitoring (see Fig. 1). Thus overall, CDF provides the means to perform SHM without a local source, since the extracted coherent guided waves between the sensors are similar to those obtained from conventional pitch-catch measurements between a source and receiver pair (i.e., the GF). For instance, for practical applications, ultrasonic sources may be located remotely from the sensing region (e.g., in easily accessible areas during routine maintenance operations) while the sensors are embedded in hard-to-reach areas (e.g., in the vicinity of a hidden structural "hot-spots") where structural failure (e.g., fatigue cracks) is likely to occur. Hence CDF offers an inherently safe sensing technique for monitoring structural hot-spots in hazardous regions (e.g., fuel transfer holes in the wing risers), since it does not require a local active source but only sensors.

Indeed several experimental studies have demonstrated the feasibility of extracting estimates of the GF (also referred to as impulse response or structural response) from cross-

[a)]Author to whom correspondence should be addressed. Electronic mail: karim.sabra@me.gatech.edu

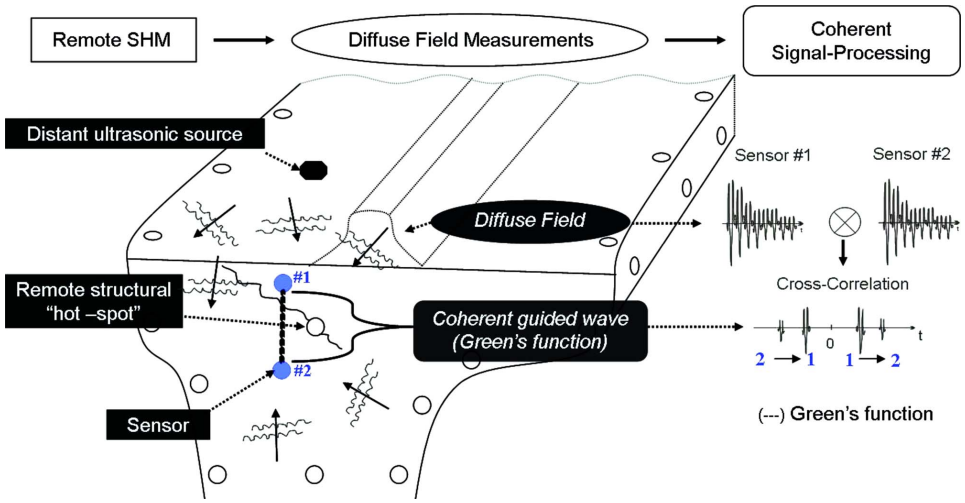


FIG. 1. (Color online) Principle of the CDF technique.

correlations of (acoustic or elastic) diffuse fields or ambient noise records in applications relevant to seismology (Shapiro *et al.*, 2005; Sabra *et al.*, 2005c), underwater acoustics (Roux *et al.*, 2005; Sabra *et al.*, 2005a, 2005b), civil engineering (Farrar and James, 1997; Snieder and Cafak, 2006), low-frequency (<5 kHz) flexural properties identification of hydrofoils (Sabra *et al.*, 2007), high frequency ultrasonics (approximately megahertz) (Weaver and Lobkis, 2001; Larose *et al.*, 2006; Van Wijk, 2006; Langley, 2007), and guided wave (or Lamb wave) measurements (approximately kilohertz) (Larose *et al.*, 2007; Sabra *et al.*, 2008).

A practical issue in the implementation of the CDF technique resides in the selection of the number N of secondary ultrasonic sources and their locations, in order to generate a fully diffuse field in the structure after superposition of their individual contributions at the receiver. Furthermore, in complex structures, the measured signals are typically nonstationary and feature an exponentially decaying coda which is dominated by multiply scattered waves (see schematic on Fig. 1). On one hand, it has been shown that the diffuse field regime is more likely to be achieved within the late coda waves (after multiple reflections have occurred) than during the early ballistic direct arrivals (Weaver 1986; Larose *et al.*, 2004; Paul *et al.*, 2005). On the other hand, coda waves have an overall exponential decay, mainly determined by elastic attenuation, which can quickly be contaminated by measurement noise or electronic noise. Thus selecting the optimal portion of duration T of the recorded signals for implementing CDF may not be straightforward and indeed depends on experimental conditions.

The main objective of this article is to investigate experimentally, using a scanning laser vibrometer, the emergence of coherent Lamb waves from the CDF technique and thus complement previous theoretical studies (Sabra *et al.*, 2005b; Larose *et al.*, 2008). The ratio of coherent arrivals to residual temporal fluctuations as well as the accuracy of the phase and group velocities of these coherent guided wave arrivals were measured to quantify (1) the performance of various processing schemes of the recorded diffuse fields, (2) the effect of the spatial configuration of the secondary ultrasonic sources, and (3) the influence of incoherent measurement noise (e.g., sensor noise) on the performance of the CDF technique. Additionally a criterion was developed to determine whether late arrivals of the cross-correlation waveforms obtained from CDF are actually dominated by weak coherent coda arrivals (thus potentially usefully for SHM purposes) or by residual fluctuations instead (thus indicating that more spatial or temporal averaging is required).

Following this introductory section, Sec. II summarizes theoretical derivations related to performance of the CDF technique in the presence of incoherent measurement noise. Section III describes the experimental setup and diffuse field measurements in a plate with complex geometry. Section IV presents the experimental parametric study of the CDF technique. Finally, the conclusions drawn from this study are summarized in Sec. V.

II. THEORY

A. Definition of mean and variance of the CDF estimate

As stated in the Introduction, the CDF technique is typically implemented by using one or several remote sources distributed throughout the structure of interest away from the sensors' region to generate a diffuse field in the sensors' region (see Fig. 1). The cross-correlation $C_{12}^j(t)$ between the reverberating signals $S_1^j(t)$ and $S_2^j(t)$ generated by the jth remote source and recorded, respectively, by two sensors 1 and 2 is given by

$$C_{12}^j(t) = \frac{1}{T}\int_0^T S_1^j(\tau) S_2^j(t+\tau) d\tau, \tag{1}$$

where T is the duration of the reverberating recordings. The temporal integration in Eq. (1) physically corresponds to a temporal averaging operation over the finite duration T. The expected value of the cross-correlation between sensors 1 and 2, noted $C_{12}(t)$, can be constructed from an ensemble average of the cross-correlations $C_{12}^j(t)$ obtained from each secondary source j:

$$C_{12}(t) = \frac{1}{N}\sum_{j=1}^{N} C_{12}^j(t). \tag{2}$$

The ensemble average over all remote sources helps establishing a diffuse regime at the receiver's locations by cumulating multiple realizations of the scattered field in the structure of interest. Simply put, the more complex and reverberant the structure, the fewer are the sources needed to generate a diffuse field throughout the structure in practice.

It should be noted that computing Eqs. (1) and (2) involves cross-correlating signals originating from the same source event at both sensors (i.e. a coherent processing). Thus only the square of the amplitude spectrum of the signal excitation matters and not its absolute phase (which cancels out in the cross-correlation operation). Hence, various source excitations could indeed be used in practice (e.g., pulse, frequency sweeps, and coded sequences) depending on the types of sensors and the desired improvement in the signal-to-noise ratio of the recordings $S_1^j(t)$ and $S_2^j(t)$.

In practice, the mean cross-correlation $C_{12}(t)$ [see Eq. (2)] is only an estimate of the actual local Green's function $G_{12}(t)$ between sensors 1 and 2. Consequently, $C_{12}(t)$ always contains some residual temporal fluctuations which can blur the identification of the symmetric arrivals of $C_{12}(t)$ (e.g., for insufficient number N of secondary source). High levels of residual fluctuations may potentially prevent the accurate extraction of the local Green's function $G_{12}(t)$ from CDF. The level of these temporal fluctuations can be estimated from the square root of the variance $V_{12}(t)$ of the CDF estimate $C_{12}(t)$ which is defined as (Weaver and Lobkis, 2005; Sabra *et al.*, 2005b; Larose *et al.*, 2008)

$$V_{12}(t) = \frac{1}{N-1}\left(\frac{1}{N}\sum_{j-1}^{N}(C_{12}^j(t))^2 - \left(\frac{1}{N}\sum_{j=1}^{N}C_{12}^j(t)\right)^2\right). \quad (3)$$

One contribution to the variance of the cross-correlation waveforms is the imperfect averaging over the N secondary sources or time duration of the measurements T when computing the ensemble-averaged cross-correlation $C_{12}(t)$ as compared to the ideal situation where fully diffuse wavefield actually exist in the tested structure. One metric commonly used to asses the performance of CDF is the "signal-to-noise" ratio (SNR) of the emerging coherent arrivals from the cross-correlation waveform $C_{12}(t)$ defined as (Larose *et al.*, 2004; Weaver and Lobkis, 2005; Sabra *et al.*, 2005b; Larose *et al.*, 2008)

$$\mathrm{SNR}(t) = \frac{C_{12}(t)}{\sqrt{V_{12}(t)}}. \quad (4)$$

Here the "signal" level of interest is the value of the cross-correlation waveform $C_{12}(t)$ (or its envelope), and the corresponding "noise" level is the standard deviation of the residual temporal fluctuations of the cross-correlation [i.e., the square root of the variance, see Eq. (3)]. The value of SNR(t) is a common metric used to assess the performance of CDF for estimating the local Green's function $G_{12}(t)$ between sensors 1 and 2.

B. Theoretical prediction of the coherent SNR

Theoretical expressions for the mean $C_{12}(t)$ and variance $V_{12}(t)$ of the CDF estimate have been previously derived for various recorded signal models (e.g., stationary or nonstationary elastic field) (Weaver and Lobkis, 2005; Sabra *et al.*, 2005b; Larose *et al.*, 2008). To the first order, these theoretical expressions reduce to (Larose *et al.*, 2008)

$$C_{12}(t) \approx F(t) \otimes (G_{12}(t) - G_{21}(-t)) \cdot \int_0^T I(u)I(t+u)du, \quad (5)$$

where $G_{12}(t)$ is the actual Green's function between sensors 1 and 2, $I(t)$ is the envelope of the recorded signal averaged over the N sources (i.e., the square root of the averaged intensity), T is the recording duration, $F(t)$ is a function taking into account the combined effect of transducer's response (in phase and amplitude) and the normalized autocorrelation of the source signal (close to a Dirac function for impulsive sources), and the symbol $\otimes$ denotes a convolution operation. When using identical transducers at locations 1 and 2 and impulsive remote sources (e.g., laser impulse) to generate the elastic wavefields, $F(t)$ simply acts as band-pass filter on the true Green's function $G_{12}(t)$. Equation (5) states that for a fully diffuse field, the cross-correlation is a symmetric function of time. Hence, in practical applications, symmetric arrivals of $C_{12}(t)$ can be used as reliable estimates of actual paths of the local Green's function $G_{12}(t)$.

Similar to Eq. (5), the theoretical expression of the variance $V_{12}(t)$ can be simplified to

$$V_{12}(t) \approx \frac{1}{2BN}\int_0^T I^2(u)I^2(t+u)du, \quad (6)$$

where N is the number of remote sources, T is the recording duration, and B is the frequency bandwidth of the recorded signals. Equation (6) was derived under the main assumptions of large lag time and distant receivers compare to the wavelength of interest as discussed in the previous literature (Weaver and Lobkis, 2005; Sabra *et al.*, 2005b; Larose *et al.*, 2008). Furthermore, as discussed by Larose *et al.* (2008), the simple model of the diffuse signals yielding Eqs. (5) and (6) is not valid if long-range correlations occur in the multiply scattered signals, for instance, due to the existence of closed loops or recurrent scattering, which may exist in a highly disorganized propagating medium.

In complex structures, the measured signals at the sensor locations generated by an impulsive source are typically nonstationary signals with an exponentially decaying coda (see Fig. 1). This coda is dominated by multiply scattered waves over long reverberation time. In this case, the temporal evolution of the averaged intensity of the decaying elastic wavefield, noted $I_c(t)$ (i.e., the signal's envelope), can be fitted by (see Fig. 5)

$$I_c(t) = I_0 e^{-t/\tau_D}, \quad (7)$$

where I_0 is the maximum value of the averaged intensity (at $t=0$) and τ_D is the decay time of the elastic field whose value depends on the scattering properties and elastic attenuation of the structure (Evans and Cawley, 1999; Derode *et al.*, 1999; Larose *et al.*, 2008).

Furthermore, in practice, the recorded signals always contain some amount of non-propagating noise (e.g., experi-

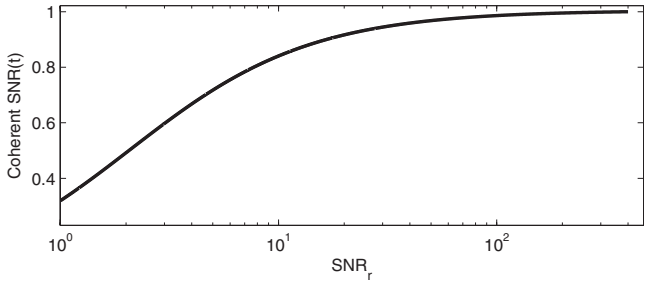


FIG. 2. Normalized variations of the coherent signal-to-noise ratio (SNR(t)) vs signal-to-noise ratio of the recordings (SNR$_r$) based on theoretical predictions of Eq. (10) for the parameters: T=6 ms, τ_D=5 ms, t^*=148 μs (A_0 arrival time).

mental sensor noise or electrical noise) which hinders the accurate recording of weak multiply scattered signals at longer times. The simplest noise model for this incoherent noise is the result of contributions from an additive stationary homogeneous flat-spectrum band-limited zero-mean random field that is uncorrelated between sensors' locations and uncorrelated with the diffuse elastic wavefield. If σ_N^2 is the variance of the idealized uncorrelated noise field (Proakis, 1995) (assumed identical for all measurements), the total recorded intensity $I^2(t)$ can then be approximated by $I_c^2(t) + \sigma_N^2$.

$$I^2(t) \approx I_c^2(t) + \sigma_N^2. \tag{8}$$

On the other hand, since the incoherent noise-field is assumed to be spatially uncorrelated and with zero mean, it does not contribute, to the first order, to the mean coherent estimate $C_{12}(t)$. Hence in the numerator of $C_{12}(t)$, the envelope $I(t)$ (associated with the amount of coherent energy flowing between sensors 1 and 2) can then simply be approximated by $I_c(t)$. Consequently, after combining Eqs. (5)–(8), the theoretical prediction of the signal-to-noise ratio SNR(t) [see Eq. (4)] of the CDF estimate $C_{12}(t)$ can be expressed as

$$\mathrm{SNR}(t) = \sqrt{2BN}\frac{\int_0^T I_c(u)I_c(t+u)du}{\sqrt{\int_0^T (\sigma_N^2 + I_c^2(u))(\sigma_N^2 + I_c^2(t+u))du}} \cdot (F(t) \otimes (G_{12}(t) - G_{21}(-t))), \tag{9}$$

which can be rewritten explicitly using the modeled averaged intensity $I_c(t)$ from Eq. (7) as

$$\mathrm{SNR}(t) = \sqrt{2B\tau_D N}\frac{\mathrm{SNR}_r(e^{-t/\tau_D} - e^{-(2T+t)/\tau_D})}{\sqrt{(\mathrm{SNR}_r)^2(e^{-2t/\tau_D} - e^{-2(2T/t)/\tau_D}) + 2\mathrm{SNR}_r(1+e^{-2t/\tau_D})(1-e^{-2T/\tau_D}) + \frac{4T}{\tau_D}}} \cdot (F(t) \otimes (G_{12}(t) - G_{21}(-t))), \tag{10}$$

where $\mathrm{SNR}_r = I_o^2/\sigma_N^2$ is defined as the signal-to-noise ratio of the recorded diffuse waveforms.

Equations (9) and (10) can be used to estimate the influence of various parameters such as N (number of sources), B (signal frequency bandwidth), T (recording duration), τ_D (the decay time of the diffuse field signals), and SNR$_r$ on the coherent signal-to-noise ratio SNR(t) of the cross-correlation waveform. Based on Eqs. (9) and (10), it appears that the parameters T and N do not have the same influence on the increase in the coherent SNR in the case of noisy recorded waveforms. Indeed, since the incoherent noise component is assumed to be spatially uncorrelated [Eq. (8)] among receivers 1 and 2, it can be noticed that SNR(t) theoretically grows as $\sqrt{N}$. On the other hand, Eqs. (9) and (10) show that low values of SNR$_r$ for the recorded signals limit the achievable value of the coherent SNR, especially if the recording duration T is selected to be longer than τ_D. Figure 2 illustrates the result of the coherent SNR in function of SNR$_r$ using parameters from the experiment described in Sec. III: T=6 ms, τ_D=5 ms, and t^*=148 μs is the arrival time of the A_0 first anti-symmetric Lamb mode. Furthermore, in the case of high values of the recorded SNR$_r$ (i.e., weak incoherent noise level), it can be noted that the expression of the coherent SNR in Eq. (10) is, as expected, independent of SNR$_r$ and the classical dependency of $\sqrt{B\tau_D N}$ (for exponentially decaying diffuse signals, that is, $T \sim \tau_D$) or $\sqrt{BTN}$ (for stationary signals—such as continuous ambient noise—that is, $\tau_D \to \infty$) is recovered, in agreement with previous studies (Weaver and Lobkis, 2005; Sabra *et al.*, 2005b; Larose *et al.*, 2008).

Finally, for practical SHM applications, the recorded signals can actually have a very short coda due to high level of elastic attenuation, (i.e., small decay time τ_D). Consequently, previous experimental studies have proposed to increase the coherent SNR by clipping the amplitude of the recorded signal to an arbitrary threshold in order to artificially enhance the contribution of the weak arrivals in the late coda [Larose *et al.*, 2004; Sabra *et al.*, 2008]. This clipping procedure allows maximizing the amount of useful information in the recorded signals which contributes to the cross-correlations $C_{12}(t)$. However, the amplitude clipping also artificially enhances the relative importance of the incoherent noise component and thus effectively lowers the recorded SNR$_r$, especially in the late coda. Consequently, the clipping level should be, as expected, significantly larger than the level of incoherent noise (σ_N^2) in order for the amplitude clipping procedure to actually enhance the coherent SNR as shown from experimental results in Sec. IV.

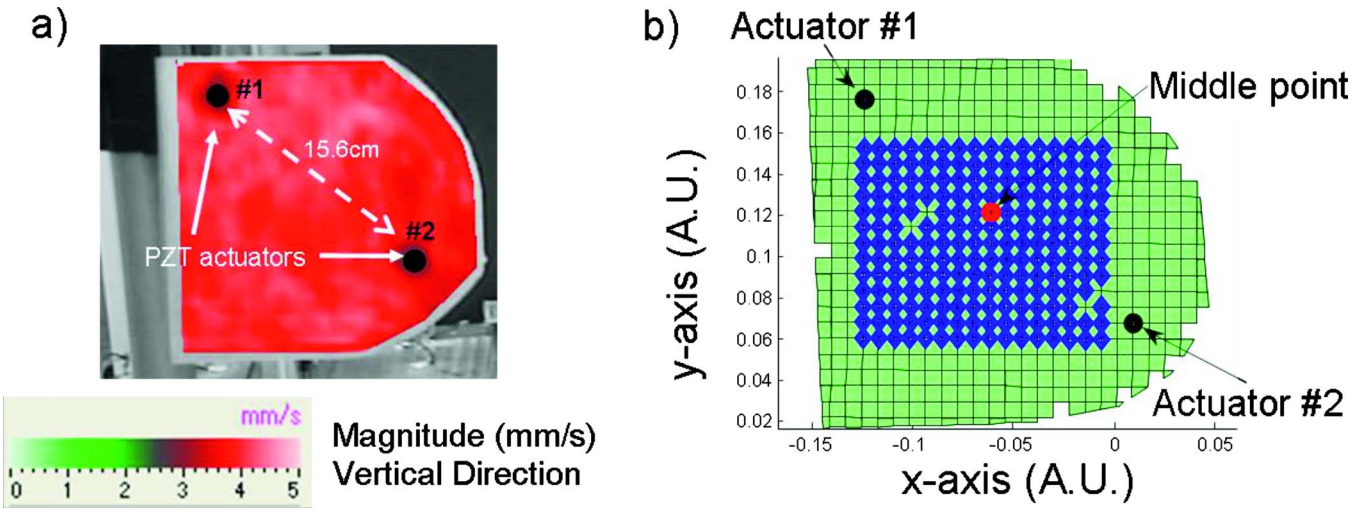


FIG. 3. (Color online) (a) 2-mm-thick aluminum plate test sample with random geometry with details of source locations. The color scale indicates the measured small variations in the average rms velocity evaluated over the plate with the SLDV. (b) Measurement grid of the SLDV (610 points). The square pixel size is 0.6 cm. The locations of the two actuators as well as the grid point equidistant from the sensors (middle point) are also indicated. The diamond shapes indicate the sensing points located in the central area, used for the diffuse field analysis, of the scanning grid.

III. DIFFUSE FIELD MEASUREMENTS

A. Experimental setup

Experiments were conducted to estimate the decay time (or reverberation time) τ_D [see Eq. (7)] on an aluminum plate of thickness 2 mm with complex geometry [Fig. 3(a)]. Cutting the plate edges irregularly enhances the generation of diffuse field through multiple scattering and reflections from the plate's edges and other discontinuities (Weaver, 1986; Evans and Cawley, 1999). Two piezoelectric (PZT) ceramic actuators disks (Steminc, model SMD15T09F2275S) were mounted 15.6 cm apart on the plate at two arbitrary locations. Each actuator was driven by a waveform generator (Agilent model 33220A) transmitting 1 cycle of a sinusoidal signal of period 20 μs, of frequency 50 kHz, and of peak amplitude 1 V. A scanning laser Doppler velocimeter (SLDV) [Polytec PI, Model PSV400M2] was used to record the first 12.8 ms of the the plate's velocity response (2.5 MHz, sampling frequency) over a fixed grid of 610 points [see Figs. 3(b) and Fig. 4(a)]. The SLDV is mainly sensitive to the out-of-plane displacements of the plate dominated here by the A_0 guided wave modes given the thin plate geometry and the use of PZT actuators. Furthermore, due to the limited bandwidth and geometry of the PZT actuators, the energy spectrum extends in the frequency band $B=4-90$ kHz [see Fig. 4(b)], with most signal energy concentrated around 40–50 kHz. No significant frequency dispersion was observed for the A_0 mode arrival for the recorded signals in the selected frequency band $B=4-90$ kHz.

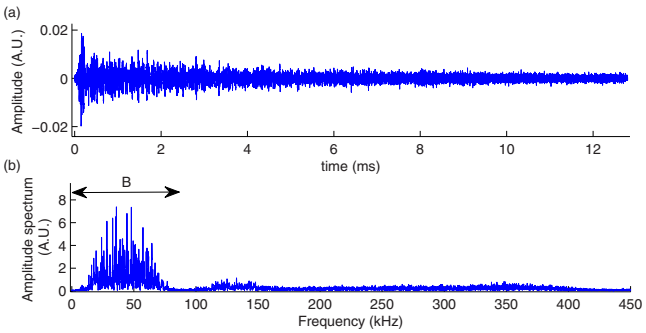


FIG. 4. (Color online) (a) Typical recorded reverberant wavefield generated by the piezoelectric actuator 1 recorded at location 2 (see Fig. 3). After the direct (or ballistic) arrival, the signal exhibits a long lasting coda (up to 10.5 ms) resulting from multiple scattering at the plate edges. (b) Amplitude spectrum of the recorded signal, with energy mostly concentrated in the frequency band 4–90 kHz.

B. Diffuse field analysis

The temporal variations in the elastic energy distribution in the plate were investigated in order to determine the overall onset of the diffuse field regime using plate's velocity measurements obtained with the SLDV. In the context of this study, energy equipartition should occur among the different guided wave modes supported by the structure which are primarily the lowest order modes S_0 and A_0 given the thin plate geometry and low frequency excitations (Akolzin and Weaver, 2004).

The decay time τ_D of the elastic energy for the recorded reverberant wavefield was determined by fitting the exponential decaying model [see Eq. (7)] to the smoothed envelopes of the measured signals [see Fig. 5]. The fitting procedure was only performed in the central portion of the signals, for $3.3\ \text{ms} < t < 10.6\ \text{ms}$ in order to exclude the ballistic and singly scattered portion of the signal. The last 2.2 ms of the recorded signals were excluded from the analysis since they appeared to be dominated by incoherent measurement noise (see Fig. 4). This fitting procedure yields very similar estimates for τ_D of 4.7 and 5 ms by using the signals generated by actuators 1 and 2, respectively. Based on the estimated group velocity of the A_0 mode in the plate (~ 1000 m/s), the decay time corresponds to an effective travel distance for the scattered waves of $D \approx 5$ m, i.e., over 30 reflections across the dimensions of the plate. These values are in quantitative agreement with previous studies concluding that diffuse field could be generated within a few tens wave transits across a regular thin plate (Evans and Cawley, 1999). Hence the

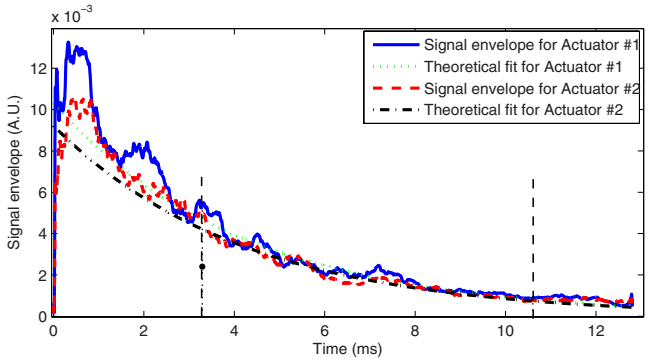


FIG. 5. (Color online) Smoothed envelopes of the reverberant signals filtered in B and recorded at the middle point between the two actuators [see Fig. 3(a)] when either actuator 1 or 2 was pulsing. The two dashed vertical lines limit the time-interval (3.3–10.6 ms) for which the theoretical exponential fit [see Eq. (7)] was computed for each of the measured signal envelope.

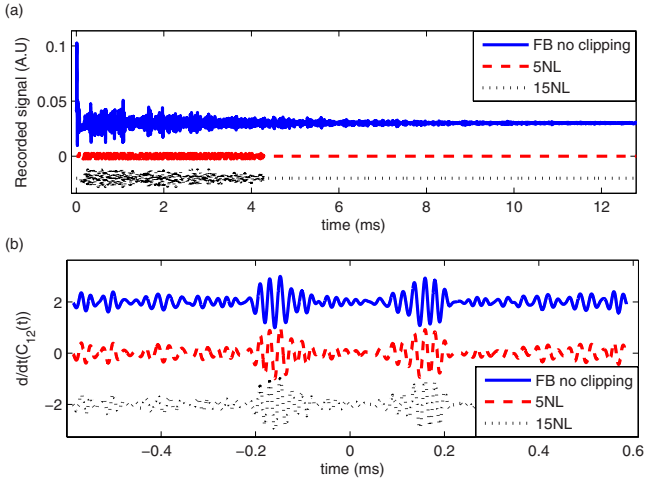


FIG. 6. (Color online) (a) Recorded signal (in the frequency band B =4–90 kHz) above actuator 1 with the SLDV when actuator 1 was active. The same signal after applying an amplitude clipping procedure, at 5 and 15 times the electronic noise level, is also displayed for the recording duration (T=4.3 ms). (b) Time-derivative of the cross-correlation waveforms obtained from CDF using the previous reverberant signals with and without amplitude clipping pre-processing measured over the whole N=610 grid points. Each cross-correlation waveform was normalized to its maximum values.

propagating wavefield is likely to exhibit characteristic features of a diffuse field regime after a few milliseconds in this thin plate with complex geometry.

IV. PARAMETRIC STUDY OF THE CDF TECHNIQUE IN A THIN PLATE

A practical issue in the implementation of CDF resides in the spatial distribution of the secondary ultrasonic sources in order to generate a fully diffuse field in the plate. To investigate these requirements, a parametric study of the CDF technique can be conducted for a given pair of sensors at location $\boldsymbol{r}_1$ and $\boldsymbol{r}_2$ by varying the spatial distribution of N uncorrelated elastic sources located at $\boldsymbol{r}_{s,j}$ $(j=1,\ldots,N)$. However, this approach would not be very practical for large values of N since it involves either using a large number of sources or repeating the experiment several times by moving a single source. Instead, by swapping the role of source and receiver based on elastic reciprocity [Achenbach, 1993], experimental parametric studies of the CDF were conducted by (1) using a SLDV to record the signals generated by two separated sources, now located at $\boldsymbol{r}_1$ and $\boldsymbol{r}_2$, over the distributed sensing locations $\boldsymbol{r}_{s,j}$ $(j=1,\ldots,N)$ (see Fig. 3) and (2) summing the cross-correlations contributions according to Eqs. (1)–(3). This reciprocal implementation of the CDF allows to simply varying the number and location of reciprocal sources.

A. Influence of the clipping level of the reverberant waveforms

Using the CDF technique [see Eqs. (1), (2), and (5)], estimates of the Green's function between the location of actuators 1 and 2 were extracted from cross-correlations of the reverberant signals recorded by the SLDV over the whole measurement grid (see Figs. 3 and 4). Clipping the amplitude of high-amplitude portion of the recorded signals can be used to assign uniform weights to the multiple reverberations (see Fig. 6) and thus artificially enhance the contribution of the late coda (tail of the signal) to the cross-correlation process (Larose *et al.*, 2004; Sabra *et al.*, 2008). The clipping threshold was determined as a multiple of the incoherent noise level NL, which was estimated from the averaged value of the standard deviation during the last 0.4 ms of the tail of all recorded signals, after filtering in the main bandwidth B =4–90 kHz. High clipping level (e.g., 70NL) barely modifies the initial reverberant waveform while a low clipping level (e.g., 5NL) removes all visual appearance of the ballistic arrival and exponential decay, as shown qualitatively on Fig. 6(a).

Figure 6(a) represents a typical recorded reverberant waveform after filtering (plain line) and the two clipped waveforms obtained after applying an amplitude clipping at both low (5NL, dashed line) and high level (15NL, dotted line) on the first 4.3 ms of the original waveform. Figure 6(b) displays the time-derivative of the resulting cross-correlation obtained after summing over the contributions of all 610 points of the scanning grid [see Eq. (2) and Fig. 3], for all three waveforms processing types shown in Fig. 6(a). The computed cross-correlation waveforms are symmetric function of time, as expected theoretically from cross-correlation of fully diffuse fields (Weaver and Lobkis, 2004; Larose *et al.*, 2008). Hence, the first high-amplitude symmetric arrival should provide a reliable estimate of the A_0 mode contribution of the local Green's function between the locations 1 and 2 (see Fig. 3). High coherent SNRs of the A_0 mode arrival (t^*=148 μs) were measured for the three waveforms for N=610 as 71 (for no clipping), 73 (for 5NL), and 75 (for 15NL) using Eqs. (2)–(4). In order to minimize the effect of local temporal fluctuations, the local variance $V_{12}(t^*)$ [see Eq. (3)] was approximated by its mean value in the interval t=70–223 μs.

Furthermore, as stated in Sec. II B, the amplitude clipping procedure is sensitive to the amount of incoherent measurement noise present in the recorded waveforms, and can potentially degrade the performance of CDF at low recorded SNR_r [see Eq. (9)]. Figure 7 displays the variations in the

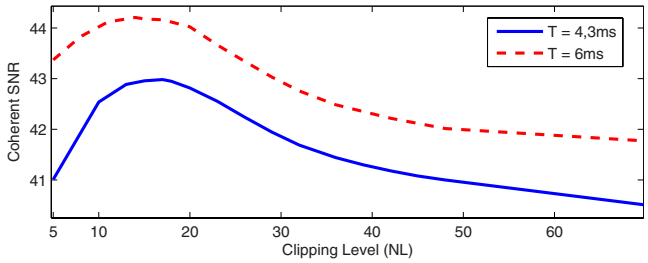


FIG. 7. (Color online) Measured evolution of the coherent SNR of the cross-correlation time function (linear scale) for increasing value of the clipping level (defined with respect to the estimated noise level, noted "NL") applied to two different epochs, T=4.3 ms (solid line) or T=6 ms (dash line), of the recorded waveforms (see Fig. 8).

coherent SNR(t) of the A_0 mode arrival (t^*=148 μs), (linear scale) for increasing clipping levels applied to the reverberant waveforms using two different recording durations T =4.3 ms (solid line) or T=6 ms (dash line) and N=200 randomly selected measurement points. The coherent SNR was computed from the symmetric part of the total correlation [i.e., $C_{12}(t)+C_{12}(-t)$]. The experimental curves displayed in Fig. 7 indicate that lower clipping improves the coherent SNR up to a maximum obtained around 15NL approximately for both selected recording of duration T. Beyond this value, any additional clipping overly emphasizes the contribution of incoherent measurement noise on the measurements which lowers the effective recorded SNR$_r$ and thus, in turn, the measured coherent SNR decreases as expected from theoretical predictions [see Eqs. (9) and (10)].

B. Influence of the number *N* and location of the secondary sources

The influence of the number N and location of secondary sources on the CDF performance is assessed experimentally in this section. First, in agreement with previous studies (Snieder, 2004; Sabra *et al.*, 2005b; Weaver and Lobkis, 2005; Larose *et al.*, 2008) and the theoretical predictions in Eqs. (9) and (10) (see Fig. 8, dash-dotted line), the coherent SNR of the A_0 arrival obtained from CDF was found to increase proportionally to $\sqrt{N}$ regardless of the specific amplitude clipping threshold apply to the reverberant waveforms (see Fig. 8, linear scale for the horizontal axis). Furthermore, results of Fig. 8 also demonstrate the improvement of coherent SNR when using a moderate clipping level (15NL, dotted line) with respect to no clipping at all (plain line), while very

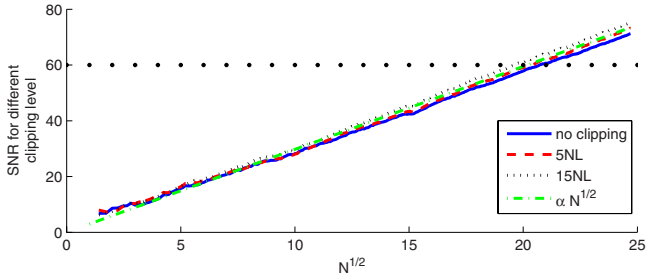


FIG. 8. (Color online) Evolution of the coherent SNR [see Eqs. (2)–(4)] vs the square root of the number of secondary sources ($\sqrt{N}$) for the same clipping level used in Fig. 8. Theoretical dependency of $\sqrt{N}$ (dotted-dashed line) is also indicated [see Eqs. (9) and (10)].

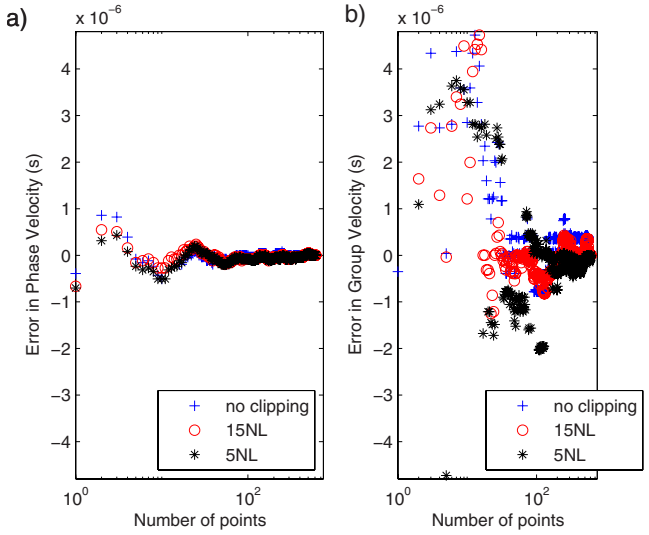


FIG. 9. (Color online) (a) Measurement error in phase velocity (computed with respect to the value obtained for N=610) vs the number of sources N (logarithmic scale) for the coherent A_0 mode arrival extracted from CDF after using the same three pre-processing amplitude clipping procedure as in Fig. 8(b). Same as (a) but for group velocity.

low clipping level (5NL, dashed line) improves only slightly the coherent SNR when compared to the values obtained in the absence of clipping (plain line). For instance, in order to obtain an arbitrary value for the coherent SNR of 60 (i.e., 35 dB), the number of secondary sources is increased from, respectively, 384 for a clipping threshold of 15NL to 410 for a clipping threshold of 5NL and 428 if no clipping is applied to the recorded waveforms.

Second, in the context of SHM applications, the precision of the phase and group velocity measurements of the A_0 mode obtained from CDF for a given set of N secondary sources is also studied thereafter, in addition to the usual coherent SNR metric. For a given value of N, the phase (respectively, group) velocity was computed, respectively, based on the arrival time of the first maximum (i.e., A_0 mode) of the symmetric cross-correlation waveform (respectively, envelope). Given the separation distance D=15.6 cm between the two actuators, the asymptotic estimate of the phase (respectively, group) velocity obtained for the largest number N=610 pts, and hence best reconstructed waveforms [see Fig. 6(b)], was estimated as C_p=1057 m/s^{-1} (respectively C_g=1020 m/s^{-1}), which is in good agreement with predicted value for the A_0 mode (wavelength $\lambda \approx 1$ cm) in this thin aluminum plate (Achenbach, 1993). Figure 9 displays the relative variations in arrival times of the maxima of the symmetric cross-correlation waveform and its envelope, with respect to the asymptotic values used to compute C_p and C_g, using the same pre-processing type of the recorded waveforms (5NL, 15NL clipping, or no clipping) as in Fig. 8. A logarithmic scale was used for the horizontal axis in order to enhance the visualization of the variations in C_p and C_g for small values of the parameter N. Overall in all cases, the error in measured arrival times becomes very small ($<$0.5 μs) after N=25 (respectively, N=205) for phase (group) velocity measurements. As expected group velocity measurements [which uses the maximum value of the envelope of $C_{12}(t)$] appear to be more sensitive to the fluctuations

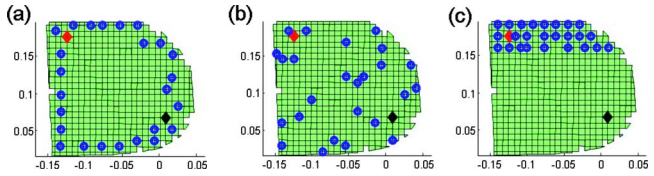


FIG. 10. (Color online) Geometric configuration of the selected $N=25$ measurement points, acting as secondary sources (indicated by circles) for the reciprocal CDF implementation, which were (a) uniformly distributed around the plate, (b) randomly distributed over the plate, or (c) aligned along one side of the plate. The locations of the actuators (see Fig. 3), acting as sensors for the reciprocal CDF implementation, are indicated by diamond shapes.

levels of the cross-correlation waveforms [i.e., variance $V_{12}(t)$ (see Eq. (6))] than phase velocity measurements [which uses the maximum value of $C_{12}(t)$ itself].

Finally, the effect of the spatial configuration of the secondary ultrasonic sources on the CDF performance was assessed in a practical SHM scenario where only 25 secondary ultrasonic sources [i.e., $j=1,\ldots,25$ in Eqs. (1) and (2)] would be used to remotely extract the coherent A_0 mode between locations 1 and 2 (indicated by diamond shapes in Fig. 10). To do so, 25 points of the measurement grids (acting as secondary sources in this reciprocal CDF implementation, (see first paragraph of this section) were selected on the plate along three different patterns: surrounding the whole plate [Fig. 10(a)], randomly located [Fig. 10(b)], or aligned along the three top lines of the measurement grid [Fig. 10(c)]. For each configuration, the measured reverberant waveforms were first filtered in the frequency band $B=4-90$ kHz and clipped at 15NL. Figure 12 shows that the cross-correlation waveforms obtained from CDF using these three different sources configurations are highly similar, especially during the first symmetric A_0 arrival ($100\ \mu s < t < 200\ \mu s$). Furthermore, the measured A_0 arrival time ($t=0.148$ ms) using these $N=25$ sources is within 0.2% of the asymptotic value obtained using for the full set ($N=610$ pts) of measurement points [see Figs. 8(b) and Fig. 9]. Overall, Fig. 11 demonstrates that practical implementations of the CDF technique for SHM applications are feasible even using a small number of secondary sources. Furthermore, Fig. 11 shows that the performance of the CDF technique appears quite insensitive to the exact source locations as long as the recorded (or processed) reverberant field of the structure approximates well a fully diffuse wavefield, as expected from previous numerical simulations (Larose *et al.*, 2008).

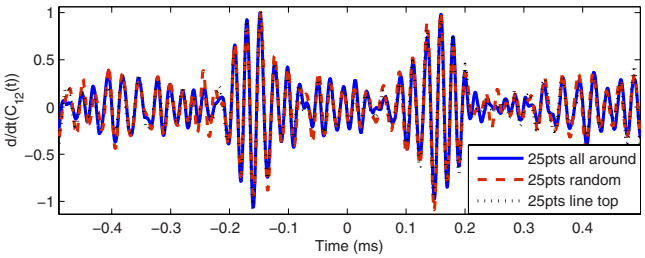


FIG. 11. (Color online) Normalized time-derivatives of the cross-correlation waveforms obtained from CDF for the three configurations of secondary sources displayed in Fig. 10.

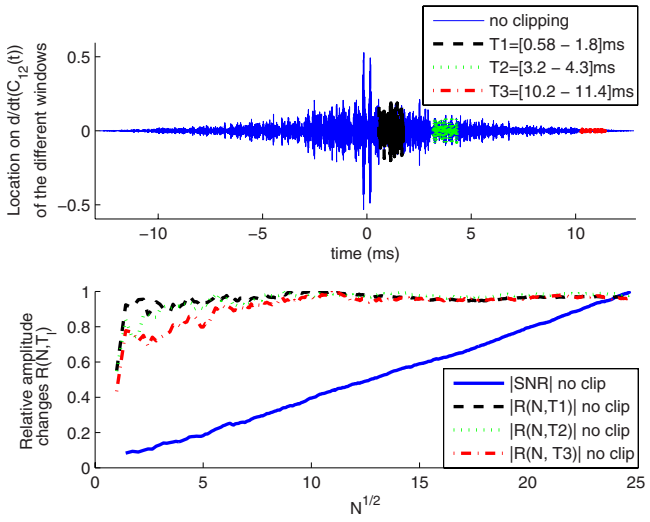


FIG. 12. (Color online) (a) The three selected time-windows: $T_1=0.58-1.8$ ms, $T_2=3.2-4.3$ ms, $T_3=10.2-11.4$ ms used for the analysis of the emergence rate of the late coherent arrivals are indicated on one of the computed cross-correlation waveforms (obtained here from $N=610$ secondary sources). (b) Normalized variations in the ratio $R(N;T_l)$ [see Eq. (11)] for increasing values of $\sqrt{N}$ for the three time-windows shown in (a). The normalized variations in the coherent SNR [see Eq. (4)] are also indicated for comparison (solid line).

C. Emergence of coherent coda arrivals from CDF

Based on the previous discussion, a clear coherent A_0 arrival emerges from cross-correlation waveform when using a sufficient number of sources N [see Figs. 6(b) and 11]. But, the computed cross-correlation waveform $C_{12}(t)$ contains also other weaker late arrivals for any given value of N [see Fig. 12(a)]. Hence a practical question for SHM applications is to determine beyond which point these late arrivals correspond to actual coherent coda arrivals of Green's function as opposed to residual temporal fluctuations due to the imperfect convergence of the CDF technique. One simple way to distinguish between these two situations is to measure the variations in relative amplitude changes $R(N;T_l)$ of the coherent A_0 arrival amplitude with respect to the standard deviation of these late arrivals averaged over a given time interval T_l, defined as

$$R(N;T_l)=\frac{\max_t(C_{12}(t;N))}{\mathrm{std}(C_{12}(t;N);t\in T_l)}, \tag{11}$$

where the cross-correlation waveform $C_{12}(t;N)$ is computed using Eq. (2) for a increasing number N of randomly selected secondary sources. No amplitude clipping was applied to the recorded waveforms. Since CDF is a coherent process, the amplitude of any coherent arrivals (e.g., related to an actual arrival of the local Green's function) would grow linearly with N. On the other hand, the amplitude of residual temporal fluctuations (e.g., due to incoherent electronic noise) only grows as $\sqrt{N}$ (Sabra *et al.*, 2005b; Weaver and Lobkis, 2005; Larose *et al.*, 2008). Hence the ratio $R(N;T_l)$ theoretically grows as $\sqrt{N}$, as long as the arrivals in the window T_l are dominated by residual fluctuations, but then plateaus at a fixed value when coherent arrivals actually emerge in the selected window T_l. Figure 12(b) displays the normalized variations in the ratio $R(N;T_l)$ vs $\sqrt{N}$ for three successive

time-windows T_l selected from the reference cross-correlation waveforms [see Fig. 12(a)]. The normalized variations of the coherent SNR(t) for increasing N (same as Fig. 8) are also indicated as a reference. As expected the ratio $R(N;T_l)$ starts to plateau more rapidly for increasing N when the selected time-window T_l contains early coherent arrivals (dashed line, T_1=0.58–1.8 ms) as opposed to a later time-window (dashed line, T_3=10.2–11.4 ms). Hence Fig. 14(b) shows that the emergence of the late coherent arrivals is thus slower than for early coherent arrivals for this experimental set-up. For instance, the ratio $R(N;T_l)$ reaches 95% of the asymptotic plateau value after N=4 for the window T_1 =0.58–1.8 ms, after N=28 for the window T_2=3.2–4.3 ms, and after N=73 for the window T_3=10.2–11.4 ms. Hence, in practice, for sufficient number of sources N, even the late coda arrivals (i.e., included in the last window T_3 =10.2–11.4 ms) appear to be genuine coherent arrivals of $C_{12}(t;N)$ which could potentially be used for structural monitoring purposes (e.g., for passive coda wave interferometry) (see Hadziioannou *et al.*, 2009).

D. Influence of the duration and signal-to-noise ratio of the recorded reverberant signals

In addition to increasing the number of sources N or reducing the clipping level of the measured waveforms, the coherent SNR of the cross-correlation waveforms can also be improved by using a longer duration T of the recorded waveforms in order to benefit from the multiple scattering effects [see Eqs. (9) and (10)]. To assess these theoretical predictions, the evolution of the coherent SNR for increasing recording duration T was experimentally measured by first pre-processing the same large portion of the filtered reverberant waveforms corresponding to the time-window (0.43–4.3 ms) of duration $T_{\max}$=3.87 ms set by the 15NL threshold limit, but using either a clipping level of 5NL, 15NL, or no clipping at all (similar to the procedure used to generate Fig. 6), and second by cross-correlating various epochs of this processed signals having increasing duration T from 0 to $T_{\max}$ in 40 μs increment. In the selected time-window (0.43–4.3 ms), the exponential decay of the reverberant field was estimated as τ_D=3.8 ms [see Eq. (7) and the procedure described in Sec. III B] and the average signal-to-noise ratio of the recorded waveforms was measured as SNR$_r$=616$\gg$1. The experimental variations of coherent SNR(t) measured at the A_0 arrival time t=0.148 ms are displayed in Fig. 13 as a function of the square-root normalized duration $\sqrt{T/T_{\max}}$ for N=25, N=100, or N=200. These measured variations in SNR(t) appear to be very similar regardless of which pre-processing types (5NL, 15NL, or no clipping) or specific number of secondary sources N. Equation (11) was used to predict the theoretical variations of SNR(t^*) (dot dashed line) computed for $0<T<T_{\max}$ obtained from Eq. (10) using $T_{\max}$=3.87 ms, t^*=148 μs, τ_D=3.7 ms, SNR$_r$=616, which appears in good agreement with the measured ones (see Fig. 14). Both experimental and measured variations in the coherent SNR(t) tend toward the expected plateau when $T>\tau_D$ (i.e., when all the energetic part of the recorded waveforms has been processed).

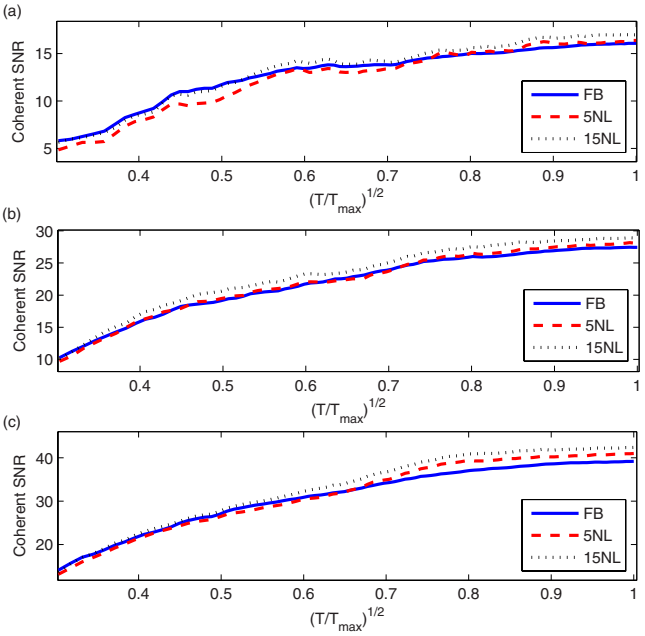


FIG. 13. (Color online) Evolution of the coherent SNR vs the normalized recording duration $\sqrt{T/T_{\max}}$ for $T_{\max}$=3.87 ms using the three different clipping level for the recorded waveforms: no clipping (plain line), 5NL (dashed line), and 15NL (dotted line) with (a) N=25, (b) N=100, or (c) N=200.

Finally, the effect of incoherent measurement noise on the performance of the CDF is then assessed quantitatively by dividing each of the reverberant signal collected over the whole measurement grid into four arbitrary time-windows T_1=0–4.6133 ms, T_2=1.3672–5.9258 ms, T_3 =2.7344–7.2930 ms, and T_4=4.1016–8.6602 ms [see Fig. 15(a)]. The recorded waveforms were further filtered in the band B=4–90 kHz and a low clipping level of 5NL was applied to each time-window [e.g., see Fig. 8(a)]. Figure 15(b) displays the four computed cross-correlation time functions from CDF obtained using all processed reverberant signals (N=610), similar to the procedure used in Sec. IV A. The coherent A_0 arrival is nearly identical for the four cross-correlations time functions at t^*=148 μs. This highlights the robustness of the CDF process for obtaining deterministic Green's function estimate and illustrates that both early (e.g., window T_1) and late coda (e.g., window T_4) portions of the diffuse recordings—of similar duration—contain a comparable amount of coherent propagating waves, or "coherent

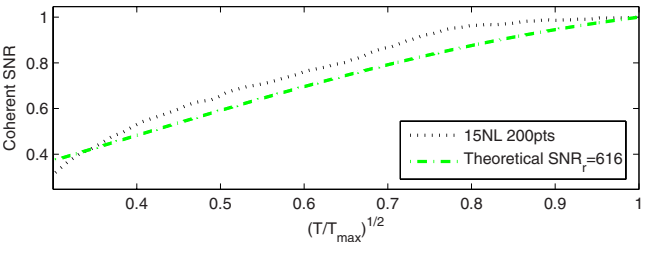


FIG. 14. (Color online) Comparison between the experimental variations in the coherent SNR of the cross-correlation waveforms, obtained using a clipping at 15NL and N=200 points [same as Fig. 13(c), dotted line], and the theoretical predictions, obtained from Eq. (10), for increasing recording duration T ($0<T<T_{\max}$) using the paramaters $T_{\max}$=3.87 ms, t^*=148 μs, τ_D =3.7 ms, SNR$_r$=616.

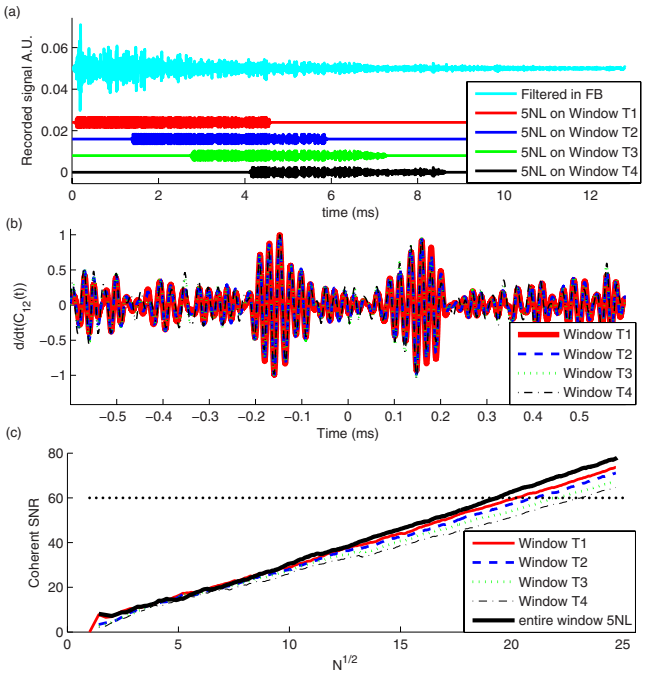


FIG. 15. (Color online) (a) Typical recorded waveform filtered in B =4–90 kHz along with the four selected time-windows T_1=0–4.6133 ms, T_2=1.3672–5.9258 ms, T_3=2.7344–7.2930 ms, and T_4=4.1016 –8.6602 ms of the clipped waveform at 5NL. (b) Coherent A_0 arrivals obtained by cross-correlating the recorded waveforms in the four selected time-windows T_1–$T4$ (N=610). The time-derivative of each cross-correlation time function are superimposed. (c) Variations in the coherent SNR(t^*) (linear scale) vs $\sqrt{N}$ for obtained from CDF using the time-windows T_1–$T4$ (t^*=148 μs). The variations in the coherent SNR(t^*) obtained using the full duration of the clipped waveforms (0–9.1 ms) is also shown for comparison (thick plain line).

information," between the actuators' locations. However, the effective recorded SNR$_r$ decreases over the four selected time-windows since the amplitude of the reverberant wavefield decays exponentially with time (see Sec. III B), while the amplitude of the incoherent measurement noise likely remains constant throughout the recording. Consequently, in agreement with the theoretical predictions from Eq. (10), the achieved coherent SNR(t) for a given number of secondary sources N is lower when using late coda waves (e.g., window T_4) when compared to early coda waves results (e.g., window T_1) [see Fig. 15(c)]. For instance, in order to obtain an arbitrary value of 60 for the coherent SNR(t^*) of the A_0 arrival, the number of secondary sources is increased from, respectively, N=408 in the window T_1, to N=445 in the window T_2, N=488 in the window T_3, and finally N=535 in the window T_4. The variations in the coherent SNR(t^*) obtained using the full duration of the clipped waveforms (0–9.1 ms) is also shown for comparison (thick plain line) and is higher for the same value of N, as expected from Fig. 13.

E. Parametric variations of the coherent SNR vs *N* and *T*

Figure 16 displays contour plots of the variations in coherent SNR(t^*) of the A_0 arrival for increasing values of T and N using three different pre-processing of the filtered (B =4–90 kHz) reverberant waveforms: 5NL or 15NL clipping level and no clipping at all (similar to the procedure used to

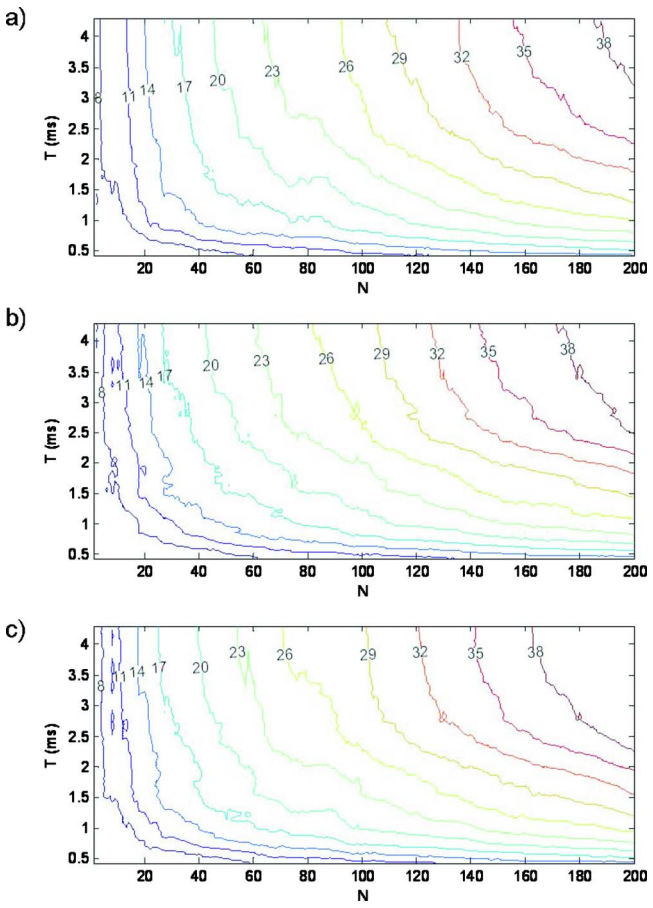


FIG. 16. (Color online) Parametric variations in the coherent SNR [linear scale, see Eq. (4)] as a function of the recording duration T and number of secondary sources N (selected randomly across the measurement plate). Each contour represents a constant SNR level at the indicated value from 8 (i.e., 18 dB) to 38 (i.e., 32 dB). Prior to cross-correlations, the recorded waveforms were first filtered in B=4–90 kHz and three different clipping levels were applied: (a) no clipping, (b) 5NL, and (c) 15NL.

generate Fig. 8). Each contour represents a constant SNR level at the indicated value from 8 (i.e., 18 dB) to 38 (i.e., 32 dB). Based on the results from Fig. 11 showing that the specific distribution of secondary sources does not significantly influence the coherent symmetric A_0 arrival computed from CDF, only one random realization of the spatial distribution of the secondary sources was used for any given value N (up to N=200 here). These parametric variations in coherent SNR displayed in Fig. 16 are a summary of the combined results displayed in Figs. 8 and 13. As expected, Fig. 16 shows that higher values of coherent SNR can be achieved by increasing either the recording duration T or the number of sources N. Furthermore, following the findings from Figs. 7 and 8, Fig. 16 also illustrates the possibility of using amplitude clipping of the recorded data to further reduce the required value of T or N to achieve a given coherent SNR value, as seen, for instance, by comparing the SNR contour levels in Figs. 16(a) and Fig. 16(c).

Furthermore, the asymmetric shape of each SNR contour curves with respect to the variables T and N also demonstrates that an increase in the recording duration T cannot always compensate for a lack of secondary sources when using noisy, multiply scattered signals. Indeed the near ver-

tical portion of these parametric curves for small value of N shows that the coherent SNR(t) tends toward a plateau when $T > \tau_D$ due to the exponential decay of the measured reverberant waveforms and the influence of the incoherent measurement noise for low recorded SNR_r [see Eq. (10) and Fig. 13]. Hence, Fig. 16 suggests that a given value of coherent SNR could potentially be achieved optimally in the "elbow" region of each curve by minimizing both the necessary recording duration T and number of sources N. For instance, inspection of Fig. 16 shows that a coherent SNR level of 35 (i.e., 30 dB) could be optimally achieved with the following parameter choices (T=3.25 ms, N=165) if no clipping of the recorded waveform is used, (T=2.82 ms, N=163) for a 5NL clipping level and (T=2.78 ms, N=152) for a 15NL clipping level.

Finally, based on the results obtained for this thin plate, it appears that a relatively large number of secondary sources N ($\sim$100) is required in order to achieve both high coherent SNR and high precision for the phase (or group) velocity measurements from the reconstructed A_0 arrival. This requirement on N may result both from the presence of strong elastic attenuation (which increases the temporal decay of the diffuse field and thus limits the effective duration of the diffuse field record) and the small number of propagating modes in this plate at low frequencies (mainly S_0 and A_0 modes here at 50 kHz) in contrary to previous experiments on small aluminum blocks at higher frequencies (Weaver and Lobkis, 2005).

V. CONCLUSIONS

The performance of the CDF technique was investigated theoretically and experimentally using a thin aluminum plate with complex geometry and boundaries, in the frequency bandwidth 4–90 kHz. Estimates of the A_0 Lamb wave component of the local GF as well as late coherent coda waves arrivals propagating between two arbitrary locations on the plate were extracted from CDF. Hence this highlights the robustness of CDF to reconstruct *remotely* the *local* elastic response (including ballistic and coda arrivals) between a pair of (embedded) sensors. Hence the proposed CDF technique may thus provide a new inherently safe alternative for structural health monitoring systems in complex structures, especially when the local use of active sources is limited or prohibited.

The coherent SNR of the Green's function estimates obtained from CDF sets the achievable accuracy for damage detections between a pair of passive sensors (e.g., for detecting the weak scattered field from a local crack). Theoretical predictions, using a simple model for the exponentially decaying reverberant wavefield, and experimental measurements of the dependency of the coherent SNR on independent parameters that the operator can potentially control—such as the recordings duration T, the number of secondary sources N, or signal-to-noise ratio of the selected portion of the recorded waveforms—were developed. These predictions and measurements of the coherent SNR could potentially be used to quantify the probability of detection of CDF for SHM applications (Achenbach, 2000).

Achenbach, J. D. (**1993**). *Wave Propagation in Elastic Solids* (North Holland, Amsterdam).

Achenbach, J. D. (**2000**). "Quantitative non-destructive evaluation," Int. J. Solids Struct. **37**, 13–27.

Akolzin, A., and Weaver, R. L. (**2004**). "Generalized Berry conjecture and mode correlations in chaotic plates," Phys. Rev. E **70**, 046212.

Derode, A., Tourin, A., and Fink, M. (**1999**). "Ultrasonic pulse compression with one-bit time reversal through multiple scattering," J. Appl. Phys. **85**, 6343–6352.

Evans, M. J., and Cawley, P. (**1999**). "Measurement and prediction of diffuse fields in structures," J. Acoust. Soc. Am. **106**, 3348–3360.

Farrar, C., and James, G. (**1997**). "System identification from ambient vibration measurements on a bridge," J. Sound Vib. **205**, 1–18.

Giurgiutiu, V. (**2007**). *Structural Health Monitoring: With Piezoelectric Wafer Active Sensors* (Academic, New York).

Hadziioannou, C., Larose, E., Coutant, O., Roux, P., and Campillo, M. (**2009**). "Stability of monitoring weak changes in multiply scattering media with ambient noise correlation: Laboratory experiments," J. Acoust. Soc. Am. **125**, 3688–3695.

Langley, R. S. (**2007**). "On the diffuse field reciprocity relationship and vibrational energy variance in a random subsystem at high frequencies," J. Acoust. Soc. Am. **121**, 913–921.

Larose, E., Derode, A., Campillo, M., and Fink, M. (**2004**). "Imaging from one-bit correlations of wide-band diffuse wavefields," J. Appl. Phys. **95**, 8393–8399.

Larose, E., Lobkis, O. I., and Weaver, R. L. (**2006**). "Passive correlation imaging of a buried scatterer," J. Acoust. Soc. Am. **119**, 3549–3552.

Larose, E., Roux, P., and Campillo, M. (**2007**). "Reconstruction of Rayleigh-Lamb dispersion spectrum based on noise obtained from an air-jet forcing," J. Acoust. Soc. Am. **122**, 3437–3444.

Larose, E., Roux, P., Campillo, M., and Derode, A. (**2008**). "Fluctuations of correlations and Green's function reconstruction: Role of scattering," J. Appl. Phys. **103**, 114907.

Paul, A., Campillo, M., Margerin, L., Larose, E., and Derode, A. (**2005**). "Empirical synthesis of time-asymmetrical Green functions from the correlation of coda waves," J. Geophys. Res. **110**, B08302.

Proakis, J. G. (**1995**). *Digital Communications* (McGraw-Hill, New York), pp. 159–163.

Roux, P., Sabra, K. G., Kuperman, W., and Roux, A. (**2005**). "Ambient noise cross correlation in free space: Theoretical approach," J. Acoust. Soc. Am. **117**, 79–84.

Sabra, K. G., Roux, P., and Kuperman, W. A. (**2005a**). "Arrival-time structure of the time-averaged ambient noise cross-correlation function in an oceanic waveguide," J. Acoust. Soc. Am. **117**, 164–174.

Sabra, K. G., Roux, P., and Kuperman, W. A. (**2005b**). "Emergence rate of the time domain Green's function from the ambient noise cross correlation," J. Acoust. Soc. Am. **118**, 3524–3531.

Sabra, K. G., Gerstoft, P., Roux, P., Kuperman, W. A., and Fehler, M. C. (**2005c**). "Surface wave tomography from microseisms in Southern California," Geophys. Res. Lett. **32**, L14311.

Sabra, K. G., Winkel, E. S., Bourgoyne, D. A., Elbing, B. R., Ceccio, S. L., Perlin, M., and Dowling, D. R. (**2007**). "Using cross correlations of turbulent flow-induced ambient vibrations to estimate the structural impulse response. Application to structural health monitoring," J. Acoust. Soc. Am. **121**, 1987–2005.

Sabra, K. G., Srivastava, A., Lanza di Scalea, F., Bartoli, I., Rizzo, P., and Conti, S. (**2008**). "Structural health monitoring by extraction of coherent guided waves from diffuse fields," J. Acoust. Soc. Am. **123**, EL8–EL13.

Shapiro, N. M., Campillo, M., Stehly, L., and Ritzwoller, M. (**2005**). "High resolution surface-wave tomography from ambient seismic noise," Science **307**, 1615–1617.

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

Snieder, R., and Cafak, E. (**2006**). "Extracting the building response using seismic interferometry; theory and application to the Millikan Library in Pasadena, California," Bull. Seismol. Soc. Am. **96**, 586–598.

Van Wijk, K. (**2006**). "On estimating the impulse response between receivers in a controlled ultrasonic model," Geophysics **71**, SI79–SI84.

Weaver, R. L. (**1982**). "On diffuse waves in solid media," J. Acoust. Soc. Am. **71**, 1608–1609.

Weaver, R. L. (**1984**). "Diffuse waves in finite plates," J. Sound Vib. **94**, 319–335.
Weaver, R. L. (**1986**). "Laboratory studies of diffuse waves in plates," J. Acoust. Soc. Am. **79**, 919–923.
Weaver, R. L., and Lobkis, O. I. (**2001**). "Ultrasonics without a source: Thermal fluctuation correlations at MHz frequencies," Phys. Rev. Lett. **87**, 134301.
Weaver, R. L., and Lobkis, O. I. (**2004**). "Diffuse fields in open systems and the emergence of the Green's function," J. Acoust. Soc. Am. **116**, 2731–2734.
Weaver, R. L., and Lobkis, O. I. (**2005**). "The mean and variance of diffuse field correlations in finite bodies," J. Acoust. Soc. Am. **118**, 3447–3456.

Prediction of noise reduction through vehicle path rerouting

Rufin Makarewicz
Institute of Acoustics, A. Mickiewicz University, Umultowska 85, 61-614 Poznań, Poland

(Received 3 June 2009; revised 28 September 2009; accepted 1 October 2009)

Traffic noise, produced by moving vehicles, is described in terms of the day-night average sound level L_{dn} and the day-evening-night level L_{den}. Both levels depend on the path of noise sources, such as road vehicles or aircraft. Usually, noise source path is rectilinear. Substituting a circle arc of radius R for the straight path segment yields noise reduction $\Delta L=\Delta L_{dn}=\Delta L_{den}$. Assuming that noise propagation is governed by geometrical spreading, air absorption, and ground effect, relationship between ΔL and R is derived. For example, replacement of a straight road at the distance of 100 m with the road of radius $R=270$ m and the angle $\Phi=68°$ yields the noise reduction of 4 dB. In the case of road traffic noise, the presented results seem to be a viable alternative to barrier construction. © *2010 Acoustical Society of America.* [DOI: 10.1121/1.3257229]

PACS number(s): 43.50.Lj [KVH] Pages: 216–222

I. INTRODUCTION

The idea of source rerouting seems to be quite new (see the literature cited in Ref. 1). It happens that the aircraft or highway noise exceeds the prescribed limit because the perpendicular distance d to the rectilinear track is too short (Fig. 1). It is the most common cause of extremely high noise level. In Ref. 1 it has been shown that the best shape for a minimum noise route is part of a lemniscate based on the receiver. However, cars on highways and aircraft have to be smooth without inflection points. The current study is more realistic because the circular segments of radius $R=d/\cos\Phi$ are considered (Fig. 2). The proposed method is focused on a single receiver position.

Substituting a circle arc of radius R for the straight path segment, $-R\sin\Phi<x<+R\sin\Phi$, seems to be an engineering practice. So far its noise reduction outcome has not been quantified. The problem is how the noise reduction ΔL and its efficiency η (expressed in \$/dB) depend on the angle Φ, i.e., on the radius R.

In the United States the day-night average sound level is used as follows:

$$L_{dn}=10\log\left(\frac{15}{24}10^{0.1L_d}+\frac{9}{24}10^{0.1(L_n+10)}\right), \tag{1}$$

where L_d and L_n are day and night average noise levels $L_{Aeq,T}$ for $T_d=15\times3600$ s (7:00–22:00) and $T_n=9\times3600$ s hours (22:00–7:00), respectively. In the European Union environmental noise is assessed in terms of the day-evening-night level

$$L_{den}=10\log\left(\frac{12}{24}10^{0.1L_d}+\frac{4}{24}10^{0.1(L_e+5)}+\frac{8}{24}10^{0.1(L_n+10)}\right), \tag{2}$$

where L_d, L_e, and L_n represent, respectively, the day, evening, and night average noise levels $L_{Aeq,T}$ for $T_d=15\times3600$ s (7:00–19:00), $T_e=4\times3600$ s (19:00–23:00), and $T_n=8\times3600$ s (23:00–7:00).

In Sec. II it will be shown that the path rerouting (Fig. 2) leads to identical reduction in L_{dn} and L_{den}

$$\Delta L=L_{dn}-\tilde{L}_{dn}=L_{den}-\tilde{L}_{den}, \tag{3}$$

where the tilde $\tilde{L}$ means the level of noise from the new path. The dependence of ΔL on the angle Φ and the perpendicular distance d is derived in Sec. III. If D, H, and h denote the horizontal distance, the source height, and receiver height (Fig. 3), then the perpendicular distance becomes

$$d=\sqrt{D^2+(H-h)^2}. \tag{4}$$

The cost of noise reduction is proportional to the difference in the path lengths before and after rerouting (Fig. 2)

$$\Delta l=l-\tilde{l}=2\frac{\Phi-\sin\Phi}{\cos\Phi}d. \tag{5}$$

The higher the noise reduction, the higher the costs. Therefore the cost-benefit analysis can be based on the rate of 1 dB noise reduction [Eqs. (3) and (5)],

$$\eta=C\frac{\Delta l(d,\Phi)}{\Delta L(d,\Phi)}\ \ \$/\mathrm{dB}, \tag{6}$$

where C denotes the cost of the increase in gasoline usage, land use costs, as well as a construction cost per extra meter of road length. A large number of η imply a higher cost and lower efficiency.

There are readily available methods to predict the noise reduction by any source path rerouting. The presented method here is developed for its mathematical simplicity as much as for its capacity to show the nature of noise propagation and noise control.

II. SOUND EMISSION AND PROPAGATION

Usually, the road traffic noise comes from a few categories of vehicles (e.g., automobiles and medium and heavy trucks). Similarly, there are many categories of aircraft. In general, for m categories of noise sources, which move along the same track, there are m values of the time average sound level $L^{(1)}_{Aeq,T},\ldots,L^{(m)}_{Aeq,T}$, which combine into the time average sound level of noise that comes from all source categories,

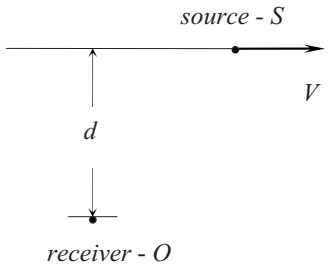


FIG. 1. Top view of the straight path of noise source S that cruises at a steady speed V. Perpendicular distance d is defined by Eq. (4).

$$L_{Aeq,T} = 10 \log\left\{ \sum_{i=1}^{m} 10^{0.1 L_{Aeq,T}^{(i)}} \right\}. \tag{7}$$

Now, we will apply the relationship between the measure of a single source noise, i.e., the A-weighted sound exposure level $L_{AE}^{(i)}$ and the corresponding time average sound level,[2]

$$L_{Aeq,T}^{(i)} = 10 \log\left\{ \frac{N_i t_o}{T} \right\} + L_{AE}^{(i)}, \quad t_o = 1 \text{ s}. \tag{8}$$

Here N_i is the number of noise sources of the ith category that passes by the receiver during the time interval T (day, evening, or night—see above). Subsequently, the A-weighted sound exposure level can be written as[2]

$$L_{AE} = 10 \log\{\varepsilon\}, \tag{9}$$

where the sound exposure, strictly speaking "relative A-weighted sound exposure," equals

$$\varepsilon = \int_{-\infty}^{+\infty} \frac{p_A^2}{p_o^2} \frac{dt}{t_o}, \quad p_o = 20 \ \mu\text{Pa}. \tag{10}$$

In the above integral, p_A^2 denotes the time pattern of the A-weighted squared sound pressure due to a single source (Fig. 4).

Before rerouting, when the source cruises at a steady speed V along a straight track from $x=-\infty$ to $x=+\infty$ (Fig. 5), the increment of the path length is $dx=(d/V\cos^2\phi)d\phi$ and the sound exposure takes the form [Eq. (10)]

$$\varepsilon = \frac{d}{Vt_o} \int_{-\pi/2}^{+\pi/2} \frac{p_A^2}{p_o^2} \frac{d\phi}{\cos^2\phi}. \tag{11}$$

After rerouting, the source moves along a straight track from $x=-\infty$ to $x=d\cdot tg(-\Phi)$, then along the arc of the circle of constant radius $R=d/\cos\Phi$, and again along a straight track from $x=d\cdot tg(+\Phi)$ to $x=+\infty$ (Fig. 2). In other words, the

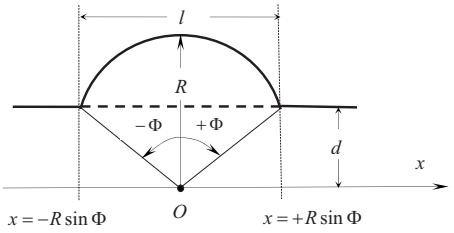


FIG. 2. Top view of the straight segment of length $l=2d\cdot tg\Phi$, which is replaced by the circular segment of the radius $R=d/\cos\Phi$ and the length $\tilde{l}=2\Phi d/\cos\Phi$ [Eq. (5)].

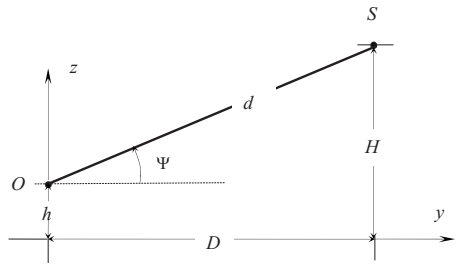


FIG. 3. Perpendicular distance d [Eq. (4)] between the straight path of the source S and receiver O, with the angle $\Psi=\arctan[(H-h)/D]$ (side view).

circle segment starts at $x=-R\sin\Phi$, ends at $x=+R\sin\Phi$, and disappears when $\Phi\to 0$. Under such circumstances, the noise reduction ΔL tends to zero as well (see Figs. 6, 8, and 9).

On the circular path of radius R, the time increment is $dt=Rd\phi/V$, and the corresponding sound exposure equals [Eq. (10)]

$$\varepsilon_c = \frac{p_{Ac}^2}{p_o^2} \frac{2\Phi d}{Vt_o \cos\Phi}. \tag{12}$$

Adding the measure of the noise from the rectilinear track segments $(-\pi/2,-\Phi)$ and $(+\pi/2,+\Phi)$ to the difference $\varepsilon_c - \varepsilon_r$, the total sound exposure of noise from the new track becomes

$$\tilde{\varepsilon} = \varepsilon - \varepsilon_r + \varepsilon_c. \tag{13}$$

Here, ε is defined by Eq. (11) and

$$\varepsilon_r = \frac{d}{Vt_o} \int_{-\Phi}^{+\Phi} \frac{p_A^2}{p_o^2} \frac{d\phi}{\cos^2\phi} \tag{14}$$

describes the noise from the rectilinear track segment $-\Phi < \phi < +\Phi$ (Fig. 2).

For the ith category of the noise sources, the relative A-weighted squared sound pressure of noise generated by a single source can be written as[2]

$$\frac{p_{Ai}^2}{p_o^2} = \frac{W_{Ai}(V)}{W_o} \frac{d_o^2}{4\pi r^2} F(r), \quad d_o = 1 \text{ m}, \tag{15}$$

where $W_o=10^{-12}$ W, $W_{Ai}(V)$ represents the speed dependent A-weighted sound power, r is the instantaneous distance between a moving source and a receiver (Fig. 5), and the function $F(r)$ quantifies sound attenuation in excess of geometrical spreading. Now, Eqs. (11)–(15) give the sound exposures before and after rerouting[3]

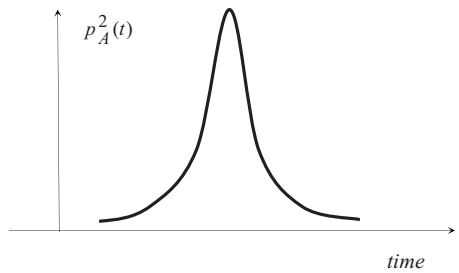


FIG. 4. Time pattern of the A-weighted squared sound pressure $p_A^2(t)$ due to a single source moving at a steady speed along the track shown in Fig. 1.

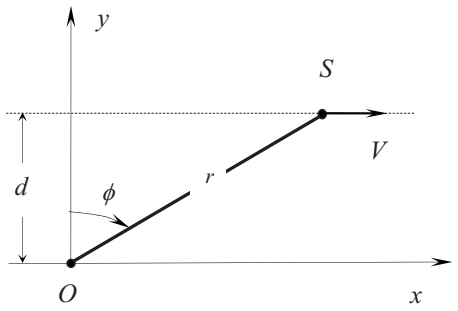


FIG. 5. Instantaneous distance between the cruising source S and the receiver O: $r=d/\cos\phi$.

$$\varepsilon_i = S_i \cdot K, \quad \tilde{\varepsilon}_i = S_i \cdot \tilde{K}, \tag{16}$$

where

$$S_i = \frac{W_{Ai}(V_i)}{W_o}\frac{d_o}{V_i t_o} \tag{17}$$

characterizes noise emission by the source of the ith category, either on the old or new path. The *propagation functions* K and $\tilde{K}$ for the old and new paths (Figs. 1 and 2) are determined by

$$K = \frac{d_o}{4\pi d}\int_{-\pi/2}^{+\pi/2} F\left(\frac{d}{\cos\phi}\right)d\phi \tag{18}$$

and

$$\tilde{K} = K - \Delta K, \tag{19}$$

respectively. Here

$$\Delta K = \frac{d_o}{4\pi d}\left[\int_{-\Phi}^{+\Phi} F\left(\frac{d}{\cos\phi}\right)d\phi - 2\Phi\cos\Phi F\left(\frac{d}{\cos\Phi}\right)\right] \tag{20}$$

describes the noise reduction due to substitution of the rectilinear track segment $-\Phi<\phi<+\Phi$ by the circular one (Fig. 2). Accordingly, by applying Eqs. (8), (16), and (17) one finds the time average sound levels for noise sources of the ith category before and after rerouting,

$$L^{(i)}_{\mathrm{Aeq},T} = 10\log\left\{N_i\frac{W_{Ai}}{W_o}\frac{d_o}{V_i T}\right\} + 10\log K \tag{21}$$

and

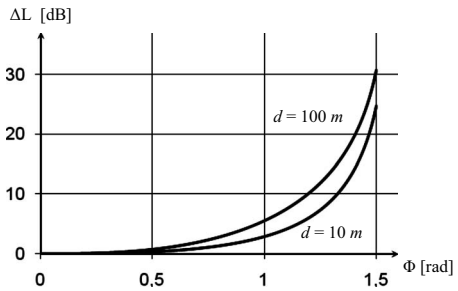


FIG. 6. Noise reduction ΔL [Eqs. (31), (39), and (40)] calculated for the heights $h+H=2$ m (Fig. 3), ground coefficient $\gamma=0.008$, and two perpendicular distances $d=10$, 100 m.

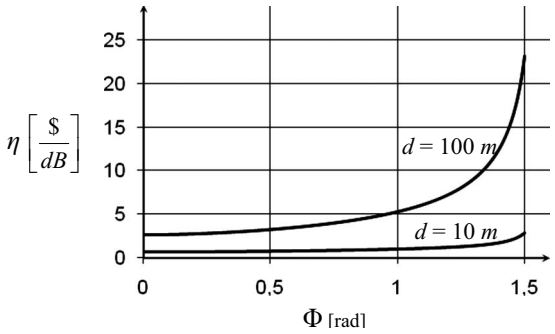


FIG. 7. Noise reduction efficiency η [Eqs. (32), (39), and (40)] calculated for $2C=1$, heights $h+H=2$ m (Fig. 3), ground coefficient $\gamma=0.008$, and perpendicular distances $d=10$, 100 m.

$$\tilde{L}^{(i)}_{\mathrm{Aeq},T} = 10\log\left\{N_i\frac{W_{Ai}}{W_o}\frac{d_o}{V_i T}\right\} + 10\log\tilde{K}. \tag{22}$$

Consequently, inserting the above two formulas into Eq. (7) one gets the time average sound level of noise from all sources,

$$\begin{aligned} L_{\mathrm{Aeq},T} &= L^{(\mathrm{eq})}_{WA} + 10\log K(d), \\ \tilde{L}_{\mathrm{Aeq},T} &= L^{(\mathrm{eq})}_{WA} + 10\log\tilde{K}(d,\Phi). \end{aligned} \tag{23}$$

The rerouting does not change the sound power level of noise radiated from the track of unit length,

$$L^{(\mathrm{eq})}_{WA} = 10\log\left\{\frac{W_{\mathrm{Aeq}}}{W_o}\right\}, \tag{24}$$

where

$$W_{\mathrm{Aeq}} = \sum_{i=1}^{m}\frac{N_i W_{Ai}}{V_i T}d_o \tag{25}$$

expresses the sound power from $d_o=1$ m.

In order to determine the values of L_{dn} [Eq. (1)] and L_{den} [Eq. (2)], we introduce the sound power W_{Aeq} for the day, evening, and night time hours [Eq. (25)],

$$W^{(d)}_{\mathrm{Aeq}} = \sum_{i=1}^{m}\frac{N_{di}W_{Ai}}{V_{di}T_d}d_o, \quad W^{(e)}_{\mathrm{Aeq}} = \sum_{i=1}^{m}\frac{N_{ei}W_{Ai}}{V_{ei}T_e}d_o, \tag{26}$$

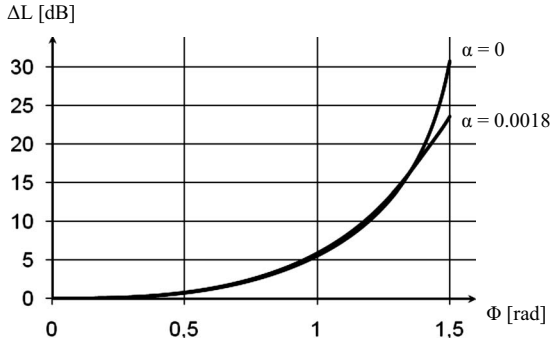


FIG. 8. Noise reduction ΔL [Eqs. (31), (44), and (46)] calculated for the heights $h+H=2$ m (Fig. 3), ground coefficient $\gamma=0.008$, perpendicular distance $d=100$ m, and two air absorption coefficients $\alpha=0$ and $\alpha=0.0018$.

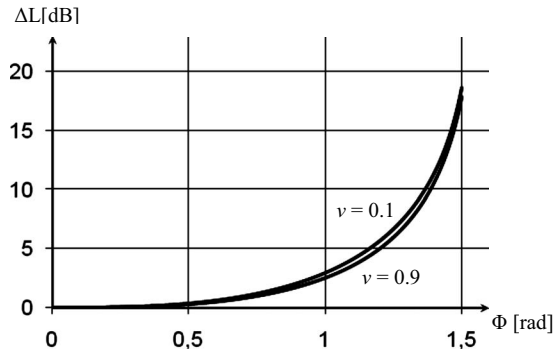


FIG. 9. Noise reduction ΔL [Eqs. (31), (58), and (59)] calculated for air absorption coefficient $\alpha=0.0018$, perpendicular distance $d=1000$ m, and two directivity coefficients $\nu=0.1$ and $\nu=0.9$.

$$W_{\mathrm{Aeq}}^{(n)} = \sum_{i=1}^{m} \frac{N_{ni} W_{Ai}}{V_{ni} T_n} d_o.$$

Finally Eqs. (1), (23), (24), and (26) imply the day-night average sound levels for the old and new routes,

$$L_{\mathrm{dn}} = L_{WA}^{(\mathrm{dn})} + 10 \log K(d), \quad \tilde{L}_{\mathrm{dn}} = L_{WA}^{(\mathrm{dn})} + 10 \log \tilde{K}(d,\Phi), \tag{27}$$

where

$$L_{WA}^{(\mathrm{dn})} = 10 \log\left\{ \frac{W_{\mathrm{Aeq}}^{(d)} + 10 W_{\mathrm{Aeq}}^{(e)}}{W_o} \right\}. \tag{28}$$

Similar relationships one gets for the day-evening-night average sound level [Eqs. (2), (23), (24), and (26)] are as follows:

$$L_{\mathrm{den}} = L_{WA}^{(\mathrm{den})} + 10 \log K(d), \quad \tilde{L}_{\mathrm{dn}} = L_{WA}^{(\mathrm{den})} + 10 \log \tilde{K}(d,\Phi), \tag{29}$$

where

$$L_{WA}^{(\mathrm{den})} = 10 \log\left\{ \frac{W_{\mathrm{Aeq}}^{(d)} + \sqrt{10} W_{\mathrm{Aeq}}^{(e)} + 10 W_{\mathrm{Aeq}}^{(n)}}{W_o} \right\}. \tag{30}$$

Notice that rerouting leads to identical reductions in L_{dn} and L_{den} [Eqs. (3), (19), (27), and (29)],

$$\Delta L = -10 \log\left\{ 1 - \frac{\Delta K(d,\Phi)}{K(d)} \right\}, \tag{31}$$

with the propagation function for the straight track K [Eq. (18) and Fig. 5] and its decrement ΔK [Eq. (20)] caused by the track modification (Fig. 2). Ultimately, the cost-benefit equation takes the following form [Eqs. (5), (6), and (31)]:

$$\eta = \frac{-0.2Cd}{\log\{1 - \Delta K/K\}} \frac{\Phi - \sin\Phi}{\cos\Phi} \ \$/\mathrm{dB}, \tag{32}$$

where C is defined in Eq. (6).

III. SOUND ATTENUATION

In excess of geometrical spreading [Eq. (15)] there are two major mechanisms of sound attenuation, i.e., ground effect and air absorption. Due to excess attenuation, the source rerouting results in noise decrease despite the decreased angle of view (see, e.g., calculation of road traffic noise applied in England).

A. Ground effect

During non-refractive propagation, ground effect can be modeled by[4]

$$F_g(r) = \beta\left[1 + \gamma \frac{r^2}{(h+H)^2} \right]^{-1}, \tag{33}$$

with two adjustable coefficients β and γ. The validity of the above equation has been confirmed.[5] For a large elevation angle Ψ such as $\gamma r^2 \ll (h+H)^2$ (Fig. 3), the only consequence of the ground effect is the noise enhancement $F_g \to \beta$. With the ground coefficient $\gamma \approx 0.010$, the function $F_g(r)$ fits quite well to the aircraft noise data.[6] In any case γ depends on the spectrum of the source power and acoustical qualities of the ground surface (reflective, little absorbing, and absorbing), e.g., above the sound absorbing grass, the value of $\gamma \approx 0.008$ describes quite well the propagation of road traffic noise. When the source is close to the ground (e.g., automobile and landing aircraft), the ground effect becomes important ($\gamma r^2 \gg (h+H)^2$).

Thus, for a non-turbulent medium above a perfectly flat surface, the ground effect function would decrease with the second power of the distance $F_g \propto r^{-2}$. Taking into account the geometrical spreading and ground effect [Eq. (15)], one obtains the squared A-weighted sound pressure $p_A^2/p_o^2 \propto r^{-4}$. This relationship corresponds to the 12 dB reduction of the A-weighted sound level L_{pA} per doubling of the distance $r \to 2r$. This prediction has been experimentally verified for the airport noise (Fig. 19.32 in Ref. 7).

B. Air absorption

Air absorption depends on the sound power spectrum, air temperature τ (°C), and humidity h (%). It has been shown that its approximation could be[8,9]

$$F_a = [1 + \alpha(\tau,h) r]^{-1}. \tag{34}$$

In Ref. 9 the air absorption coefficient α (1/m) has been calculated for a typical power spectrum of highway and railway traffics. Under standard atmospheric conditions, with $\tau = 25$ °C and $h=70\%$, one can find $\alpha = 1.8 \times 10^{-3}$ 1/m. For a wide range of temperatures and humidity, within the distance $r < 500$ m, the approximation error $\Delta L = 10 \log(F_a/F_a^*)$ is less than 0.4 dB.

A similar approximation of F_a [Eq. (34)] has been made for typical aircraft noise.[8] When $r < 5000$ m, the error in this approximation ΔL is less than 1.5 dB. Under standard atmospheric conditions, the air absorption of aircraft noise is characterized by $\alpha = 2.5 \times 10^{-3}$ 1/m.

IV. NOISE REDUCTION AND ITS EFFICIENCY

In this section the explicit form of the propagation function $K(d)$ [Eq. (18)] and its decrement $\Delta K(d,\Phi)$ [Eq. (20)] will be derived. Then, the noise reduction ΔL [Eq. (31)] and noise reduction efficiency η [Eq. (32)] will be calculated.

A. Road traffic noise

If the perpendicular distance d from the receiver to the road (Fig. 1) is less than 100 m, then geometrical spreading is modified mainly by the ground effect. With $r=d/\cos\phi$ (Fig. 5) one gets [Eq. (33)]

$$F_g(\phi)=\frac{\beta\cos^2\phi}{[\cos^2\phi+g^2]}, \tag{35}$$

where the ground effect is described by

$$g^2=\gamma\frac{d^2}{(h+H)^2}. \tag{36}$$

Equations (18) and (35) combine into the propagation function

$$K_g=\frac{\beta d_o}{4\pi d}\int_{-\pi/2}^{+\pi/2}\frac{\cos^2\phi d\phi}{\cos^2\phi+g^2}, \tag{37}$$

and its change

$$\Delta K_g=-\frac{\beta d_o}{4\pi d}\left\{\frac{2\Phi\cos^3\Phi}{\cos^2\Phi+g^2}-\int_{-\Phi}^{+\Phi}\frac{\cos^2\phi d\phi}{\cos^2\phi+g^2}\right\}, \tag{38}$$

which estimate the result of rerouting. The closed form of the above integrals is[10]

$$K_g=\frac{\beta d_o}{4d}\left\{1-\frac{g}{\sqrt{1+g^2}}\right\} \tag{39}$$

and

$$\Delta K_g=\frac{\beta d_o}{2\pi d}\left\{\Phi\left[1-\frac{\cos^3\Phi}{\cos^2\Phi+g^2}\right]-\frac{g}{\sqrt{1+g^2}}\text{arc }tg\left(\frac{g}{\sqrt{1+g^2}}\tan\Phi\right)\right\}. \tag{40}$$

Accordingly, Eqs. (31), (36), (39), and (40) lead to continuous function $\Delta L(\Phi)$ noise reduction (Fig. 6) graphed for two perpendicular distances $d=10$ and 100 m (Fig. 1), with the ground coefficient $\gamma=0.008$ and the heights $h+H=2$ m (Fig. 3). As expected, the noise reduction ΔL increases with the angle Φ (Fig. 2), i.e., with the radius $R=d/\cos\Phi$ and with the perpendicular distance d. For example, with the distance $d=100$ m, the required noise reduction $\Delta L=5$ dB gives the angle $\Phi\approx1$ rad and the radius $R\approx185$ m and $x=\pm155$ m (Fig. 2). Plot in Fig. 7 shows the noise reduction efficiency η [Eqs. (32), (39), and (40)] for $2C=1$, $d=10$, 100 m, $h+H=2$ m, and $\gamma=0.008$. It is clear that the cost per 1 dB noise reduction η increases with an increased angle Φ, i.e., radius $R=d/\cos\Phi$ (Fig. 2).

Considering together the ground effect and air absorption [Eqs. (33) and (34)] with $r=d/\cos\phi$ (Fig. 5), one concludes that

$$F_g(\phi)F_a(\phi)=\frac{\beta\cos^3\phi}{[\cos^2\phi+g^2][\cos\phi+\alpha d]}, \tag{41}$$

where the parameter g is defined by Eq. (36). Mindful of Eqs. (18), (20), and (41) one arrives at the integrals for the propagation function

$$K_{ag}=\frac{\beta d_o}{4\pi d}\int_{-\pi/2}^{+\pi/2}\frac{\cos^3\phi d\phi}{[\cos^2\phi+g^2][\cos\phi+\alpha d]}, \tag{42}$$

and its change

$$\Delta K_{ag}=-\frac{\beta d_o}{4\pi d}\left\{\frac{2\Phi\cos^4\Phi}{[\cos^2\Phi+g^2][\cos\Phi+\alpha d]}-\int_{-\Phi}^{+\Phi}\frac{\cos^3\phi d\phi}{[\cos^2\phi+g^2][\cos\phi+\alpha d]}\right\}. \tag{43}$$

Ultimately, tables of integrals[10] allow getting the closed form of Eq. (42),

$$K_{ag}=\frac{\beta d_o}{4d}\left\{1-\frac{1}{g^2+a^2}\left[\frac{g^3}{\sqrt{1+g^2}}+\frac{4}{\pi}\frac{a^3}{\sqrt{a^2-1}}\text{arc }tg\sqrt{\frac{a-1}{a+1}}+\frac{a}{\pi}\frac{g^2}{\sqrt{1+g^2}}\ln\frac{\sqrt{1+g^2}+1}{\sqrt{1+g^2}-1}\right]\right\}, \tag{44}$$

where $a=\alpha d$. If air absorption is small ($\alpha d<1$), then

$$\frac{1}{\sqrt{a^2-1}}\text{arc }tg\sqrt{\frac{a-1}{a+1}}\to\frac{1}{\sqrt{1-a^2}}\text{arc }th\sqrt{\frac{1-a}{1+a}}. \tag{45}$$

Making use again of Ref. 10 one obtains [Eq. (43)]

$$\Delta K_{ag}=\frac{\beta d_o}{2\pi d}\left\{\Phi\cdot G_1(\Phi,a,g)-\frac{G_2(\Phi,a,b)}{g^2+a^2}\right\}, \tag{46}$$

where

$$G_1=1-\frac{\cos^4\Phi}{[\cos^2\Phi+g^2][\cos\Phi+a]}, \tag{47}$$

$$G_2=\frac{g^3}{\sqrt{1+g^2}}\text{arc }tg\left(\frac{g}{\sqrt{1+g^2}}tg\Phi\right)+\frac{2a^3}{\sqrt{a^2-1}}\text{arc }tg\left[\sqrt{\frac{a-1}{a+1}}tg\frac{\Phi}{2}\right]+\frac{a}{2}\frac{g^2}{\sqrt{1+g^2}}\ln\frac{\sqrt{1+g^2}+\sin\Phi}{\sqrt{1+g^2}-\sin\Phi}. \tag{48}$$

The noise reduction ΔL [Eqs. (31), (44), and (46)] is plotted in Fig. 8 without and with the air absorption taken into account ($\alpha=0$ and $\alpha=1.8\times10^{-3}$—standard atmospheric conditions). It seems that the air absorption does not play any important role compared with the ground effect.

B. Aircraft noise

In the free field of a directional point such as any aircraft so close to the receiver the air absorption can be neglected, Eq. (15) can be rewritten as follows:

$$\frac{p_{Ai}^2}{p_o^2} = \frac{W_{Ai}(V)}{W_o}\frac{d_o^2}{4\pi r^2}Q(\phi,\nu). \tag{49}$$

Longitudinal and lateral directivities are represented by the directivity function $Q(\phi,\nu)$.[11]

The *directivity coefficient* $0 \le \nu < 1$ depends on the power setting and the aircraft speed, and its value can be estimated from noise measurements. For any aircraft, the law of energy conservation yields identity

$$\iint_{4\pi} Q(\phi,\nu)d\Omega = 4\pi \tag{50}$$

where $d\Omega = 2\pi\cos\phi d\phi$ represents the differential of the solid angle.

In this study it is assumed that the elevation angle is relatively large $\Psi > 15°$ (Fig. 3). In Ref. 6 it has been shown that under such circumstances the ground effect, called sometimes as the "lateral attenuation," can be quantified by only one coefficient β [Eq. (33)], and the A-weighted sound power can be written as a products:

$$W_A \to \beta W_A. \tag{51}$$

The approximation of the real directivity function is

$$Q = \frac{q(\nu)}{1 - \nu\sin\phi}, \tag{52}$$

where the emission angle $-\pi/2 \le \phi \le +\pi/2$. Substituting Eq. (52) into Eq. (50) gives

$$q = \frac{2\nu}{\ln[(1+\nu)/(1-\nu)]}. \tag{53}$$

Now, Eqs. (49) and (51) yield the relative A-weighted squared sound pressure

$$\frac{p_{Ai}^2}{p_o^2} = \frac{\beta W_{Ai}(V)}{W_o}\frac{d_o^2}{4\pi r^2}Q(\phi,\nu)F_a(r), \tag{54}$$

where the air absorption function $F_a(r)$ is defined by Eq. (34) and directivity function by Eq. (52). With $r = d/\cos\phi$ (Fig. 5) one gets

$$\beta Q \cdot F_g = \frac{\beta q(\nu)\cdot\cos\phi}{[\cos\phi + \alpha d][1 - \nu\sin\phi]}, \tag{55}$$

and Eqs. (18), (20), and (55) lead to

$$K_a(\nu) = \frac{\beta d_o q(\nu)}{4\pi d}\int_{-\pi/2}^{+\pi/2}\frac{\cos\phi d\phi}{[\cos\phi + \alpha d][1 - \nu\sin\phi]} \tag{56}$$

and

$$\Delta K_a(\nu) = -\frac{\beta q d_o}{4\pi d}\left\{\frac{2\Phi\cos^2\Phi}{\cos\Phi + \alpha d} - \int_{-\Phi}^{+\Phi}\frac{\cos\phi d\phi}{[\cos\phi + \alpha d][1 - \nu\sin\phi]}\right\}. \tag{57}$$

For moving aircraft, the perpendicular distance d [Fig. 3 and Eq. (4)] becomes "the slant range at closest approach to the receiver."

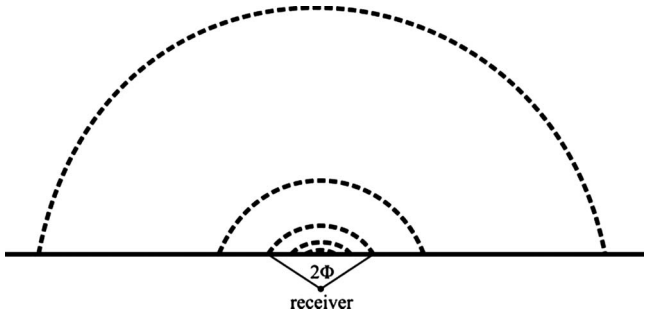


FIG. 10. The new routes, with angles $\Phi = 0.50$, 0.75, 1.00, 1.25, and 1.45 rad, used for INM calculations of noise reduction ΔL [Eq. (61)].

Finally, making use again of the tables of integrals[10] one arrives at

$$K_a(\nu) = \frac{\beta q d_o}{4d[1 - \nu^2 + \nu^2(\alpha d)^2]}\left\{\sqrt{1-\nu^2} + \frac{\nu\alpha d}{\pi}\ln\frac{1+\nu}{1-\nu} - \frac{4}{\pi}\frac{\alpha d}{\sqrt{(\alpha d)^2 - 1}}\arctan tg\left(\sqrt{\frac{\alpha d - 1}{\alpha d + 1}}\right)\right\} \tag{58}$$

and

$$\Delta K_a(\nu) = \frac{\beta q d_o}{4\pi d}\left\{G(\Phi,\alpha d,\nu) - \frac{2\Phi\cos^2\Phi}{\cos\Phi + \alpha d}\right\}, \tag{59}$$

where

$$G = \frac{1}{[1 - \nu^2 + \nu^2(\alpha d)^2]}\left\{2\sqrt{1-\nu^2}\arctan tg(\sqrt{1-\nu^2}tg\Phi) + \nu\alpha d\ln\frac{1 + \nu\sin\Phi}{1 - \nu\sin\Phi} + -\frac{4\alpha d}{\sqrt{(\alpha d)^2 - 1}}\arctan tg\left(\sqrt{\frac{\alpha d - 1}{\alpha d + 1}}tg\frac{\Phi}{2}\right)\right\}, \tag{60}$$

In order to quantify the noise reduction due to the rerouting of the aircraft track, we apply Eqs. (31), (58), and (59). The results of ΔL calculations are shown in Fig. 9. The dependence of ΔL on the directivity coefficient ν is clear. The required noise reduction, e.g., $\Delta L = 4$ dB, with $\nu = 0.1$ gives the angle $\Phi = 63°$ and the radius $R \approx 2200$ m (Fig. 2).

To validate the proposed method in this study, the INTEGRATED NOISE MODEL (INM) 7.0A has been used for the helicopter Sikorsky S-70 Blackhawk (UH-60 A) flying at the distance $d = 1000$ m (Fig. 1) (temperature of 14.6 °C, pressure of 759.97 mm Hg, and headwind of 14.8 km/h). The new routes, with the angles $\Phi = 0.50$, 0.75, 1.00, 1.25, and 1.45 rad, are shown in Fig. 10. The corresponding INM noise reduction (Fig. 11)

$$\Delta L = L_{AE} - \tilde{L}_{AE}, \tag{61}$$

is in consonance with the previously obtained results (Fig. 9).

V. CONCLUSIONS

Substitution of a circle arc for a straight path segment (Fig. 2) yields noise reduction ΔL in terms of the day-night

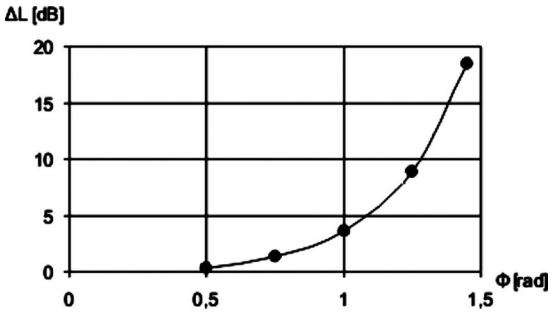


FIG. 11. Noise reduction ΔL [Eq. (61)] calculated with the integrated noise model 7.0a for the helicopter Sikorsky S-70 Blackhawk (UH-60 A) flying at the perpendicular distance $d=1000$ m (temperature of 14.6 °C, pressure of 759.97 mm Hg, and headwind of 14.8 km/h).

average sound level L_{dn} and the day-evening-night level L_{den} [Eq. (3)]. ΔL is a function of the angle Φ, i.e., the radius $R=d/\cos\Phi$ (Fig. 2). In order to determine ΔL [Eq. (31)], the values of the propagation function K [Eq. (18)] and its decrement ΔK [Eq. (20)] are needed. For road traffic noise, when the ground effect dominates, one can find K_g and ΔK_g from Eqs. (39) and (40), respectively. Taking into account both the air absorption and ground effect, the values of K_{ag} and ΔK_{ag} have to be calculated from Eqs. (44) and (46). Finally, for the aircraft noise generated by a directional source, the values of $K_a(\nu)$ and $\Delta K_a(\nu)$ yield Eqs. (58) and (59). The method proposed here was validated against the INM. It seems that the results are in agreement (Figs. 9 and 10).

To simplify the final results, it was assumed that the sound rays are straight or bent to the ground, so there is no shadow zone. It is typical of restless atmosphere on a clear night following a warm day. The results presented may have some use for road planning and aircraft noise reduction. In the former case, presumably, it would be cheaper than building noise barriers.

ACKNOWLEDGMENTS

The author wishes to thank the reviewers and editor for their thorough comments. The assistance from Mr. L. Lisek and Mr. M. Zoltowski is gratefully acknowledged.

[1]R. Makarewicz, "Noise reduction through source rerouting," J. Acoust. Soc. Am. **116**, 2735–2737 (2004).
[2]R. Makarewicz and M. Żółtowski, "Variations of road traffic noise in residential areas," J. Acoust. Soc. Am. **124**, 3568–3575 (2008).
[3]R. Makarewicz and R. Golebiewski, "Uncertainty of traffic noise prediction," Acta Acust. Acust. **92**, 578–582 (2006).
[4]R. Makarewicz and P. Kokowski, "Simplified model of ground effect," J. Acoust. Soc. Am. **101**, 372–376 (1997).
[5]K. Attenborough, K. M. Li, and K. Horoshenkov, *Predicting Outdoor Noise* (Taylor & Francis, London, 2007).
[6]H. Olsen, R. Ranberg, S. Storeheier, I. Granoien, and A. Ustad, "Validation of source distant propagation algorithms for aircraft noise prediction," in Inter-Noise 2002, Dearborn, MI (2002).
[7]M. E. House, "Measurement and prediction of aircraft noise," in *Transportation Noise- Reference Book*, edited by P. Nelson (Butterworths, London, 1987).
[8]R. Makarewicz, "Influence of air absorption and ground effect on aircraft noise," in Proceedings of the 3rd Joint Meeting of ASA and ASJ, Honolulu, HI (1996).
[9]R. Makarewicz, "Attenuation of outdoor noise due to air absorption and ground effect," Appl. Acoust. **53**, 133–151 (1998).
[10]I. S. Gradshteyn and I. M. Ryzhik, *Tables of Integrals Series and Products* (Academic, New York, 1980).
[11]A. D. Pierce, *Acoustics: An Introduction to Its Physical Principles and Applications* (Acoustical Society of America, New York, 1989).

Optimum absorption and aperture parameters for realistic coupled volume spaces determined from computational analysis and subjective testing results

David T. Bradley[a] and Lily M. Wang
Architectural Engineering Program, Peter Kiewit Institute, University of Nebraska–Lincoln, Omaha, Nebraska 68182-0681

(Received 16 June 2009; revised 22 October 2009; accepted 3 November 2009)

This project utilizes computational modeling to study the effects of varying two architectural parameters, absorption ratio and aperture size, in a realistic coupled volume concert hall. Coupled volumes have been shown to exhibit non-exponential sound energy decay profiles, referred to as double slope effect. A number of objective metrics (T30/T15, LDT/T10, decay ratio, and ΔL) have been used to quantify the double slope effect of the profiles generated in the virtual hall. T30/T15 and LDT/T10 showed similar trends across all hall configurations, indicating decreasing double slope effect with increasing coupled volume absorption ratio for each aperture size, and producing highest values at a specific aperture size for each absorption ratio. Generally, LDT/T10 provides finer resolution than T30/T15 when analyzing the decay profiles in this study. Results from the two metrics derived from Bayesian analysis, decay ratio and ΔL, seem less consistent. Subjective testing has also been conducted to determine the effect of varying the two architectural parameters in the hall, and multidimensional scaling analysis shows that, in general, listener preference is inversely proportional to the level of double slope effect, with the highest levels of preference occurring at low and medium levels of double slope effect. Recommended design guidelines for coupled volume halls are provided based on these computational and subjective results.
 [DOI: 10.1121/1.3268604]

PACS number(s): 43.55.Br, 43.55.Ka, 43.55.Hy, 43.55.Fw [ADP] Pages: 223–232

I. INTRODUCTION

Spaces exhibiting acoustically coupled volumes are typically composed of a highly reverberant secondary room connected to a main room via an acoustically transparent aperture. This geometrical configuration is of particular interest due to the potential for the creation of temporally disparate decay rates in the main room. This so-called non-exponential decay is atypical in single volume spaces, and can lead to unique acoustic conditions. In particular, coupled volume systems can be designed to produce a non-exponential decay that exhibits a steep initial decay followed by a long reverberant tail. This type of decay is assumed to produce high perceived levels of both clarity and reverberance, qualities that are typically desirable in performing arts spaces. The current work focuses on determining the acoustic effect of architectural parameters such as the amount of absorption in the coupled spaces and the size of the aperture opening between them. Particular attention is given to studying how the sound energy decay within coupled volume spaces varies in response to changing architectural parameters, and how human listeners respond to these variations. This research is an extension of previous work in which the authors examined coupled volumes consisting of two rectangular boxes connected to one another.[1] The current study explores a more complicated and realistic geometry such as those found in existing coupled volume performing arts spaces. Furthermore, psychoacoustics testing has been conducted to determine the preference of listeners for various decay profiles from the virtual coupled volume hall. The results are used to provide a design rubric for implementing coupled volume geometries in auditoria.

A detailed theoretical development of the sound energy decay in coupled volumes can be found in literature.[1–6] To summarize, for the particular case when a highly reverberant secondary volume is coupled with a less reverberant main volume, sound energy from the secondary space can feed back into the main space through the aperture opening between them. When the sound source and receiver are located in the main volume, this delayed feedback of sound energy can result in a decay profile with more than one slope. Decay profiles are typically plotted on a logarithmic scale, such that a single slope decay depicts an exponential decay. A double slope decay is referred to as non-exponential; a representative decay profile is shown in Fig. 1. This double slope decay can be represented by the two linear functions of time shown as solid and dashed bold lines in Fig. 1, such that[3]

$$L_1(t) = -\left(\frac{60}{T_1}\right)t, \tag{1}$$

$$L_2(t) = -\left(\frac{60}{T_2}\right)t - 10\log\left(\frac{A_{1S}A_{2S}}{S^2}\right), \tag{2}$$

where $A_{is}=A_i+S$, A_i is the equivalent absorption area of the ith space (inclusive of the aperture opening area), S is the

[a]Author to whom correspondence should be addressed. Present address: Physics+Astronomy Department, Vassar College, Poughkeepsie, New York 12604-0745. Electronic mail: dabradley@vassar.edu

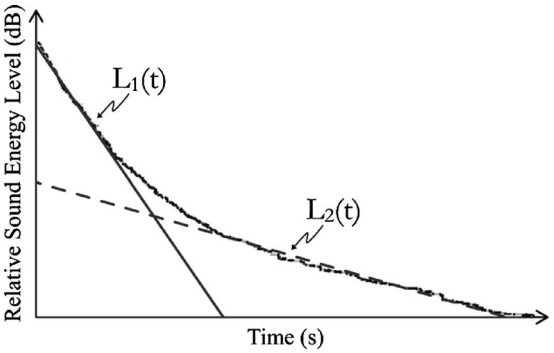


FIG. 1. A representative decay outline exhibiting a double slope, where $L_1(t)$ is shown by the solid line and $L_2(t)$ by the dashed line.

surface area of the aperture opening, and T_i is the decay time for room i alone. Equations (1) and (2) are in the form of $y=-mx+b$. This allows for the decay slope of the main volume, when coupled, to be analyzed based on its two temporal components, early decay ($L_1(t)$) and late decay ($L_2(t)$). Each portion of the decay is given by a line with a slope and y-intercept that can be calculated from architectural parameters of the coupled volume system.[1] Double slope effect (DSE) will be used in this paper to refer to the phenomenon when the second slope is larger than the first, and dominates the decay profile during the late portion (as shown in Fig. 1).

For this investigation, a coupled volume concert hall was designed based on an amalgamation of characteristics from existing coupled volume halls. The design of this imagined hall was then implemented as a computer model using computer aided drafting (CAD) software. An acoustics computer modeling program, ODEON (version 6.5), was used to predict room impulse responses (RIRs) for several configurations of the virtual hall computer model. The RIRs were studied to determine the effect of changes in the architectural parameters on sound decay in the virtual hall. Then, the response of listeners to the sound fields created in the computer models was determined from subjective testing and subsequent statistical analysis.

II. PREVIOUS RESEARCH

Several investigations have been conducted on the acoustics of coupled volume systems, using both physical and computational models based on geometrical acoustics.[7–14] Of particular note, Nijs *et al.*[9] compared scale and computational model results of coupled volume spaces and listed several best practice techniques for increasing the fidelity of the computational simulations. These techniques have been implemented in the computational work described in this paper. Also, Bradley and Wang[14] validated the use of ODEON (version 6.5) for such studies by comparing simulated and measured impulse responses in a coupled volume concert hall and finding good correlations in the data.

In some of this previous work, metrics have been proposed for quantifying DSE in coupled volume halls. Harrison and Madaras[10] used the ratio T30/T15 to give the relationship between two portions of the reverberant decay in the main room. T30 is defined as the decay time with sound pressure level (SPL) limits of −5 and −35 dB in the energy decay function, multiplied by a factor of 2.[15] Similarly, T15 is given as the decay time from −5 to −20 dB, multiplied by a factor of 4. T30/T15 gives the relationship between the overall decay curve and the early portion of the decay, therefore providing an indication of the relationship between the two slopes in a non-exponential decay. Harrison and Madaras[10] found that T30/T15 values were highest for smaller aperture opening sizes, and increased with larger coupled volume sizes. Ermann[13] used a similar ratio of T60/T15 in his work, wherein he studied the effect of aperture size and coupled volume absorption on DSE using numerical and CATT ACOUSTIC (version 8.0b) models of an imagined coupled volume concert hall. His work showed that having an average absorption coefficient ($\bar{\alpha}$) in the coupled space less than 0.02 ($\bar{\alpha}<0.02$) produced significantly higher T60/T15 values than $0.02<\bar{\alpha}<0.05$. Also T60/T15 values were largest when the aperture size was relatively small (1% of the total main volume surface area).

While these DSE quantifiers, T30/T15 and T60/T15, can describe general trends in DSE, they do not offer a method for differentiating between different double slope profiles. Particularly, specific T30/T15 and T60/T15 values can be used to describe several different double slope profiles. Furthermore, the decay measures in the numerator and denominator of each ratio may not be a good linear fit of the decay curve over the range covered, which can misrepresent the two slopes in the decay. The two decay measure ranges also overlap, which can obfuscate the double slope nature of the decay. These ambiguities compromise the effectiveness of these descriptors as quantifiers of DSE. Bradley and Wang[1] tested two quantifiers that can theoretically identify DSE profiles more uniquely: the decay ratio (T_2/T_1) and ΔL. ΔL is defined as the difference in decibels between the two y-intercepts in the DSE profile (referred to as ΔdB in previous studies), calculated from data produced by a Bayesian analysis algorithm developed by Xiang *et al.*[5,6,16] In this investigation, decay ratio and ΔL showed trends similar to T30/T15, except for cases with the largest coupled volume size in which the results were harder to interpret.[1]

In another study, Bradley and Wang[17] analyzed computational models of an existing coupled volume concert hall, and compared the DSE quantifiers listed above as well as a newly proposed quantifier, late decay time/early decay time (LDT/EDT). EDT is defined as the decay time from 0 to −10 dB, multiplied by a factor of 6.[15] LDT was proposed by the authors as a measure of the late portion of the decay. This decay measure is defined as the decay time with SPL limits of −25 and −35 dB in the energy decay function, multiplied by a factor of 6. This SPL range was chosen to correspond with the bottom of the established T30 range, which ensures that the decay values are above the noise floor for measurements taken according to ISO 3382-1.[15] The LDT/EDT ratio, unlike T30/T15, takes the ratio of two distinct portions of the decay, and more closely approximates the ratio of the two slopes found in a double slope decay than ratios that cover overlapping decay ranges. This distinction makes LDT/EDT likely to quantify DSE more accurately than T30/T15 or T60/T15. Indeed, Bradley and Wang found that LDT/EDT seemed to differentiate DSE with more resolution than the

other ratios. The two metrics calculated from Bayesian analysis results, decay ratio and ΔL, surprisingly did not indicate the presence of a double-sloped decay curve in this study, though.

Upon subsequent analysis, the authors have replaced LDT/EDT with a revised metric, LDT/T10, for the work described in this paper; T10 is defined as the decay time from -5 to -15 dB, multiplied by a factor of 6. One can expect that LDT/T10 should resolve to a value of 1 when the decay curve is exponential. This may not necessarily occur with LDT/EDT since EDT is much more influenced by early reflection patterns than T10.

None of the metrics introduced so far has been previously linked to subjective preference in a suitable manner. Some previous subjective studies have focused on perception of exponential decays versus non-exponential decays generated by splicing together signals from computer-simulated single-sloped decays.[18,19] Picard's results, in particular, showed that subjects could more easily perceive a difference between the decays when the y-intercepts of the two slopes of the decay were minimized (i.e., smaller ΔL).[19] Additionally, subjects were more likely to recognize a difference between two non-exponential decay curves when the difference between the first and second slopes in the non-exponential decay was increased (i.e., larger decay ratio T_2/T_1).

Bradley and Wang[1] conducted a subjective study using a range of decay curves generated from simplistic representations of coupled volume performance spaces, created from two rectangular boxes connected to one another. This study sought to determine if varying degrees of double slope effect were perceived to have greater or less reverberation and/or clarity. Results showed that increasing the coupled volume size and the aperture size correlated with higher perceived reverberance, which matched objective T30/T15 results as well as possibly higher decay ratios and lower ΔL values. No statistically significant results on perceived clarity were found, though, in part due to the simplistic geometry used. Note that subjective preference between the decay curves was not investigated.

More recently, Ermann[20] conducted subjective testing using exponential and non-exponential decays created from computational models of coupled volume spaces in an attempt to determine preference of double slope versus single slope decays. The results from this investigation agreed with Picard's results, showing that difference perception increased as the difference between the two slopes in the non-exponential decays was increased. However, no conclusion on preference could be made, as the sample size was small and many subjects could not correctly identify when two identical stimuli were presented in a paired comparison, indicating a problem with the reliability of the testing.

The research described in this paper advances the field of coupled volume room design by providing statistically significant data on subjective preference concerning coupled volume decays while linking that preference to objective measures of DSE. A computational model of a realistic coupled volume geometry has been systematically varied in terms of the absorption in the coupled spaces and the aperture opening between the spaces. The resulting room impulse responses have been analyzed for the degree of DSE, according to the metrics discussed above. Auralizations of these impulse responses have then been generated and used in subjective testing to determine relative preference of the decay curves. This work provides a comprehensive comparative analysis of objective and subjective data from these types of spaces, concluding with architectural design recommendations for achieving maximum subjective preference.

TABLE I. Architectural parameters and calculated volume ratio of the complex coupled volume system concert hall.

Volume (m^3) [main]	22 145
Total surface area (m^2) [main]	5 140
Average absorption coefficient α [main]	0.39
Equivalent absorption area $S\alpha$ [main]	2 005
Volume (m^3) [coupled]	10 444
Total surface area (m^2) [coupled]	5 075
Volume ratio	47%

III. COMPUTATIONAL ANALYSIS

A. Methodology

The design for the virtual coupled volume concert hall in this study was based on a combination of architectural features from several existing concert halls with coupled volumes incorporated in their design, including Culture and Congress Centre Concert Hall (Lucerne, Switzerland), Theatres on the Bay Concert Hall (Esplanade, Singapore), McDermott Concert Hall (Dallas, TX), and Symphony Hall (Birmingham, UK). The pertinent architectural parameters and the volume ratio, calculated by dividing the main volume size by the secondary volume size, of the virtual hall are listed in Table I. Since previous work by the authors found that larger coupled volume sizes were more likely to generate double slope effect across different combinations of absorption characteristics and aperture opening sizes,[1] a larger volume ratio from among the range in the existing halls surveyed was utilized in this investigation.

The shoebox-shaped main volume and the secondary volume of this virtual hall are connected through 236 aperture doors as shown in Fig. 2. The coupled volume extends the entire height of the hall on both sides of the stage and audience area, and is shown as light gray shading in Fig. 2. The doors are located on three levels in the hall, as illustrated by the dark gray shading in Fig. 2. The hall has an organ behind the stage, an acoustical canopy array hanging over the stage, and audience seating consisting of a main orchestra level and two stacked balcony levels.

Two architectural parameters were varied to create several configurations of the hall, as was done in previous work by the authors.[1] The first architectural parameter is referred to in this paper as the absorption ratio, which quantifies the equivalent absorption area in the secondary volume as a portion of the equivalent absorption area in the main volume

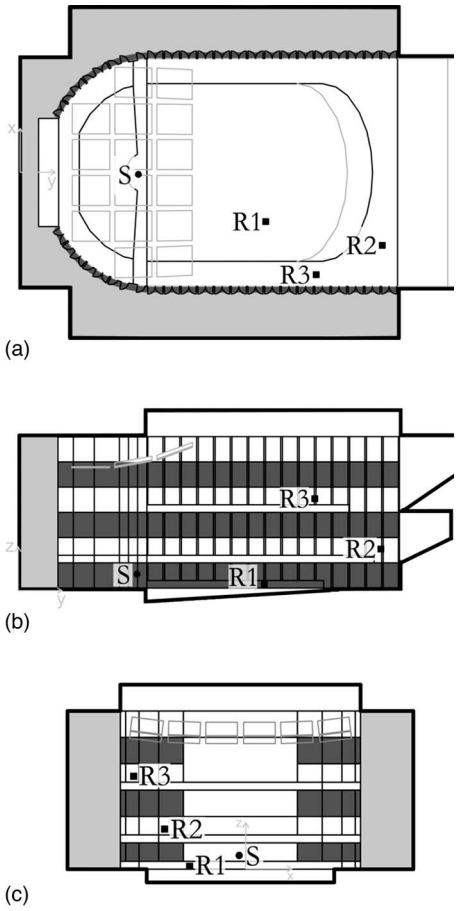


FIG. 2. (a) Plan view, (b) cross section, and (c) transverse cross section of the virtual hall with the secondary volume shaded in light gray and aperture doors in dark gray. Source position (s) is designated with a circle and receivers (R1–R3) are marked by squares.

$$\text{absorption ratio} = \frac{\sum_i (\alpha_i S_i)_{\text{secondary}}}{\sum_j (\alpha_j S_j)_{\text{main}}}, \tag{3}$$

where the i and j summation counters represent the individual surfaces in the secondary and main volumes, respectively. The second parameter is aperture opening size expressed as a percentage of the total surface area of the main volume

$$\text{aperture opening} = \frac{S_{\text{aperture}}}{\sum_j (S_j)_{\text{main}}} \times 100\%. \tag{4}$$

The aperture opening, as defined above, is normalized by the main volume surface area so that comparison across various coupled volume systems accounts for different volume sizes.

TABLE II. Four levels of absorption ratio used in the virtual hall configurations.

Level	Absorption ratio
1	0.04
2	0.08
3	0.12
4	0.16

TABLE III. Seven levels of aperture size used in the virtual hall configurations.

Level	% of main volume surface area
0	0
1	0.5
2	1.0
3	1.5
4	3.0
5	6.5
6	10.0

Varying these two parameters can be expected to vary the shape of the double slope profile, as apparent from equations for decay ratio and ΔL, presented by Bradley and Wang.[1]

The ranges of values for these architectural parameters were based on results from previous research and those typical in existing coupled volume halls. The parameter value ranges were also chosen to achieve a broad spectrum of DSE values. As such, four levels of absorption ratio were chosen, ranging from 0.04 to 0.16, as shown in Table II. Additionally, seven levels of aperture size were chosen, ranging from 0% to 10% of the main volume surface area, as shown in Table III. The combination of these architectural parameter levels produced 25 configurations: 1 single volume system and 24 coupled volume systems.

Each configuration of the virtual hall was created as a CAD model and then imported into ODEON (version 6.5). These computer models were analyzed using 200 000 rays from an omni-directional source with an impulse response length of 4500 ms. Most surfaces were assigned a scattering coefficient of 0.3; surfaces likely to produce more nonspecular reflections, such as the canopy array and the organ, were given a scattering coefficient of 0.7. These scattering coefficient values are based on the recommendations given for the ODEON program,[21,22] and are similar to those used by the authors in a previous validation study.[14]

One source position and three receiver positions were studied in the models. The receivers were distributed evenly throughout one-half of the hall, taking advantage of symmetry. The source and receiver positions are shown as circles and squares, respectively, in Fig. 2. The coordinates of these positions are given in Table IV, with reference made to the origin shown in Fig. 2.

B. Results

An impulse response was generated for each of the 25 hall configurations for each source-receiver combination.

TABLE IV. Source and receiver coordinates for complex coupled volume system computer model configurations.

	x (m)	Y (m)	z (m)
Source	14.5	0	2
Receiver 1	30	−6	0.6
Receiver 2	44	−8	4.4
Receiver 3	35	−12.3	10.5

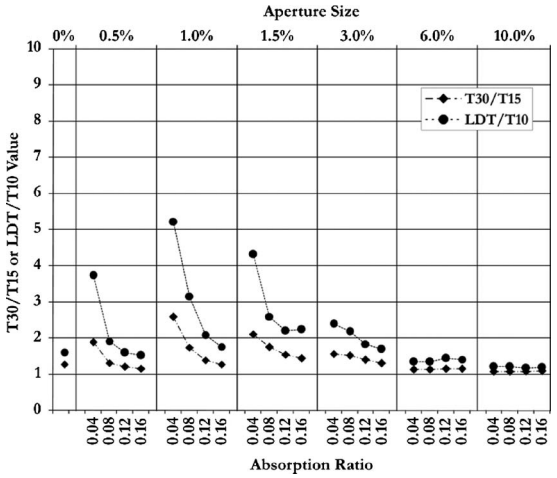


FIG. 3. T30/T15 and LDT/T10 data for the 25 virtual hall configurations

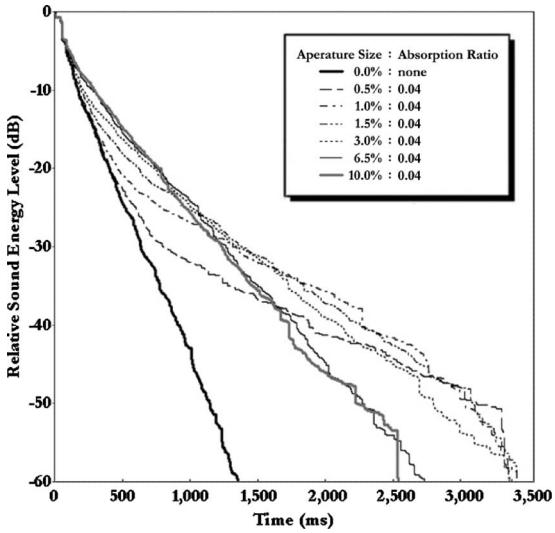


FIG. 4. Plot of energy decay curves for absorption ratio 0.04 across seven aperture sizes.

The impulse responses were exported from ODEON and analyzed to obtain several reverberation decay measures at the 1000 Hz octave band, including T15, T30, T10, and LDT. All decay measures were calculated from a least-squares fit line as prescribed in the ISO 3382-1 standard.[15] The average difference between the receiver positions was taken across all configurations for each decay measure. The overall average difference across receiver positions for each decay measure was relatively small: less than 7%. Therefore, only the results from receiver 1 will be discussed further in this paper.

Two DSE quantifiers, T30/T15 and LDT/T10, were calculated for the 25 virtual hall configurations, and are shown in Fig. 3. For all configurations, the T30/T15 and LDT/T10 data show similar trends for the varying architectural parameters. These DSE quantifiers decrease with increasing absorption ratio for each aperture size. The data across aperture size indicate that the two smaller absorption ratios (0.04 and 0.08) behave similarly, with maximum values for the DSE quantifiers found at aperture size 1.0%. The two larger absorption ratios (0.12 and 0.16) behave similarly, with the largest values found at aperture size 1.5%. In general, the DSE values are much lower for the two larger aperture sizes (6.0% and 10.0%). The DSE values for large aperture sizes closely match those for the single volume condition (0%), indicating that the decays from these configurations behave like single slope decay.

The energy decay curves corroborate the conclusions drawn from the data for these two DSE quantifiers. Figure 4 shows a representative example of the decay curves for seven aperture size levels with the absorption ratio held constant at 0.04. The single volume case (0%) has a single slope decay, while the first four coupled configurations (0.5%, 1.0%, 1.5%, and 3.0%) exhibit a distinct double slope in their decay. However, the two largest aperture sizes (6.0% and 10.0%) have decays with less evident double slope character. In fact, the decay from aperture size 10.0% can be approximated as a single slope decay with a decay time much larger than the single volume case (0%). In general, the larger aperture sizes produce decays with a single slope, as would be produced by a larger single volume space with a blend of acoustic characteristics from the main and secondary volumes.

Although the general trends of these two DSE quantifiers are similar, the way in which T30/T15 and LDT/T10 quantify DSE is different, as shown in Fig. 3. The LDT/T10 values are higher than those for T30/T15, and show larger changes in DSE across configuration. The LDT/T10 values for the coupled configurations are higher than the single volume values. However, the T30/T15 values for aperture sizes 6.0% and 10.0% are lower than the single volume (0%) value. The difference between the results of the two quantifiers suggests that LDT/T10 is able to quantify DSE more accurately and with a finer resolution than T30/T15.

The results from these two DSE quantifiers match well with a new quantity given in ISO 3382-2, the degree of curvature, which is used to determine the level of non-linearity of a decay curve.[23] This metric is calculated by the ratio of T20 over T30 given as a percentage deviation from a perfectly exponential decay. Values higher than 10% are said to indicate a decay curve that significantly deviates from exponential decay. The four coupled configurations with the smallest aperture sizes (0.5%, 1.0%, 1.5%, and 3.0%) all have degrees of curvature higher than 10%, indicating a significantly non-linear decay, which corresponds with the T30/T15 and LDT/T10 quantifier results.

Two metrics, decay ratio and ΔL, were also calculated for all the impulse responses filtered at the 1000 Hz octave band. These values were provided to the authors by Dr. Ning Xiang and his research group at Rensselaer Polytechnic Institute, by using their proprietary Bayesian analysis algorithm for room acoustics. The values calculated at receiver 1 are shown in Fig. 5. As found in previous studies,[1,17] the decay ratio and ΔL data do not match well with the other DSE quantifiers. Only 5 of the 16 configurations that T30/T15, LDT/T10, and the degree of curvature data indicate as being double slope are shown as double slope by decay ratio and ΔL. Furthermore, decay ratio and ΔL values for the two largest aperture sizes (6.0% and 10.0%) show significant

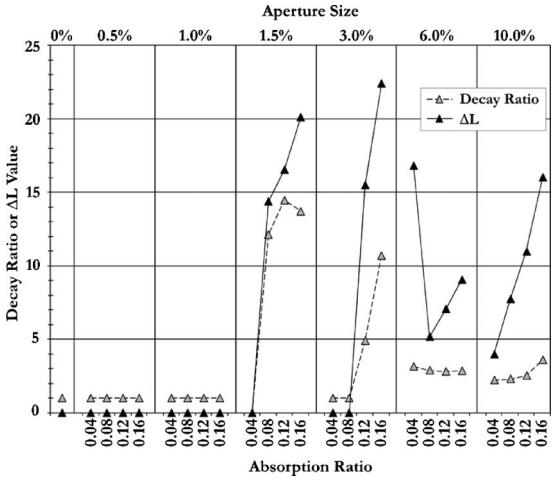


FIG. 5. Decay ratio and ΔL data for the 25 virtual hall configurations. Decay ratio equals 1 and ΔL equals 0 when only one slope has been detected by the Bayesian analysis.

DSE. By contrast the T30/T15 and LDT/T10 data for these aperture sizes showed single slope behavior, and the degree of curvature values were all less than 10%. Theoretically, decay ratio and ΔL should most uniquely describe these double slope decay curves, but the authors have found that, in practice, the results are not always as expected. Further study is certainly required in calculating decay ratio and ΔL from Bayesian analysis of decay curves generated in realistic coupled volume spaces to determine the source of this inconsistency.

Among the DSE quantifiers reviewed in this paper, LDT/T10 is the one that produces the most comprehensible results while discerning the presence of DSE more finely than the other decay time ratios. Consequently, the remainder of this paper will use the LDT/T10 ratio as the primary quantifier for DSE.

IV. SUBJECTIVE ANALYSIS

A. Methodology

In addition to the computational analysis, listener perception of the double slope decay profiles resulting from varying the architectural parameters of aperture size and absorption ratio was determined. Impulse responses from a subset of the virtual hall configurations were auralized. The results from the computational analysis were used to inform the configuration choice. In particular, configurations that produced disparate DSE results according to LDT/T10 were chosen, and configurations that produced similar DSE data were eliminated from the subjective testing subset. The data in Fig. 3 show similar DSE trends across absorption ratio within two groups of aperture sizes: group 1, consisting of aperture sizes 1.5% and 3.0%, and group 2, consisting of sizes 6.0% and 10.0%. One aperture size from each group was chosen for the subjective testing subset: aperture size 3.0% from group 1 and aperture size 10.0% from group 2. The remaining configurations were retained for the subjective testing, resulting in 17 configurations, a combination of 4 absorption ratios and 4 aperture sizes (0.5%, 1.0%, 3.0%, and 10.0%) plus the single volume case (0%). To clarify the presentation of data from the subjective testing, these 17 configurations will be referred to using an alpha-numeric labeling system as given in Table V.

TABLE V. Architectural parameter values of the 17 configurations used in the subjective testing with corresponding alpha-numeric labeling system.

Aperture size (%)		Absorption coefficient 0.04 *a*	0.08 *b*	0.12 *c*	0.16 *d*
0	00	…	…	…	…
0.5	1	1*a*	1*b*	1*c*	1*d*
1.0	2	2*a*	2*b*	2*c*	2*d*
3.0	4	4*a*	4*b*	4*c*	4*d*
10.0	6	6*a*	6*b*	6*c*	6*d*

ODEON (version 6.5) was used to produce auralizations of the 17 configurations. Binaural RIRs were convolved with an anechoic music signal provided with ODEON (Bach's Toccata et Fuga in D Minor), and a KEMAR head related transfer function (HRTF) was applied to each convolution.[24] The Bach piece was chosen because of its wide pitch range, which excited the virtual hall at a variety of frequencies. Additionally, the 15 s sample had two runs of short notes played in quick succession with decreasing pitches, each concluding with a stop chord, allowing the listener to experience both stopped and running reverberation in the virtual space.

The convolution process resulted in 17 auralized sound tracks corresponding to the 17 configurations. The tracks were presented in a paired comparison test over electrostatic Sennheiser HE-60 headphones to 30 human test subjects. The paired comparison method is a form of psychometric scaling used to order stimuli along a given dimension (e.g., preference), based on a series of independent judgments made between all possible pairs of items.[25] Subjects were asked to make a forced choice preference for each of the 136 pairs of auralized sound tracks. Specifically, for each paired comparison, the participant was told to "please listen to the audio samples and then answer the question: which track do you prefer?" The participant would then choose the first track or the second track using a radio button interface. Forced choice paired comparisons are used to create a binary response from the subject: Either track A is preferred or track B is preferred. Previous research by Parizet *et al.*[26] examined the differences among several listening test methods; they showed that forced choice comparisons had the greatest range in scores, magnifying the size of the effect for subjective response to acoustic stimuli. Additionally, paired comparison tests were found to provide a higher discrimination power than tests in which subjects rate sounds individually.

In the current study, each subject was presented with the 136 aural stimuli in three sessions. Each session lasted approximately 1 h, with a rest period of at least 2 h between sessions. The presentation order, both within pair and absolute order, was randomized for each participant to avoid bias error.

Subjects were members of the University of Nebraska community, ranging in age from 19 to 30 years, with an

equal number of males and females. They all indicated that they had some experience listening to classical music in performing arts spaces. Only subjects with hearing thresholds lower than 25 dB hearing level (HL) across 125–8000 Hz in both ears were allowed to participate in the testing.

Prior to testing, the participants underwent a training session in which the general concepts of coupled volume systems and DSE were outlined. No indication was made regarding the types of acoustic conditions that should be preferred. Instead, the training module simply focused the participants on what they could expect to hear changing in the samples, and then the subjects were expected to base preference on their own internal scale that later could be mapped to external variables through subsequent multidimensional scaling (MDS) analysis. Upon completion of the training module, the subjects were presented with a graphical user interface (GUI), which directed them to choose their preferred track from each pair.

B. Results

The data from the subjective testing described above were analyzed using MDS analysis. This method is described below along with the results from the analysis. First, a discussion of the reliability of the subjective testing data is presented.

1. Reliability of paired comparisons: Circular triads

When more than three stimuli are presented to a subject, a circular triad (CT) response can occur wherein a valid preference hierarchy cannot be established. A paired comparison study resulting in a high rate of CTs suggests that subjects find it difficult to respond to the stimuli regarding the attribute in question (e.g., preference).[25] This difficulty may be due to a high level of similarity among the stimuli, or the inappropriateness of the scale being used. Another possible cause for a high number of CTs is that the individual subject is unreliable or inconsistent in their responses. The consistency of individual subjects can be determined by Kendall's coefficient of consistency (ξ), a metric used to determine if a participant has significantly fewer CTs than is expected by chance.[27] Kendall and Babington-Smith showed that the maximum number of circular triads for an odd number of stimuli is given by

$$\mathrm{CT}_{\max} = \frac{(N^3 - N)}{24}, \tag{5}$$

where N is the number of stimuli. The coefficient of consistency is calculated using

$$\xi = 1 - \frac{d}{\mathrm{CT}_{\max}}, \tag{6}$$

where d is the number of CTs produced by a subject. If a subject were to produce the maximum number of CTs, $\xi=0$, while with no CTs, $\xi=1$.

The probability of a given ξ value based on the assumption that a subject's responses are completely random can be tested using a chi-square (χ^2) distribution. For a particular d, the p-value associated with the χ^2 distribution indicates the probability of obtaining a number of CTs equal to or greater than d based on chance, where a typical significance value of $p \leq 0.05$ is chosen for this type of test.[28] The degree of consistency among *all* subjects can be represented by Kendall's coefficient of agreement (u).[27] The development of u is similar to that of ξ, except that summations are made over the responses from all subjects. The coefficient ranges from 0 to 1, with higher coefficient values indicating higher agreement among participants. A χ^2 distribution can also be calculated for u, along with a corresponding probability. A significant p-value ($p \leq 0.05$) for the coefficient is indicative of at least a minimal level of agreement among subjects.[25]

2. Circular triad results

A circular triad analysis was conducted on the paired comparison responses of all 30 subjects. The resulting coefficient of agreement was $u=0.0463$ ($p=0.136$), indicating that the agreement among subjects did not reach the minimum level. An investigation of the individual subjects' ξ p-values revealed that the non-significance of u can be attributed to the unreliability of certain subjects. Ten subjects produced a higher number of CTs than expected by chance ($p > 0.05$). However, the majority of participants (66%) produced lower CT rates than expected by chance ($p \leq 0.05$). The fact that most of the subjects were able to respond with a high coefficient of consistency suggests that individual participant unreliability, rather than a high similarity among stimuli or inappropriate scale usage, was the cause of the low agreement among subjects. Therefore, the "unreliable" subjects can be removed to increase the significance of u.[25]

The ten subjects whose ξ p-value exceeded 0.05 were removed from the data set and a circular triad analysis was conducted once again. This modified subject set produced a $u=0.1117$ ($p<0.05$), indicating at least a minimum level of agreement among the participants. The data from the 20 "reliable" subjects were used in the MDS analysis discussed below.

3. Multidimensional scaling

MDS is a technique used to represent similarities between objects (or stimuli) as distances in a Euclidean space.[29] The more similar two objects are, the smaller the distance between them. While MDS assembles the Euclidean distance structure of the data, the researcher must assess the meaning for directions in that space, such as interpretation of the axes.[29]

MDS is an iterative process in which the accuracy of the statistical model is quantified by the output parameter *stress*, which is a numerical measure of the "badness-of-fit" between the current configuration and the input data.[30] The MDS process is repeated until the difference between the calculated coordinates and the input data is minimized according to a predetermined stop criterion, such as a minimum stress level. *Stress* can also be used to determine the number of dimensions of the solution. In general, the stress level will decrease as the number of dimensions is increased. However, increasing the number of dimensions in the MDS solution can lead to visual obscurity and difficulty in inter-

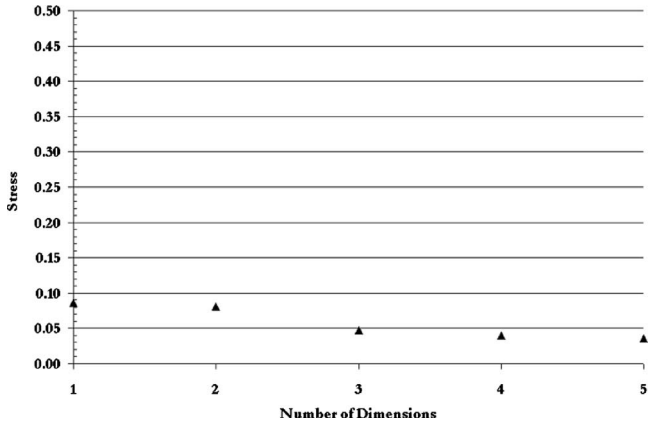


FIG. 6. A plot of number of dimensions versus stress. The stress value for dimension 1 is below 0.15, with a small amount of change in stress as number of dimensions is increased, indicating dimension 1 is the correct dimension.

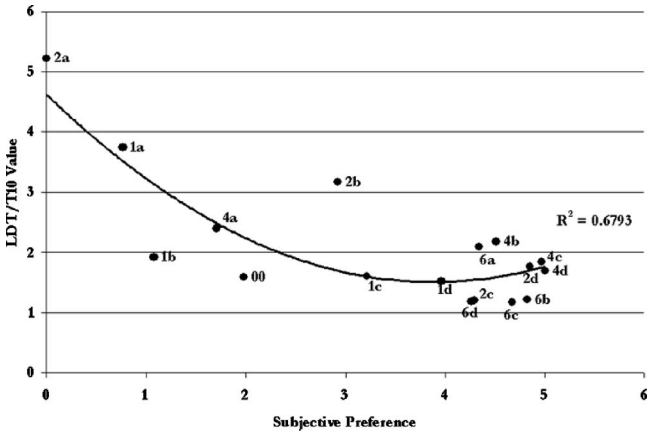


FIG. 7. Plot of relative subjective preference versus LDT/T10 data with a second-order polynomial trendline and correlation value. Higher subjective preference values indicate greater preference by listeners.

preting the coordinate axes; therefore, the Euclidean space with the fewest dimensions that best approximates the data is typically chosen as the final dimensional space for the solution.[30] MDS is particularly useful in representing data describing stimuli whose categorization dimensions are not well understood.[30] The limited body of research on subjective impression of DSE and the imprecise nature of most of the DSE quantifiers make MDS well suited for analyzing the psychoacoustic preference data to sound fields in coupled volume spaces.

4. Multidimensional scaling results

Each subject's data set was organized into a two-dimensional $N \times N$ half matrix, where N is the number of auralized virtual hall configurations ($N=17$). Twenty matrices were generated, 1 for each of the 20 reliable subjects. Individual preference MDS analysis was conducted for one- to five-dimensional solutions. A true statistical method for determining which dimensionality yields the "correct" solution does not exist. However, since MDS is being used here as a descriptive model, considerations such as interpretability, ease of use, and stability can be used to determine an appropriate dimension value.[29] Additionally, each dimensional solution yields a unique final stress value, which can be used to determine the solution that provides the best interpretation of the data. The stress values for each of the five solutions found through MDS analysis are shown in Fig. 6.

If the stress for dimension 1 is below 0.15, then this strongly suggests that the one-dimensional solution is the most appropriate choice.[29] The stress for dimension 1 in this case is 0.086, indicating that the data should be interpreted in this dimension. Additionally, Fig. 6 shows little improvement in stress values with increasing number of dimensions, suggesting that the one-dimensional solution yields a good approximation of the data.

The MDS relative preference values, normalized to range from zero to six, are plotted against the LDT/T10 values in an x-y scatter plot in Fig. 7. The configurations on the low end of the preference scale exhibit high levels of DSE (e.g., 1*a* and 2*a*), while on the high end, the DSE levels are in the mid-range (e.g., 4*c* and 4*d*). The lowest DSE values are located at the upper end of the middle of the preference scale. A second-order polynomial trendline corresponds with this increase in preference with decreasing DSE, and a slight peak in preference for medium levels of DSE. Also note that many of the coupled volume scenarios are preferred over the single main volume case (00).

The effect of variation in the virtual hall's architectural parameters on the MDS data set can be seen in Fig. 8, in which relative preference and DSE are plotted across configurations. The preference values are again normalized to range from zero to six so that they match the range of LDT/T10 values. This normalization is acceptable since the MDS results have no inherent range (that is, the values are relative). The y-axis of Fig. 8 therefore has both the LDT/T10 and relative preference values shown together. In general, subjective preference increases with increasing absorption ratio, except for the aperture size level 6 (10.0%) case. For this aperture size, preference peaks for absorption ratio b (0.08). However, the changes in the MDS results and the LDT/T10 data for aperture level 6 are relatively small. Preference levels are largest for aperture level 4 (3.0%). In general, the preference values are inversely proportional to DSE

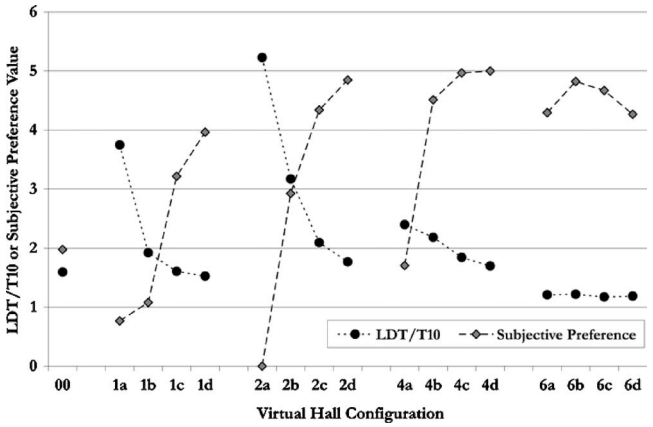


FIG. 8. Plot of LDT/T10 data versus relative MDS preference values for 17 configurations of virtual hall. Higher subjective preference values indicate greater preference by listeners.

TABLE VI. Four levels of preference with the corresponding virtual hall configurations and architectural parameters

Preference level	Configuration	Volume ratio (%)	Aperture size (%)	Absorption ratio
1	4*d*	47	3.0	0.16
2	4*c*	47	3.0	0.12
3	2*d*	47	1.0	0.16
4	6*b*	47	10.0	0.08

level as quantified by LDT/T10. Overall, the data suggest that hall configurations producing low and medium levels of DSE are most preferred by listeners, with the highest preference occurring at medium sized aperture openings (3% of the total main volume surface area) with larger absorption ratios (0.16). The lowest preference value is for configuration 2*a* (aperture size 1.0%: absorption ratio 0.04), which has the highest DSE level.

To further elucidate the effect of architectural variation on listener preference, the virtual hall configurations for the four highest levels of preference are listed in Table VI. The data from Table VI were used to develop the flow chart of design guidelines in Fig. 9(a). Each of the configurations in Table VI has a relatively large volume ratio (47%). However, previous results have indicated that DSE decreases with decreasing volume ratio.[1,10] Therefore, to generalize the design criteria for additional volume ratios, the other two architectural parameters have been adjusted accordingly in Figs. 9(b) and 9(c), which show suggested design guidelines for medium and small volume ratios.

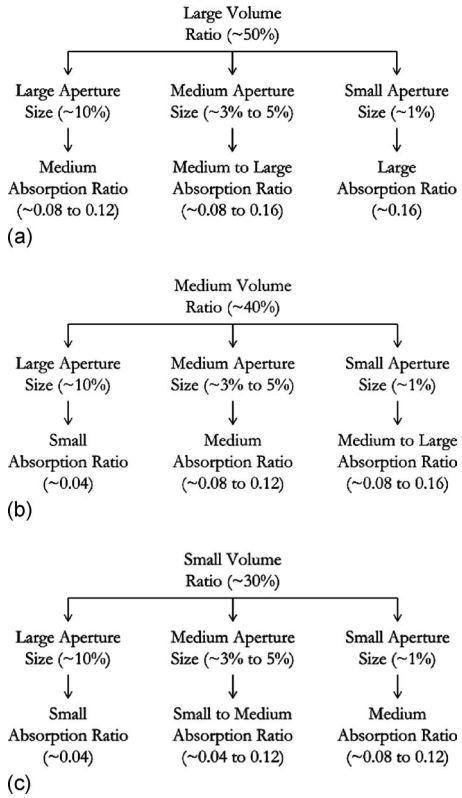


FIG. 9. Recommended design guideline flow charts for coupled volume concert halls with (a) large volume ratio, (b) medium volume ratio, and (c) small volume ratio.

The lowest level of preference from the subjective testing of the virtual hall was associated with the highest level of DSE, even lower than preference for the single main volume case (00). This level of DSE was produced by a configuration with a small aperture size (0.5%) and a small absorption ratio (0.04). Therefore, these architectural parameter configurations, and others that would produce high levels of DSE as quantified by LDT/T10 values greater than 3, should be avoided in coupled volume design.

V. CONCLUSIONS AND FUTURE WORK

A virtual coupled volume concert hall was studied using computational modeling, in which two architectural parameters were varied: absorption ratio and aperture size. Four levels of absorption ratio and seven levels of aperture size were analyzed. Two decay ratios, T30/T15 and LDT/T10, were used to quantify the DSE, and the data from the analysis of the virtual hall show similar trends across the various configurations for both quantifiers. DSE decreases with increasing absorption ratio for each aperture size. DSE is highest at a specific aperture size for each absorption ratio. The high DSE aperture sizes were typically a small percentage of the total main volume surface area. Generally, LDT/T10 seemed to be a more discriminating quantifier of DSE than T30/T15. The metrics of decay ratio and ΔL were also calculated via a Bayesian analysis algorithm programed by Dr. Ning Xiang at Rensselaer Polytechnic Institute, but those results did not appear consistent with what was expected; in many of the cases tested, the presence of a double slope decay was not found. Further work is suggested in calculating decay ratio and ΔL from Bayesian analysis of decay curves generated in realistic coupled volume spaces to determine the source of this inconsistency.

This study has also provided a comprehensive analysis of preference related to realistic non-exponential decays with an appropriate subject pool size leading to significant results. Subjective testing was conducted to determine the effect on listener preference of varying aperture size and absorption ratio in the virtual hall. MDS analysis was used to determine the relative preference results for different virtual hall configurations. In general, listener preference was inversely proportional to DSE level, with the highest levels of preference occurring at low and medium levels of DSE. Many of these decay curves were preferred over the single volume case of the main hall alone, though. Based on the computational and subjective results, recommended design guidelines for coupled volume halls have been provided in Fig. 9.

Further work could focus on other coupled volume parameters, such as the location of coupling apertures, which was found to have some effect on DSE in a previous study.[10] The distribution of absorption in the space warrants further investigation, as previous studies of single volume spaces have indicated that different absorption distributions cause significant changes in sound energy behavior in the space, such as a higher dependency on scattering coefficient values.[31] The effect of surface scattering values could also be studied in more detail. For example, the scattering values of

surfaces surrounding the coupling apertures may need to be higher than the ones used in this study to account for diffraction occurring at the aperture.

The realistic hall geometry studied in this research was based on the typical coupled volume hall, which is characterized by a rounded shoebox shape. Alternative geometries commonly seen in single volume halls, such as fan-shape and vineyard styles, could also be studied. Different modeling methods, such as physical scale modeling, radiative transfer models,[32,33] diffusion models,[34,35] and finite difference time-domain methods,[36] could be used to make additional conclusions about the influence of architectural parameters on DSE and its subjective impression. Aspects of the subjective response to DSE could be further investigated as well. In particular, different music selections could be tested, or additional subjective preference tests could be run comparing double slope decays to a wider range of single slope decays. Furthermore, other subjective data collection methods, such as multiple level or continuous scale tests, might be used to gain additional insight into the preference levels for coupled volume spaces.

ACKNOWLEDGMENTS

The authors would like to thank Damian Doria of Artec Consultants, Inc. and Dr. Ning Xiang for their assistance with this project. This work was partially supported by an Acoustical Society of America Graduate Fellowship for Minorities and the Ford Foundation Diversity Fellowship Program.

[1]D. T. Bradley and L. M. Wang, "The effects of simple coupled volume geometry on the objective and subjective results from non-exponential decay," J. Acoust. Soc. Am. **118**, 1480–1490 (2005).
[2]H. A. Davis, "Reverberation equations for two adjacent rooms connected by an incompletely soundproof partition," Philos. Mag. **50**, 75–80 (1925).
[3]L. Cremer and H. A. Muller, *Principles and Applications of Room Acoustics* (Applied Science, New York, 1982), Vol. **1**.
[4]A. Munro and R. Johnson, "Calculating reverberation time in coupled rooms," Workshop on Design of Concert Chambers, Tanglewood '99 (Concert Hall Research Group, Summer Institute, Lenox, MA, 1999).
[5]N. Xiang and P. M. Goggans, "Evaluation of decay times in coupled spaces: Bayesian parameter estimation," J. Acoust. Soc. Am. **110**, 1415–1420 (2001).
[6]N. Xiang and P. M. Goggans, "Evaluation of decay times in coupled spaces: Bayesian decay model selection," J. Acoust. Soc. Am. **113**, 2685–2697 (2003).
[7]C. F. Eyring, "Conditions under which residual sound in reverberant rooms may have more than one rate of decay," J. Soc. of Motion Picture Engineers **15**, 528–538 (1930).
[8]C. F. Eyring, "Reverberation time measurements in coupled rooms," J. Acoust. Soc. Am. **3**, 181–206 (1931).
[9]L. Nijs, G. Jansens, and M. van der Voorden, "The prediction of sound pressure levels in coupled rooms using ray-tracing models," Inter-Noise 98, Christchurch, New Zealand (1998).
[10]B. Harrison and G. Madaras, "Computer modeling and prediction in the design of coupled volumes for a 1000-seat concert hall at Goshen College, Indiana," J. Acoust. Soc. Am. **109**, 2388(A) (2001).
[11]J. E. Summers, R. R. Torres, and Y. Shimizu, "Statistical-acoustics models of energy decay in systems of coupled rooms and their relation to geometrical acoustics," J. Acoust. Soc. Am. **116**, 958–969 (2004).
[12]J. E. Summers, R. R. Torres, and Y. Shimizu, "Adapting a randomized beam-axis-tracing algorithm to modeling of coupled rooms via late-part ray tracing," J. Acoust. Soc. Am. **118**, 1491–1502 (2005).
[13]M. Ermann, "Coupled volumes: Aperture size and the double-sloped decay of concert halls," Build. Acoust. **12**, 1–14 (2005).
[14]D. T. Bradley and L. M. Wang, "Comparison of measured and computer-modeled objective parameters for an existing coupled volume concert hall," Build. Acoust. **14**, 79–90 (2007).
[15]ISO 3382-1, "Acoustics—Measurement of Room Acoustic Parameters—Part 1: Performance Spaces," 2009.
[16]N. Xiang, P. M. Goggans, T. Jasa, and M. Kleiner, "Evaluation of decay times in coupled spaces: reliability analysis of Bayesian decay time estimation," J. Acoust. Soc. Am. **117**, 3707–3715 (2005).
[17]D. T. Bradley and L. M. Wang, "Quantifying the double slope effect in coupled volume room systems," Build. Acoust. **16**, 105–123 (2009).
[18]B. S. Atal, M. R. Schroeder, and G. M. Sessler, "Subjective reverberation time and its relation to sound decay," Proceedings of the Fifth International Congress on Acoustics, Liege, Belgium (1965), Paper No. G32.
[19]D. Picard, "Audibility of non-exponential reverberation decays," Rapport de Stage D'Option Scientifique Rensselear Polytechnic Institute, 2003.
[20]M. Ermann, "Double sloped decay: Subjective listening test to determine perceptibility and preference," Build. Acoust. **14**, 91–107 (2007).
[21]C. L. Christensen, ODEON room acoustics program, User Manual v6.5, Industrial, Auditorium and Combined Editions, Odeon A/S, Denmark, 2003.
[22]C. L. Christensen, "ODEON—A design tool for auditorium acoustics, noise control and loudspeaker systems," Proceedings of the Institute of Acoustics: Reproduced Sound 17, Measuring, Modeling, or Muddling, Vol. **23**, pp. 137–144 (2004).
[23]ISO 3382-2, "Acoustics—Measurement of Room Acoustic Parameters—Part 2: Reverberation Time in Ordinary Rooms," 2008.
[24]M. D. Burkhard and R. M. Sachs, "Anthropometric manikin for acoustic research," J. Acoust. Soc. Am. **58**, 214–222 (1975).
[25]H. A. David, *The Method of Paired Comparisons* (Oxford University Press, New York, 1988).
[26]E. Parizet, N. Hamzaoui, and G. Sabatie, "Comparison of some listening test methods: a case study," Acta Acust. **91**, 356–364 (2005).
[27]M. G. Kendall and B. Babington-Smith, "On the method of paired comparisons," Biometrics **31**, 43–62 (1939).
[28]G. Knezek, S. Wallace, and P. Dunn Rankin, "Accuracy of Kendall's chi-square approximation to circular triad distributions," Psychometrika **63**, 23–34 (1998).
[29]J. B. Kruskal and M. Wish, *Multidimensional Scaling* (Sage, Newbury Park, CA, 1978).
[30]S. S. Schiffman, M. L. Reynolds, and F. W. Young, *Introduction to Multidimensional Scaling* (Academic, New York, 1981).
[31]L. M. Wang and J. Rathsam, "The influence of absorption factors on the sensitivity of a virtual room's sound field to scattering coefficients," Appl. Acoust. **69**, 1249–1257 (2008).
[32]E. Nosal, M. Hodgson, and I. Ashdown, "Improved algorithms and methods for room sound-field prediction by acoustical radiosity in arbitrary polyhedral rooms," J. Acoust. Soc. Am. **116**, 970–980 (2004).
[33]M. Hodgson and E. Nosal, "Experimental evaluation of radiosity for room sound-field prediction," J. Acoust. Soc. Am. **120**, 808–819 (2006).
[34]A. Billon, V. Valeau, A. Sakout, and J. Picaut, "On the use of a diffusion model for acoustically coupled rooms," J. Acoust. Soc. Am. **120**, 2043–2054 (2006).
[35]Y. Jing and N. Xiang, "Visualizations of sound energy across coupled rooms using a diffusion equation model," J. Acoust. Soc. Am. **124**, EL360–EL365 (2008).
[36]J. M. Botts and N. Xiang, "An application of the finite difference time-domain method to the study of coupled volumes," J. Acoust. Soc. Am. **125**, 2735(A) (2009).

The ensemble variance of pure-tone measurements in reverberation rooms[a)]

Finn Jacobsen[b)]
Department of Electrical Engineering, Acoustic Technology, Technical University of Denmark, Building 352, Kgs. Lyngby DK-2800, Denmark

Alfonso Rodríguez Molares
E.T.S.E. Telecomunicación, Universidade de Vigo, Campus Lagoas-Marcosende, Vigo E-36310, Spain

(Received 22 June 2009; revised 11 November 2009; accepted 11 November 2009)

Reverberation rooms are often used for measuring the sound power emitted by sources of sound. At medium and high frequencies, where the modal overlap is high, a fairly simple model based on sums of waves from random directions having random phase relations gives good predictions of the ensemble statistics of measurements in such rooms. Below the Schroeder frequency, the relative variance is much larger, particularly if the source emits a pure-tone. The established theory for this frequency range is based on ensemble statistics of modal sums and requires knowledge of mode shapes and the distribution of modal frequencies. This paper extends the far simpler random wave theory to low frequencies. The two theories are compared, and their predictions are found to compare well with experimental and numerical results.
 [DOI: 10.1121/1.3271034]

PACS number(s): 43.55.Cs, 43.58.Bh [LMW] Pages: 233–237

I. INTRODUCTION

Statistical models are well established in room acoustics. For example, more than half a century ago Schroeder[1] developed a stochastic model that predicts how the sound pressure at a given position in a reverberant room varies with the frequency. A closely related stochastic theory based on sums of coherent waves arriving from random directions and having random phases gives reliable predictions of the spatial fluctuations of the sound pressure in a reverberant room driven with a pure-tone above the Schroeder frequency.[2] An alternative theory for this and other phenomena is also stochastic in nature but based on sums of modes with random distributions of the modal frequencies.[3,4] The modal theory, which is more complicated than the random wave theory and requires more information about the room, is generally regarded as valid also below the Schroeder frequency.[4] However, there is surprisingly little experimental evidence of its validity in this frequency range.

This paper attempts to extend the simpler theoretical approach to low frequencies. The two theories are compared with experimental and numerical results.

II. OUTLINE OF EXISTING MODELS

The random wave theory is essentially due to Schroeder,[1] Andres,[5] Waterhouse,[2] and Lubman.[6] The alternative modal theory is essentially due to Lyon[3] and Davy,[4,7] but later modified by Weaver.[8]

A. The random wave theory

Above the Schroeder frequency, a harmonic sound field in a reverberation room can be modeled as a sum of waves

$$p(\mathbf{r}) = \lim_{N\to\infty} \frac{1}{\sqrt{N}} \sum_{n=1}^{N} A_n e^{j(\omega t + \mathbf{k}_n \cdot \mathbf{r})}, \tag{1}$$

where $p(\mathbf{r})$ is the sound pressure at position $\mathbf{r}$, A_n is a complex random amplitude the phase angle of which is uniformly distributed between 0 and 2π, and $\mathbf{k}_n$ is a random wave number vector with a uniform distribution over all angles of incidence. Note that no information about the particulars of the room has entered into this stochastic pure-tone diffuse field theory at this stage. It is easy to show that the corresponding expression for the mean square pressure is a sum of two independent squared Gaussian variables (random sums) with zero mean.[2] A sum of two squared Gaussian variables with zero mean has a chi-square distribution with two degrees of freedom,[2,9] from which it follows that the relative ensemble variance is 1,

$$\varepsilon^2\{p_{\mathrm{rms}}^2\} = 1. \tag{2}$$

In this expression, p_{rms}^2 is the time average of the squared sound pressure at a given position in a given room, and ε^2 is the relative ensemble variance (i.e., the ensemble variance normalized by the squared ensemble average). Above the Schroeder frequency, one can expect the same statistics with respect to room, position, and frequency, and a relative spatial variance of unity has been validated experimentally many times, e.g., in Refs. 2 and 10.

About 30 years ago, Jacobsen[11] and Pierce[12] independently used Eq. (1) to calculate the ensemble variance of the sound power output of a monopole that emits a pure-tone in

[a)]Portions of this work were presented in "The uncertainty of pure tone measurements in reverberation rooms below the Schroeder frequency," Proceedings of the 16th International Congress on Sound and Vibration, Krakow, Poland, July 2009.

[b)]Author to whom correspondence should be addressed. Electronic mail: fja@elektro.dtu.dk

a reverberation room. They showed that the reverberant part of the sound field expressed by Eq. (1) because of the resulting random contribution to the radiation impedance gives rise to a relative variance of the sound power output of

$$\varepsilon^2\{P_a\} = \frac{1}{M_s}, \tag{3}$$

where M_s is the statistical modal overlap of the room (the product of the modal density and the statistical bandwidth of the modes). The modal overlap, which is proportional to the square of the frequency and the total absorption area in the room, is large above the Schroeder frequency, and therefore, the sound power emitted by a monopole essentially equals its free field sound power in this frequency range.

B. The modal theory

The original version of the alternative theory, due to Lyon,[3] is based on the analytical Green's function in a rectangular room, which is a modal sum.[12] Lyon assumed that the modal frequencies have a Poisson distribution (i.e., are distributed independently of each other) and predicted the ensemble variance of the sound power output of a monopole and of the mean square pressure. Some years later, Davy[4] extended Lyon's theory by deriving a more general expression of the power transmission function averaged over multiple source and receiver positions, assuming a "nearest neighbor" distribution of the modal frequencies. The assumption of the modal frequencies having a nearest neighbor distribution rather than a Poisson distribution came from evidence of a "modal repulsion" effect already discussed by Lyon.[3] In 1989, Weaver[8] modified Davy's expression so as to take account of a modal frequency spacing described by the Gaussian orthogonal ensemble theory, which is now generally accepted.[13] Finally, in 1990, Davy[14] described and discussed the modified theory. The resulting expression for the relative ensemble variance of the sound power of a monopole became

$$\varepsilon^2\{P_a\} = \frac{K-1}{M_s}, \tag{4}$$

where the spatial factor K is the normalized fourth moment (the kurtosis) of the modal functions ψ as follows:

$$K = \frac{E\{\psi^4\}}{E^2\{\psi^2\}}. \tag{5}$$

In a rectangular room, the mode shape ψ is a product of cosines. This gives a value of $(3/2)^3$ for oblique modes. With this value of K, Eq. (4) is similar to but somewhat larger than the prediction given by Eq. (3) ($K-1 \simeq 2.38$).

The corresponding expression for the relative ensemble variance of the mean square pressure became[8,14]

$$\varepsilon^2\{p_{\mathrm{rms}}^2\} = 1 + \frac{K^2-3}{M_s}, \tag{6}$$

which asymptotically approaches Eq. (2) for high values of the modal overlap. [With $K=(3/2)^3$ Eq. (6) becomes $\varepsilon^2\{p_{\mathrm{rms}}^2\} \simeq 1+8.4/M_s$.] According to Davy[15] the right-hand side of Eq. (6) can be expected to be modified at very low modal overlap

$$\varepsilon^2\{p_{\mathrm{rms}}^2\} = 1 + \frac{K^2}{M_s}. \tag{7}$$

Although there is little doubt that there is an increase in the relative variance of the mean square pressure in a frequency region below the Schroeder frequency compared with the asymptotic value of unity at high modal overlap,[16] there is, for some reason, surprisingly little, if any, experimental evidence in direct support of Eqs. (6) and (7).

It should be mentioned that other authors have suggested lower values of K. Weaver[8] and Lobkis *et al.*[17] are in favor of a value of 3, which corresponds to normally distributed modal amplitudes; experimental and numerical results have led Langley and Brown[18] to suggest a value of 2.7. See discussions in Refs. 8, 13, and 17–19.

C. Extension of the random wave theory

Recently, Jacobsen and Rodríguez Molares[20] modified Eq. (3) by taking into account the local increase in the reverberant part of the sound field at the source position due to "weak Anderson localization," as predicted by Weaver and Burkhardt.[21] Equation (3) now became

$$\varepsilon^2\{P_a\} = \frac{F(M_s)}{M_s}, \tag{8}$$

where the "concentration factor" $F(M_s)$ is a function that goes smoothly from 3 to 2 as the modal overlap is increased.[13] The modified expression was validated experimentally in various reverberation rooms as well as by numerical calculations. As pointed out in Ref. 20, it is interesting and somewhat surprising that the modal overlap enters into a theory that does not make use of the concept of modes. Since Eqs. (8) and (4) are very similar [with $F=2$ and $K=(3/2)^3$] these results also validated the modal prediction. With $F=2$ and $K=3$ the two expressions are identical.

The finite ensemble variance of the sound power emitted by the source at low and medium modal overlaps means that the ensemble average of $|A_n|^2$ varies from outcome to outcome of the stochastic process. Such variations are not taken into account by Eq. (2). Thus one might expect additional variations in the mean square pressure corresponding to the variations in the sound power below the Schroeder frequency. This can be modeled by multiplying the original exponentially distributed mean square pressure by another random variable that represents the relative variations in the emitted sound power. The latter is a normally distributed independent random variable with an average of 1 and a variance given by Eq. (8). It follows that Eq. (2) becomes

$$\begin{aligned}\varepsilon^2\{p_{\mathrm{rms}}^2\} = \varepsilon^2\{x(1+y)\} &= \frac{E\{x^2(1+y)^2\}}{E^2\{x(1+y)\}} - 1 \\ &= \frac{E\{x^2\}E\{(1+y)^2\}}{E^2\{x\}E^2\{(1+y)\}} - 1 = \frac{2E\{(1+y)^2\}}{E^2\{(1+y)\}} - 1 \\ &= 2(1+E\{y^2\}) - 1 = 1 + \frac{2F(M_s)}{M_s},\end{aligned} \tag{9}$$

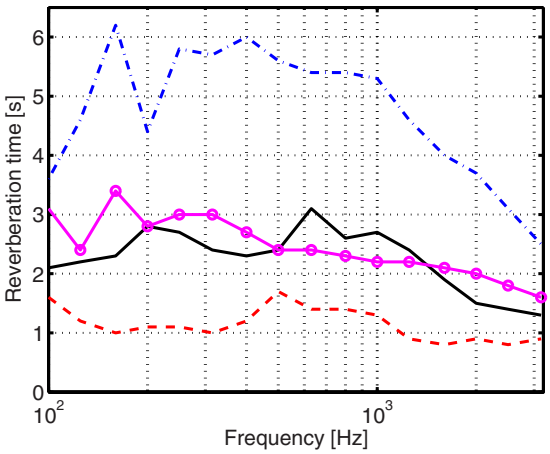


FIG. 1. (Color online) Reverberation time of the test rooms. Solid line: small lightly damped room; dashed line: small damped room; dashed-dotted line: large reverberation room; and line with circle markers: large damped reverberation room.

where x is the original mean square pressure and y is an independent, normally distributed variable with zero mean that represents the relative fluctuations of the emitted sound power. Note that Eq. (9) is similar to Weaver's equation (6) although the "correction" to the asymptotic high frequency value of unity is somewhat smaller. If $K=3$ is used in Weaver's expression rather than the value for oblique modes in a rectangular room it becomes $\varepsilon^2\{p_{\mathrm{rms}}^2\}\simeq 1+6/M_s$, which is quite similar to Eq. (9).

III. EXPERIMENTAL RESULTS

To test the validity of Eqs. (6) and (9) some experiments have been carried out in various rooms at the Technical University of Denmark: a small room (40 m^3) with almost bare walls, the same room with absorption added, a large reverberation room (245 m^3) with very little damping, and the same room with added absorption. Both rooms are essentially rectangular, but there are large (fixed) diffusers in the reverberation room, and the absorption was in all cases irregularly distributed. Figure 1 shows the reverberation time of the rooms, measured in one-third octave bands using the interrupted noise method and a Brüel & Kjær (B&K) "PULSE" analyzer. The Schroeder frequency of the small bare room is about 500 Hz; with added absorption, it is reduced to about 330 Hz. In the large reverberation room, the Schroeder frequency is about 310 Hz; and with added absorption, it is reduced to 200 Hz.

The monopole generating the sound field was a B&K OmniSource fitted with a "volume velocity adapter" with two matched quarter-inch microphones,[22] and the sound pressure was measured using the pressure microphone of a "ultimate sound probe" (USP), a three-dimensional pressure-velocity probe produced by Microflown (Zevenaar, The Netherlands). The frequency response between the volume velocity of the source and the sound pressure in the room was measured with the same B&K analyzer but in the fast Fourier transform (FFT) mode, using pseudorandom noise (6400 spectral lines) in the frequency range of up to 3.2 kHz. A similar technique was used recently to validate Eqs. (4) and (8).[20] In order to approach the full variation associated with ensemble statistics both the source and receiver positions were varied, and in the postprocessing of the results (obtained at 25 pairs of positions), additional variations over 8 Hz frequency bands (16 neighboring frequencies) were also taken into account.

Figure 2 compares the results of the measurements of the mean square pressure with predictions calculated using Eq. (6) using a weighted average of K for oblique, tangential, and axial modes, as suggested by Davy,[14] a similar prediction using $K=3$, as suggested by Weaver,[8] and a prediction calculated using Eq. (9) and $F=2$. As can be seen, the two predictions calculated from Eq. (6) are practically identical, whereas Eq. (9) gives slightly lower values. The measured relative space-frequency standard deviation of the mean square pressure fluctuates somewhat with the frequency, but follows the predicted tendency fairly closely. Above 1 kHz it approaches unity as expected. In general, the experimental data seem to agree equally well with Eqs. (6) and (9).

IV. NUMERICAL RESULTS

It is not practical to determine the full ensemble standard deviation experimentally, but it can be done with a numerical model. In this case, the matter has been examined with the finite element method (FEM), using FEM models of 25 different rooms. The rooms were rectangular and they had a uniform locally reacting wall impedance. The FEM was constructed using the commercial software packet ACTRAN. The dimensions of the rooms were chosen as uniform random variables varying between 2 and 6 m. The source position was placed at random but at least 0.4 m away from any wall. The calculations were carried out from 200 to 300 Hz with a frequency step of 2 Hz. The element size was chosen so as to provide a low numerical pollution in the examined frequency range. The mean square pressure was calculated at 50 000 randomly chosen nodal points of the mesh. Nodes closer than 0.4 m away from the walls or closer than 1 m from the source were not used. In order to determine the relative ensemble standard deviation as a function of the modal overlap, the data were sorted into appropriate modal overlap intervals. A similar technique was used recently in Ref. 20.

In Fig. 3 the results of the FEM calculations are compared with predictions determined using Eq. (9) with $F=2$ and $F=3$. [The latter case is identical to Eq. (6) with $K=3$.] A prediction calculated using Eq. (7) and $K=(3/2)^3$ is also shown. On the whole, the FEM results take values corresponding to Eq. (9) with F between 2 and 3, or Eq. (6) with K between 2 and 3. On the other hand, Eq. (7) overestimates significantly, probably because the modal overlap is too high.

Figure 4 shows the results of similar calculations with the boundary element model (BEM) using the Open Source software "OPENBEM" (University of Southern Denmark, Odense, Denmark) in an ensemble of rectangular two-dimensional "rooms" with uniform wall impedance in the frequency range between 200 and 400 Hz. The same parameters were used as in the FEM calculations, but since there were no nodes in the room, in this case, the mean square pressure was calculated at 1000 random positions at least 0.4

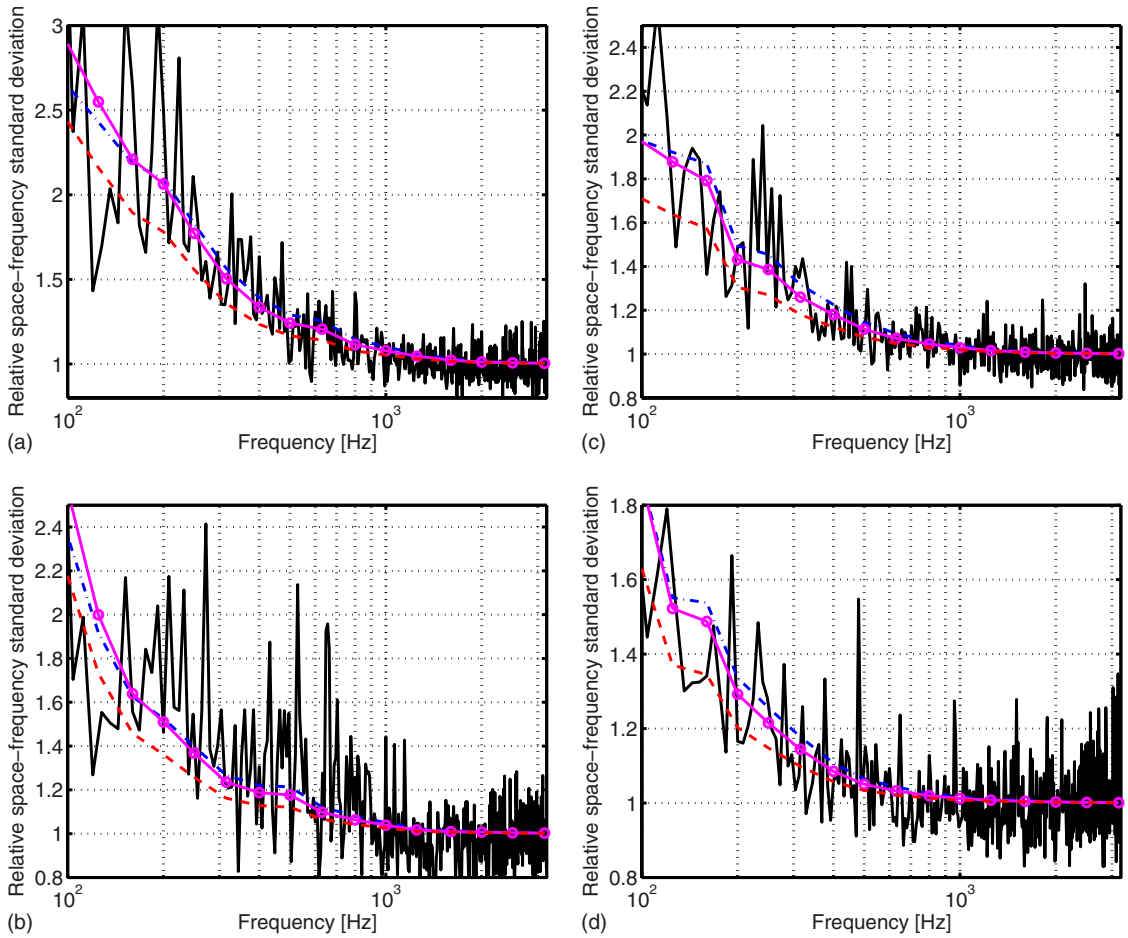


FIG. 2. (Color online) Relative space-frequency standard deviation of mean square pressure in (a) small lightly damped room, (b) small damped room, (c) large reverberation room, and (d) large damped reverberation room. Solid line: measured standard deviation; dashed line: prediction based on Eq. (9) with $F=2$; dashed-dotted line: prediction based on Eq. (6) with a weighted average for K; and line with circle markers: prediction based on Eq. (6) with $K=3$.

m from the walls and 1 m from the source. The BEM results are compared with predictions determined using Eq. (9) with $F=2$ and 3, and with Eq. (7) and $K=(3/2)^2$. The numerical results are in fairly good agreement with the predictions, particularly those based on Eq. (9) with $F=2$.

V. DISCUSSION

It should be mentioned that the extended random wave theory relies on a result from the modal theory: the concentration factor F. Nevertheless, it is surprising that such a simple theory gives results that, from a practical point of

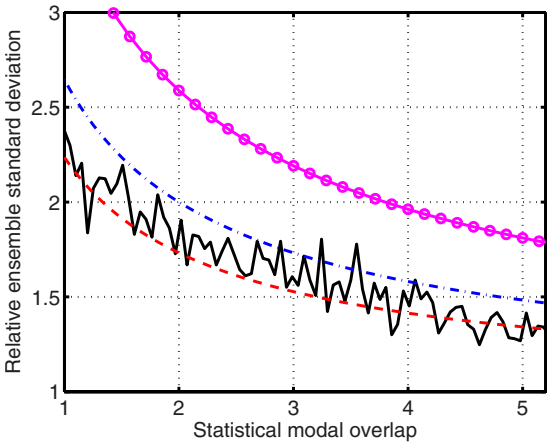


FIG. 3. (Color online) Relative ensemble standard deviation of mean square pressure in three-dimensional rooms. Solid line: FEM calculation; dashed line: prediction based on Eq. (8) with $F=2$; dashed-dotted line: prediction based on Eq. (9) with $F=3$ (or Eq. (6) with $K=3$); and line with circle markers: prediction based on Eq. (7) with $K=(3/2)^3$.

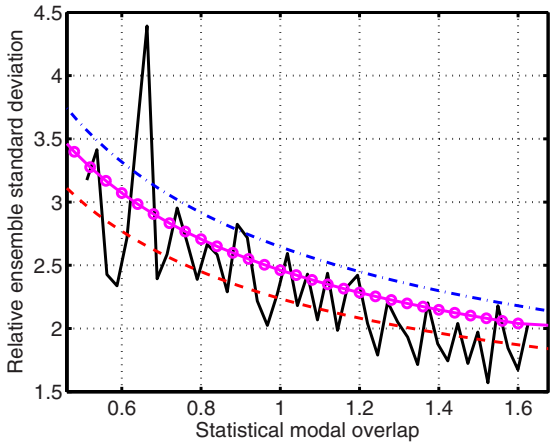


FIG. 4. (Color online) Relative ensemble standard deviation of mean square pressure in two-dimensional "rooms." Solid line: BEM calculation; dashed line: prediction based on Eq. (9) with $F=2$; dashed-dotted line: prediction based on Eq. (9) with $F=3$ (or Eq. (6) with $K=3$); and line with circle markers: prediction based on Eq. (7) with $K=(3/2)^2$.

view, are simply identical to the results of the far more complicated modal theory. On the other hand, in view of the fact that modes can be decomposed into waves, it is perhaps not *that* surprising that the two approaches lead to comparable ensemble statistics.

Equation (8) is arguably a first order approximation that might be improved by taking into account the increased variance of the reverberant part of the sound pressure predicted by Eq. (9). However, the experimental and numerical results presented in Ref. 20 confirm Eq. (8) and do not support any "higher order correction."

VI. CONCLUSION

Experimental and numerical results confirm that there is a substantial increase in the relative ensemble variance of the mean square pressure in a reverberation room driven by a monopole that emits a pure-tone below the Schroeder frequency. Above this frequency, the relative variance approaches unity; below this frequency, there is an increase in the variance that is inversely proportional to the modal overlap, that is, proportional to the ratio of the reverberation time to the room volume and inversely proportional to the square of the frequency.

Waterhouse's simple random wave theory has been extended to the frequency range below the Schroeder frequency and has shown to give predictions of the relative ensemble variance of the mean square pressure in good agreement with the more complicated statistical modal theory due to Lyon, Davy, and Weaver, and these predictions are confirmed by the experimental and numerical results.

[1]M. R. Schroeder, "Statistical parameters of the frequency response curves of large rooms," J. Audio Eng. Soc. **35**, 299–305 (1987); originally published in German in Acustica **4**, 594–600 (1954).

[2]R. V. Waterhouse, "Statistical properties of reverberant sound fields," J. Acoust. Soc. Am. **43**, 1436–1444 (1968).

[3]R. H. Lyon, "Statistical analysis of power injection and response in structures and rooms," J. Acoust. Soc. Am. **45**, 545–565 (1969).

[4]J. L. Davy, "The relative variance of the transmission function of a reverberation room," J. Sound Vib. **77**, 455–479 (1981).

[5]H. G. Andres, "Über ein gesetz der räumlichen zufallsschwankung von rauschpegeln in räumen und seine anwendung auf schalleistungsmessungen (The spatial variations of noise levels in rooms and applications for sound power measurements)," Acustica **16**, 279–294 (1965/66).

[6]D. Lubman, "Fluctuations of sound with position in a reverberant room," J. Acoust. Soc. Am. **44**, 1491–1502 (1968).

[7]J. L. Davy, "Improvements to formulae for the ensemble relative variance of random noise in a reverberation room," J. Sound Vib. **115**, 145–161 (1987).

[8]R. L. Weaver, "On the ensemble variance of reverberation room transmission functions, the effect of the spectral rigidity," J. Sound Vib. **130**, 487–491 (1989).

[9]A. Papoulis and S. U. Pillai, *Probability, Random Variables, and Stochastic Processes*, 4th ed. (McGraw-Hill, New York, 1991).

[10]K. J. Ebeling, K. Freudenstein, and H. Alrutz, "Experimental investigation of statistical properties of diffuse sound fields in reverberation rooms," Acustica **51**, 145–153 (1982).

[11]F. Jacobsen, "The diffuse sound field," Report No. 27, Acoustics Laboratory, Technical University of Denmark, Kongens Lyngby, Denmark, 1979.

[12]A. D. Pierce, *Acoustics, An Introduction to Its Physical Principles and Applications* (McGraw-Hill, New York, 1981).

[13]R. S. Langley and V. Cotoni, "The ensemble statistics of the vibrational energy density of a random system subjected to single point harmonic excitation," J. Acoust. Soc. Am. **118**, 3064–3076 (2005).

[14]J. L. Davy, "The distribution of modal frequencies in a reverberation room," in Proceedings of the Inter-Noise '90, Gothenburg, Sweden (1990).

[15]J. L. Davy, "The variance of pure tone reverberant sound power measurements," in Proceedings of the Fifth International Congress on Sound and Vibration, Adelaide, Australia (1997).

[16]D. Lubman, "Precision of reverberant sound power measurements," J. Acoust. Soc. Am. **56**, 523–533 (1974).

[17]O. I. Lobkis, R. L. Weaver, and I. Rozhkov, "Power variances and decay curvature in a reverberant system," J. Sound Vib. **237**, 281–302 (2000).

[18]R. S. Langley and A. W. M. Brown, "The ensemble statistics of the energy of a random system subjected to harmonic excitation," J. Sound Vib. **275**, 823–846 (2004).

[19]O. I. Lobkis and R. L. Weaver, "Complex modes in a reverberant dissipative body," J. Acoust. Soc. Am. **108**, 1480–1485 (2000).

[20]F. Jacobsen and A. Rodríguez Molares, "Sound power emitted by a pure-tone source in a reverberation room," J. Acoust. Soc. Am. **126**, 676–684 (2009).

[21]R. L. Weaver and J. Burkhardt, "Weak Anderson localization and enhanced backscatter in reverberation rooms and quantum dots," J. Acoust. Soc. Am. **96**, 3186–3190 (1994).

[22]S. Gade, N. Møller, J. Hald, and L. Alkestrup, "The use of volume velocity source in transfer measurements," in Proceedings of the Inter-Noise 2004, Prague, Czech Republic (2004).

Sound absorption of a micro-perforated panel backed by an irregular-shaped cavity

Chunqi Wang[a] and Li Cheng[b]
Department of Mechanical Engineering, The Hong Kong Polytechnic University, Kowloon, Hong Kong

Jie Pan
School of Mechanical Engineering, The University of Western Australia, 35 Stirling Highway, Crawley, Western Australia 6009, Australia

Ganghua Yu
Department of Mechanical Engineering, The Hong Kong Polytechnic University, Kowloon, Hong Kong

(Received 23 April 2009; revised 6 October 2009; accepted 7 October 2009)

In the pursuit of more effective noise control devices, the cavity backed micro-perforated panel absorber (CBMPPA) is developed in this study. A CBMPPA differs from the conventional micro-perforated panel (MPP) absorber in that the MPP is backed by a trapezoidal cavity, which allows more effective vibroacoustic coupling between the MPP and the cavity. A two-dimensional theoretical model is established and tested both numerically and experimentally. Based on the verified theoretical model, sound absorption performance of a trapezoidal CBMPPA is investigated numerically. Results show that the shape of the backing cavity can significantly alter the sound absorption mechanisms and frequency distribution of overall sound absorption coefficient of the absorber. Further analyses show that acoustic modes that are initially decoupled from the MPP in the rectangular configuration are coupled with the air motion in the MPP, which accounts for the change in absorption pattern of the trapezoidal CBMPPA. By the same token, it also provides the flexibility for tuning the effective absorption range of the absorber. Due to the varying impedance matching conditions, the absorption performance of the trapezoidal CBMPPA exhibits obvious local characteristics over the MPP surface, which contrasts with the spatially uniform absorption in the conventional MPP absorber. © *2010 Acoustical Society of America.* [DOI: 10.1121/1.3257590]

PACS number(s): 43.55.Ev, 43.50.Gf [NX] Pages: 238–246

I. INTRODUCTION

Micro-perforated panel (MPP) has been used as an alternative to traditional porous sound absorbing materials for a number of years. A MPP consists of a sheet panel with a lattice of sub-millimeter size perforations distributed over its surface. By reducing the sizes to sub-millimeter scale, the perforations themselves provide acoustic resistance and low acoustic mass reactance necessary for an absorber without the use of any porous material (Maa, 1975, 1987). As a clean and efficient sound absorbing material, MPP is finding more and more applications in areas such as room acoustics (Fuchs and Zha, 1997; Kang and Fuchs, 1999; Droltleff and Zhou, 2001; Kang and Brocklesby; 2005), environmental noise abatement (Asdrubali and Pispola, 2007), and duct noise control (Wu, 1997). Conventionally, a typical MPP absorber takes the form of a MPP fitted in front of a backing wall, as illustrated in Fig. 1(a). The backing wall is parallel to the MPP with an air gap between them. The air gap provides an acoustic-stiffness, which is controlled by its depth. Together with the perforation, resonance-type absorption is generated. In all the existing work mentioned above, the depth of the air gap between the MPP and the backing wall is assumed to be a constant value. Previous studies on the MPP absorber mainly focus on the properties of the MPP itself, e.g., orifice diameter, perforation rate, panel thickness and materials, etc. Little attention is paid to the cavity depth as it is usually assumed to be a constant according to the target resonance frequency of the absorber (Maa, 1987).

This work aims to study the sound absorption characteristics of an irregular-shaped cavity backed micro-perforated panel absorber (CBMPPA) with varying cavity depths (a MPP backed by a trapezoidal cavity), as shown in Fig. 1(b). The motivation for changing the constant cavity depth in the conventional MPP absorbers to a varying value is explained as below. From the vibroacoustic viewpoint, the sound absorption of the MPP absorber is dominated by the resonance of the air mass vibration of both the perforated holes and the backing cavity (or the air gap), and the absorption efficiency is limited to the resonance frequency region. Forming a closed space behind the MPP, multi-modal acoustic resonances can be created. For MPP absorbers with rectangular-shaped cavities, however, only a few selected acoustic modes play a role in the sound absorption due to the perpendicular relationship among their three adjacent walls. Other modes, although abundant in number, are decoupled from the air motion inside the micro-perforations of the MPP and hence have no contribution toward the sound absorption. This selective coupling pattern can be altered by setting the MPP

[a]Present address: Department of Mechanical Engineering, The University of Hong Kong, Pokfulam Road, Hong Kong.
[b]Author to whom correspondence should be addressed. Electronic mail: mmlcheng@inet.polyu.edu.hk

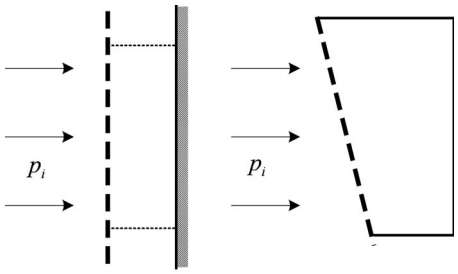


FIG. 1. Schematic of the MPP absorbers. (a) Conventional configuration in which a MPP is backed by a rigid wall or a rigid rectangular cavity (as illustrated by the dotted lines). (b) Irregular-shaped configuration in which a MPP is backed by a rigid trapezoidal cavity.

inclined to the back wall to form a so-called irregular-shaped cavity as opposed to the rectangular one. Previous work showed that the coupling between a structure and an enclosure can be greatly enhanced due to the inclination of the cavity wall (Li and Cheng, 2004; Sum and Pan, 2006). In the present case, acoustic modes that are initially decoupled from the MPP in the rectangular configuration can also be coupled with the air motion in the MPP. The expected result of the increased vibroacoustic coupling is the enhanced sound absorption performance over a wider bandwidth.

The idea of developing the irregular-shaped CBMPPA is also motivated by the practical need in the pursuit of more efficient noise barriers for environmental noise control. Noise barrier design and performance have been well documented in literature (Bies and Hansen, 1996). Since the present study does not claim any contribution in this regard, a literature review on noise barrier simulation and design is omitted here. Despite the significant progress made in the past, measurements have shown that performance of sound barriers is much less than that predicted by theory, especially for parallel barriers and barriers installed in front of noise sources with large reflective surfaces (Fuchs, 2001; Watts, 1996). This phenomenon is attributed to the multiple reflection and over-the-top scattering of the sound waves. To tackle the problem, an innovative wave tapping barrier was developed by Pan *et al.* (2004). The wave trapping barrier uses a cavity-backed perforated surface tilled toward the road. The panel is composed of a series of wedges, which can take the shape of a trapezoid or, as a degenerated case, a triangle box. The tilted profile allows noise to be trapped for a better absorption and, by the same token, minimizes the noise escaping from the top.

The success of this promising technology is, however, still limited by a number of factors. The first is the lack of a thorough understanding of the sound absorption mechanism of the trapezoidal CBMPPA. The second is the lack of a suitable design tool to optimize and simplify the development of such devices. Both of them require a reliable theoretical model that can describe the vibroacoustic behavior of the CBMPPA accurately. Most previous studies modeled the MPP absorber as an equivalent electrical circuit. Such a method surely applies to the conventional construction as shown in Maa (1975, 1987), but fails when the cavity depth is varying. The transfer matrix method, described in Lee and Kwon (2004) and Atalla and Sgard (2007), also assumes a constant cavity depth, which becomes powerless in dealing with the irregular-shaped backing cavity. In studying the sound absorption of a finite flexible MPP backed by an air cavity, Lee *et al.* (2005) developed a solution procedure based on the modal analysis approach. The theoretical model established in Lee *et al.* (2005) takes into account the full coupling between the cavity acoustics and the panel vibration. However, as it assumes a rectangular cavity and the exterior acoustic loading on the MPP surface is simplified as a uniformly distributed sound pressure, that model cannot be adopted to investigate the irregular-shaped CBMPPA directly.

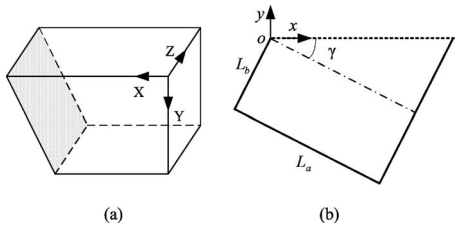


FIG. 2. Theoretical model of the irregular-shaped CBMPPA. (a) The three-dimensional configuration. The inclined wall is covered by a MPP while the other five walls of the trapezoidal cavity are acoustically rigid. (b) The two-dimensional theoretical model.

In what follows, Sec. II outlines the theoretical formulation for the sound absorption of the irregular-shaped CBMPPA based on a two-dimensional (2D) model. A method of solution is established to solve the fully coupled system among the MPP, the irregular-shaped backing cavity, and the exterior sound field. The absorption performance of an irregular-shaped CBMPPA is investigated numerically in Sec. III. Compared with the rectangular CBMPPA of the same cavity volume, the trapezoidal CBMPPA has a quite different absorption pattern with more spectral peaks. With proper geometrical configuration, the trapezoidal CBMPPA can achieve fairly good absorption performance over a wider continuous bandwidth. In addition, the absorption of the trapezoidal CBMPPA demonstrates obvious local characteristics, which is caused by the distorted acoustic modes of the trapezoidal cavity. In Sec. IV, the theoretical model is validated experimentally.

II. THEORETICAL MODELING

Figure 2(a) shows the three-dimensional (3D) configuration of an irregular-shaped CBMPPA, which consists of a trapezoidal cavity and a MPP covering on the inclined wall. The MPP itself can be either flexible or rigid, but it is assumed to be rigid for simplification in this study. The other five cavity walls are assumed to be acoustically rigid. When an acoustic wave p_i is incident on the MPP, some of the acoustic energy is reflected, and the rest is absorbed by the MPP as well as the sound absorbing materials inside the cavity (if any). As long as the inclined wall is parallel to the Z direction and the incident wave is parallel to the X-Y plane, it is appropriate to simplify the 3D configuration as a 2D theoretical model, as shown in Fig. 2(b). For convenience, a proper rotation is performed so that the MPP is laid out hori-

zontally in Fig. 2(b). The trapezoidal cavity is completely defined by the side length L_a, L_b and the inclination angle γ. The local coordinate system xoy defined in Fig. 2(b) is adopted throughout the following presentation.

The theoretical formulation starts with the analytical treatments on the aerial motion and the acoustic impedance of the micro-perforations. With a sound pressure difference applied between the two sides of the MPP, air mass vibration occurs inside the orifice. As a measure to simplify the theoretical modeling, the discrete air particle velocity over each orifice is averaged across the adjacent unperforated region so that a quasicontinuous particle velocity field $\bar{u}(x)$ is obtained. Such a simplification procedure is valid as long as the distance between orifices R is short enough compared with the acoustic wavelength λ, say, $R<\lambda/4$ (Putra, 2007). The acoustic impedance due to the micro-perforations is also averaged over the whole panel so that the relative (to the characteristic impedance $\rho_0 c$ in air) acoustic impedance of the MPP is found as (Maa, 1998)

$$Z = \frac{32\eta t}{\sigma\rho_0 c d^2}\left[\left(1+\frac{K^2}{32}\right)^{1/2}+\frac{\sqrt{2}}{32}K\frac{d}{t}\right] + i\cdot\frac{\omega t}{\sigma c}\left[1+\left(1+\frac{K^2}{2}\right)^{-1/2}+0.85\frac{d}{t}\right], \tag{1}$$

where $\omega=2\pi f$ is the angular frequency, $K=d\sqrt{\omega\rho_0/4\eta}$, d is the orifice diameter, t is the panel thickness, η is the coefficient of viscosity, and σ is the perforation ratio. Equation (1) has been shown to be accurate enough when the sound pressure level is less than 100 dB and the orifices are spaced more than a diameter apart (Maa, 1987, 1998). Thus, for normal incident wave p_i with unit amplitude, the acoustic boundary condition on the inclined MPP can be described as

$$Z\cdot\bar{u}(x) = \frac{1}{\rho_0 c}[p_{\text{cav}} - (2p_i + p_{\text{rad}})], \tag{2}$$

where p_{cav} is the sound pressure inside the backing cavity, $2p_i$ is the blocked pressure, and p_{rad} is the external pressure radiated by the particle velocity field $\bar{u}(x)$. Since both p_{cav} and p_{rad} are due to the radiation of the particle velocity field $\bar{u}(x)$, Eq. (2) describes the coupled behavior between the acoustic field and the air motion in the micro-perforations. A method of solution to this coupled equation is presented in Sec. II A.

A. Method of solution

Following the standard Galerkin procedure (Meirovitch, 2001), the particle velocity over the MPP surface $\bar{u}(x)$ is expanded as a series of sine functions

$$\bar{u}(x) = \sum_{j=1}^{J} u_j \sin(j\pi x/L), \tag{3}$$

where u_j is the complex amplitude and L is the length of the inclined wall. With a proper truncation of the decomposition series, Eq. (3) can represent the averaged particle velocity field $\bar{u}(x)$ accurately except at the boundary edges. As far as the vibroacoustic analysis in this study is concerned, the approximation errors of $\bar{u}(x)$ at the boundary edges (the two ending points in the 2D model) can be neglected. Substituting Eq. (3) into the coupled Eq. (2), multiplying $2\sin(j\pi x/L)$ to both sides, and integrating over the inclined wall $(0<x<L)$ give

$$Zu_j + \frac{2}{\rho_0 cL}\int_0^L (p_{\text{rad}} - p_{\text{cav}})\sin(j\pi x/L)dx = \frac{4}{\rho_0 cL}\int_0^L p_i \sin(j\pi x/L)dx. \tag{4}$$

The sound pressure inside the cavity due to $\bar{u}(x)$, p_{cav}, can be expressed in terms of acoustic modes of the rigid-walled cavity (Cheng, 2005):

$$p_{\text{cav}}(x,y) = \sum_{m=0}^{M} A_m\psi_m(x,y) = -\frac{i\rho_0 c^2\omega}{V}\sum_{m=0}^{M}\frac{\psi_m(x,y)}{\Lambda_m(\omega^2-\omega_m^2)}\int_0^L \bar{u}(x')\psi_m(x',0)dx', \tag{5}$$

where $\psi_m(x,y)$ is the mth acoustic mode of the trapezoidal cavity, A_m is the complex amplitude for each cavity mode $\psi_m(x,y)$, V is the cavity volume, ρ_0 is the air density, c is the speed of sound in air, and Λ_m is the normalized factor of the mth cavity mode. The acoustic modes of the trapezoidal cavity are found numerically with the commercially available software COMSOL® Multiphysics. In the conventional MPP absorber backed by a rectangular cavity, the cavity modes can be expressed analytically as

$$\psi_{m,n}(x,y) = \cos\left(\frac{m\pi x}{L_a}\right)\cos\left(\frac{n\pi y}{L_b}\right), \tag{6}$$

where L_a and L_b are the side lengths shown in Fig. 2(b) with $\gamma=0$. Accordingly, Eq. (5) becomes

$$p_{\text{cav}}(x,y) = \sum_{m,n=0}^{M,N} A_{m,n}\psi_{m,n}(x,y). \tag{7}$$

The absolute value of the complex amplitude A_m in Eq. (5) or $A_{m,n}$ in Eq. (7) measures the contribution of each cavity mode toward the sound field in the cavity. In the following presentation, the acoustic modes of the rectangular cavity will be called RC modes and the modes of the trapezoidal cavity will be called TC modes for brevity.

The external pressure imposing on the MPP includes the blocked pressure $2p_i$ and the external radiation by the micro-perforations p_{rad}. For a large array panel which is composed of a group of MPP absorbers arranged periodically, e.g., the sound barrier developed by Pan *et al.* (2004), the radiated pressure p_{rad} consists of both self-radiation and mutual radiation from other MPP absorbers. Assuming that the array panel is large enough, an approximate periodic (or rigid) boundary condition may be applied to each single MPP absorber, in which the external radiation by each MPP absorber is confined to a virtual rigid duct with infinite length. Another reason for adopting such treatment is that such a boundary condition conforms to the measurement method for the normal incidence absorption coefficient using impedance

tubes, as detailed in Sec. IV. Thus, given the averaged particle velocity distribution $\bar{u}(x)$, the radiated pressure p_{rad} is found as (Doak,1973)

$$p_{\text{rad}} = \frac{\rho_0}{L}\sum_{n=0}^{\infty} c_n \varphi_n(x) \int_0^L \varphi_n(x')\bar{u}(x')dx', \tag{8}$$

which simulates the situation when the MPP absorber is fixed to one end of a standing wave tube to measure its sound absorption characteristics. In Eq. (8), $\varphi_n(x)$ is the duct acoustics modes of a 2D channel of width L,

$$\varphi_n(x) = \sqrt{2-\delta_{0n}}\cos(n\pi x/L), \tag{9}$$

where δ_{0n} is the Kronecker delta and c_n is the modal phase speed,

$$c_n = \frac{ic}{\sqrt{(n\pi c/\omega L)^2 - 1}}, \quad n = 1,2,3,\ldots. \tag{10}$$

In order to solve Eq. (4), both p_{cav} and p_{rad} are related to the complex amplitudes of the sine series u_j via a modal impedance matrix. Denote the cavity pressure caused by $\sin(j\pi x/L)$ as $p_{\text{cav},j}$, that is,

$$p_{\text{cav},j} = -\frac{i\rho_0 c^2 \omega}{V}\sum_{m=0}^{M}\frac{\psi_m(x,y)}{\Lambda_m(\omega^2-\omega_m^2)}\int_0^L \sin(j\pi x'/L) \times \psi_m(x',0)dx'. \tag{11}$$

Then, the total cavity pressure p_{cav} becomes

$$p_{\text{cav}} = \sum_j u_j \cdot p_{\text{cav},j}. \tag{12}$$

Similarly, the total external radiation p_{rad} is

$$p_{\text{rad}} = \sum_i u_i \cdot p_{\text{rad},j}. \tag{13}$$

Substitute Eqs. (12) and (13) into Eq. (4) and denote

$$Z_{\text{cav},jl} = \int_0^L p_{\text{cav},l}(x)\cdot\sin(j\pi x/L)dx \tag{14}$$

and

$$Z_{\text{rad},jl} = \int_0^L p_{\text{rad},l}(x)\cdot\sin(j\pi x/L)dx. \tag{15}$$

Thus, Eq. (4) is transformed to a truncated set of linear equations for the unknown coefficients u_j,

$$Z\cdot\{u_j\} + \frac{2}{\rho_0 cL}[Z_{\text{rad},jl} - Z_{\text{cav},jl}]\cdot\{u_j\} = \{I_j\}, \tag{16}$$

where

$$I_j = \frac{4}{\rho_0 cL}\int_0^L p_i \sin(j\pi x/L)dx. \tag{17}$$

Equation (16) can be solved via the inversion of matrix. In the actual calculation, modal truncation is needed. In Eq. (4), 400 TC modes are used to calculate the sound field in the trapezoidal cavity. For the duct acoustic modes $\varphi_n(x)$ in Eq. (8), the subscript n ranges from 0 to 50. The finite sine series in Eq. (3) is truncated to $J=80$. Numerical results show that the number of modes is normally enough as further increase in the number does not make significant difference for the purpose of this study.

With the matrix equation (16) solved, the averaged particle velocity over the MPP surface $\bar{u}(x)$ is obtained through the summation defined in Eq. (3). The total external pressure p_e is given as the sum of the blocked pressure $2p_i$ and the radiated pressure p_{rad},

$$p_e = 2p_i + p_{\text{rad}}, \tag{18}$$

where p_{rad} is found with Eq. (8). Thus, the normal incidence sound absorption coefficient for the MPP absorber can be determined as

$$\alpha_N = \frac{\rho_0 c\int_0^L \text{Re}[p_e^* \cdot \bar{u}(x)]dx}{|p_i|^2 L}, \tag{19}$$

where Re means the real part of a complex value and the asterisk denotes the complex conjugate. The surface impedance of the MPP absorber is given as

$$Z_s = \frac{p_e}{\bar{u}(x)}. \tag{20}$$

In the above modeling procedure, the sound wave p_i is assumed to be of normal incidence on the MPP surface. As a result, only normal incidence absorption coefficient α_N is predicted in Eq. (19), which can be conveniently measured using the impedance tube method. The overall modeling procedure, however, applies to situations of oblique incidence and random incidence. In the case of oblique incidence, the blocked pressure $2p_i$ in Eq. (2) becomes a function of the incidence angle, and the external radiation pressure p_{rad} changes accordingly. In the random sound field (diffuse field), sound waves impinge on the MPP at all angles, and the absorption coefficient can be determined as $\bar{\alpha} = \int_0^{\pi/2}\alpha_\theta \sin 2\theta d\theta$, where θ is the incident angle. In order to fully characterize the acoustic properties of the CBMPPA, sound absorption at oblique and random incidences is necessary, but this issue is left to future studies.

B. Model validation

The current theoretical model is first verified against a well-established method in literature where conventional MPP absorbers are involved, for instance, the equivalent electrical circuit method in (Maa, 1987, 1998). The MPP absorber considered here consists of a micro-perforated panel fitted in front of a solid wall with a constant air gap of $D=0.06$ m. The properties of the MPP are $d=t=0.2$ mm and $\sigma=0.5\%$, where d is the orifice diameter, t is the panel thickness, and σ is the perforation ratio. The dashed curve in Fig. 3(a) shows the normal incidence absorption coefficients predicted by the equivalent electrical circuit method. Note that the predicted absorption curve has been shown to compare well with the measured results obtained with a standing wave tube (Maa, 1998). In order to predict the performance of the same MPP absorber using the present theoretical model, the constant air gap between the MPP and the solid wall is simulated as a rectangular cavity of the size 0.24×0.06 m^2, and

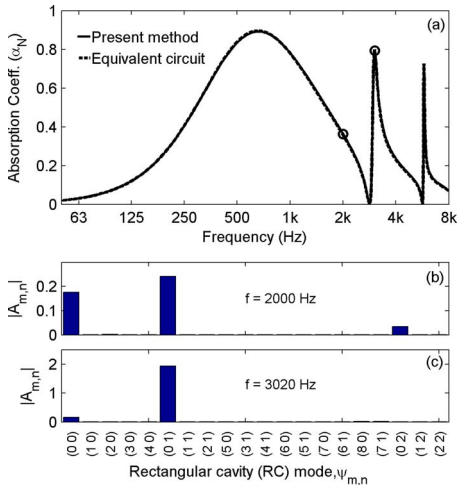


FIG. 3. (Color online) Predicted performance of a rectangular CBMPPA. The rectangular cavity is of the size 0.24×0.06 m^2. The properties of the MPP are $d=t=0.2$ mm and perforation ratio $\sigma=0.5\%$. (a) Comparison of the normal incidence absorption coefficients by the present model and the equivalent electrical circuit method in literature (Maa, 1998). (b) Amplitudes of the first 20 RC modes at frequencies $f=2000$ Hz. (c) Amplitudes of the first 20 RC modes at frequencies $f=3020$ Hz. The two frequencies are marked by open circles in (a) for easy identification.

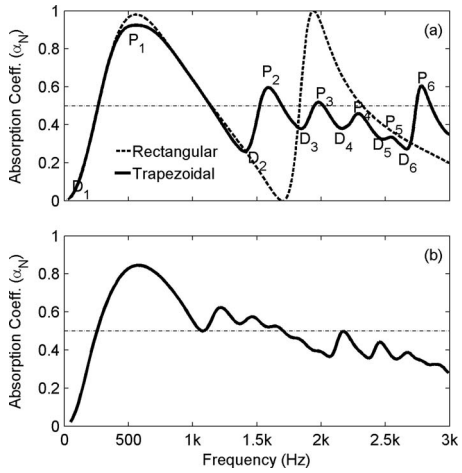


FIG. 4. Predicted absorption performance of the trapezoidal CBMPPAs. (a) Comparison between the trapezoidal CBMPPA (solid curve) and the rectangular CBMPPA of the same cavity volume. The configuration for the trapezoidal cavity is $L_a=0.4$ m, $L_b=0.06$ m, and $\gamma=11.3°$; the rectangular cavity is of the size $L_a=0.4$ m, $L_b=0.1$ m, and $\gamma=0$. (b) Absorption performance of the trapezoidal CBMPPA with a different configuration: $L_a=1$ m, $L_b=0.03$ m, and $\gamma=8.5°$. In both cases, the sound waves are incident on the MPP surface normally.

the infinite panel is reduced to a beam of finite length $L=0.24$ m. The simulated absorption coefficients with the present model are shown in Fig. 3(a) in solid curve. It can be seen that the predicted absorption curves by the two methods match with each other very well. Figures 3(b) and 3(c) show the amplitudes of the first 20 RC modal response at frequencies $f=2000$ and 3020 Hz, which are marked by open circles in Fig. 3(a) for reference. At $f=2000$ Hz, which lies in between the first and the second spectral peaks of the absorption curve, the (0,0) and (0,1) RC modes dominate the acoustic pressure inside the rectangular cavity. The (0,2) RC mode can also be observed, but its contribution is relatively small. At $f=3020$ Hz, which approximately corresponds to the second spectral peak of the absorption curve, only the (0,1) RC mode dominates. The (0,0), (0,1), and (0,2) RC modes correspond to the first three resonances of the infinite air gap in (Maa, 1987, 1998), indicating that the present model is consistent with the equivalent electrical circuit method. As mentioned in the Introduction, other cavity modes are almost not excited by the normal incident sound wave p_i since they are decoupled from the air motion in the micro-perforations.

III. NUMERICAL RESULTS

Based on the theoretical model developed in Sec. II, the normal incidence sound absorption properties of a MPP backed by a trapezoidal cavity are investigated numerically. The following set of parameters is used as the default values for the irregular-shaped CBMPPA:

$$d=0.3\text{ mm},\quad t=0.3\text{ mm},\quad \sigma=1,$$

$$L_a=0.4\text{ m},\quad L_b=0.06\text{ m},\quad \gamma=11.3°, \tag{21}$$

where d, t, and σ specify the properties of the perforated panel, and L_a, L_b, and γ define the geometrical configuration of the backing cavity. The choice of the default set of parameters in Eq. (21) is only to facilitate the analysis on the absorption mechanism of the irregular-shaped MPP absorber. No particular effort is made to seek optimal performance of the CBMPPA at the present stage.

A. Absorption performance

Figure 4(a) compares the absorption curves of the trapezoidal CBMPPA specified in Eq. (21) and a rectangular MPP absorber of the same cavity volume (i.e., $L_a=0.4$ m, $L_b=0.1$ m, and $\gamma=0$). The dashed curve shows the results for the rectangular absorber. Two spectral peaks can be observed within the frequency range of interest, and a dip occurs in between the two peaks, that is, around 1700 Hz. Since the absorption at the dip is negligible, it is the first peak that often determines the effective bandwidth for the conventional rectangular MPP absorber. The solid curve shows the absorption coefficients for the trapezoidal CBMPPA. In contrast with the two-peak spectrum for the rectangular MPP absorber, six distinct spectral dips (D_1–D_6) and peaks (P_1–P_6) are observed for the trapezoidal CBMPPA. The increased number of spectral peaks implies that more cavity modes are excited than in the case of the rectangular cavity. Around the first spectral peak P_1, the two absorbers demonstrate similar absorption performance. The second peak P_2 appears where the absorption performance of the conventional MPP absorber is very low, which makes it possible to achieve an absorber with wider effective bandwidth. The appearance of the following peaks P_3–P_6 maintains the absorption performance at a reasonably high level within the whole frequency range of interest. At the dips D_2–D_6, which roughly correspond to the resonances of the hard-walled trapezoidal cavity, a moderate sound absorption level is observed. This contrasts with the rectangular MPP absorber that has little absorption at the resonances of the backing cavity, as shown in Fig. 4(a) by the dashed curve at frequency f

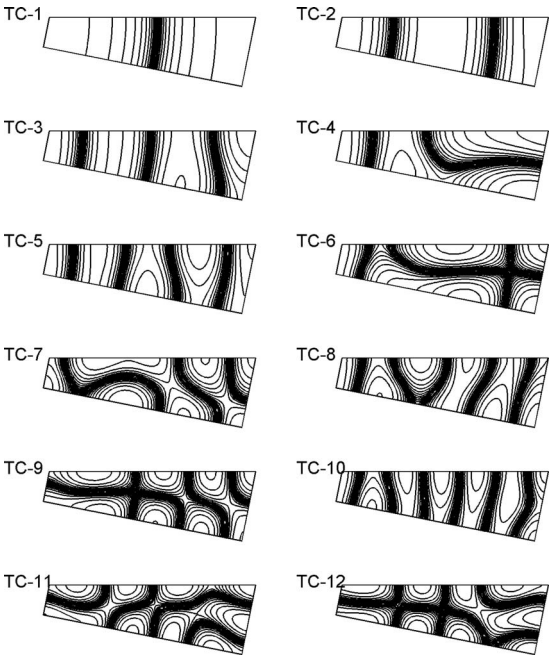


FIG. 5. The spatial distribution of pressure of the first 12 non-zero TC modes of the default trapezoidal cavity. The sound pressure level difference between contours is 3 dB.

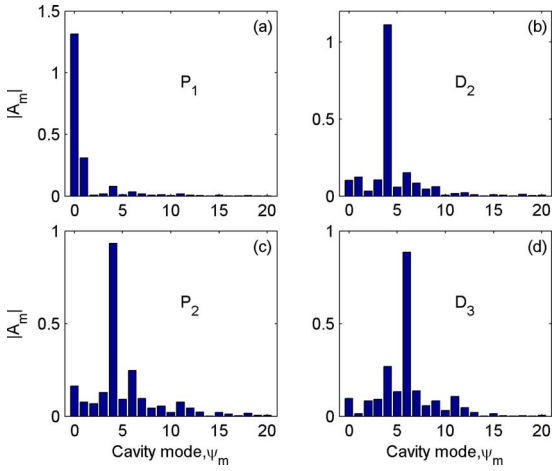


FIG. 6. (Color online) Amplitudes of the first 20 TC modes toward the total sound field in the backing cavity at the spectral peaks and dip frequencies. (a) The first peak, P_1. (b) The second dip, D_2. (c) The second peak, P_2. (d) The third peak, D_3.

=1700 Hz. It is worth noting that the occurrence of the peaks and dips can also be explained by the impedance matching conditions as in Maa (1998), but the impedance match only occurs in a local manner in the trapezoidal CBMPPA. An example of such local characteristics can be found in Fig. 7, in which the local impedance matching conditions for P_2 and D_2 are illustrated.

Note that the default set of parameters is not the optimum. A change in the geometrical configuration of the backing cavity can significantly alter the vibroacoustic coupling between the cavity and the MPP, and hence a better absorption performance might be achieved in terms of the effective bandwidth. Figure 4(b) shows the absorption curve for a different cavity configuration (L_a=1 m, L_b=0.03 m, and γ =8.5°) with the same MPP properties. The absorption coefficients at the first three dips in the curve are elevated to above half-absorption level, which gives rise to a wider absorption bandwidth. Therefore, a parametric study is necessary to systematically evaluate the effect of various design parameters on the absorption performance of the irregular-shaped CBMPPA, but this issue is left for future research.

Figure 5 depicts the first 12 nonzero modes of the trapezoidal cavity (TC modes) denoted as TC-1, TC-2,…, TC-12, respectively. It can be shown that these modes are evolved from individual or a family of RC modes of a corresponding rectangular cavity (Li and Cheng, 2004; Sum and Pan, 2006). The corresponding RC modes can be categorized into two types depending on whether the RC modes graze the inclined MPP or not. The TC modes TC-1, TC-2, TC-3, TC-5, TC-8, and TC-10 are evolved from the (1,0), (2,0), (3,0), (4,0), (5,0), and (6,0) RC modes, respectively. These RC modes graze the inclined MPP, and the evolved TC modes still nearly graze the inclined panel. The rest of the TC modes is evolved from RC modes that do not graze the inclined MPP. As more than one RC mode is involved significantly, the resulting TC mode has a relatively complicated mode shape. It is found that the two types of TC modes, i.e., the TC modes evolved from the grazing RC modes and those evolved from the nongrazing RC modes, contribute differently to the sound absorption of the trapezoidal MPP absorber.

The relationship between the free vibration characteristics of the trapezoidal cavity and the absorption performance of the trapezoidal CBMPPA is examined. Table I lists the resonance frequencies for the 12 TC modes shown in Fig. 5. It can be observed that the resonance frequencies for TC modes TC-4, TC-6, TC-7, TC-9, and TC-11 roughly coincide with the dips D_2–D_6 in Fig. 4(a). The TC mode TC-12 also corresponds to a spectral dip but it is beyond the frequency range investigated. It can also be observed that all of these TC modes are evolved from nongrazing RC modes. Figure 6 illustrates the contributions of each TC mode toward the total sound field in the backing cavity at the selected spectral peaks and dip frequencies. At the peak P_1, the zero TC mode (which is not shown in Fig. 5) dominates over other cavity modes, as shown in Fig. 6(a). At the dip D_2 and the second peak P_2, the fourth TC mode plays a significant role in the total sound field, cf. Figs. 6(b) and 6(c). Such an observation can be repeated for all the other spectral peaks and dips. That

TABLE I. The first 12 resonant frequencies for the trapezoidal cavity.

TC mode	1	2	3	4	5	6	7	8	9	10	11	12
Frequency (Hz)	437	857	1277	1385	1704	1793	2095	2143	2389	2556	2635	2677

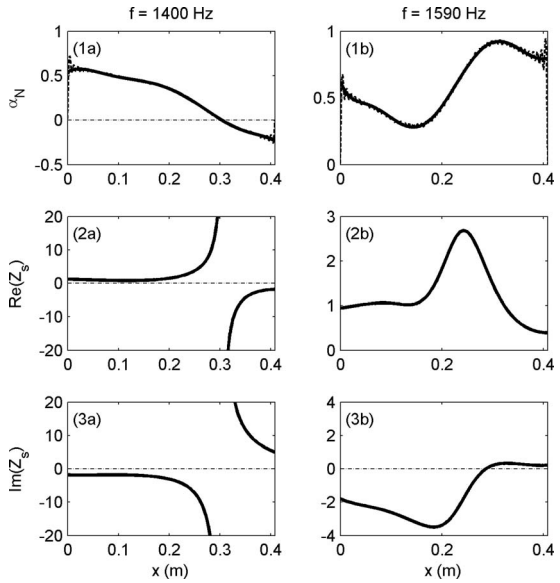


FIG. 7. Local characteristics of absorption performance of the trapezoidal CBMPPA. The first row is for the absorption coefficients. The second and third rows are for the real part and the imaginary part of the surface impedance of the MPP absorber, respectively. Left column: $f=1400$ Hz; right column: $f=1590$ Hz.

is, the mode TC-6 dominates at D_3 and P_3, TC-7 dominates at D_4 and P_4, and so on. This feature is similar to that of a conventional MPP absorber with constant air gap. However, more cavity modes participate in the sound absorption process of the trapezoidal CBMPPA. Recall that only the RC modes such as (0,0), (0,1),(0,2),…, contribute to the sound absorption of the rectangular MPP absorber, as already shown in Fig. 3. For the trapezoidal cavity, all the TC modes evolving from non-grazing RC modes participate in the sound absorption of the MPP absorber. In other words, acoustic modes that are initially decoupled from the MPP in the rectangular configuration are now coupled with the air motion in the micro-perforations, and the coupling between the structure (MPP) and the enclosure (cavity) is greatly enhanced due to the inclination of the cavity wall.

B. Local sound absorption characteristics

Another distinct feature of the trapezoidal CBMMPA is its local distribution characteristics of the sound absorption coefficients across the micro-perforated surface. For the conventional MPP absorber with constant air gap, it is known that its absorption properties are the same over the whole micro-perforated surface as long as the perforation parameters are uniform. In the trapezoidal CBMPPA, the absorption properties vary across the micro-perforated panel even though the perforation parameters are the same. By removing the integration over L in Eq. (19), the distribution of the normal incidence absorption coefficients can be calculated as

$$\alpha_N(x) = \frac{\mathrm{Re}[p_e^*(x)\cdot \bar{u}(x)]\rho_0 c}{|p_i|^2}. \tag{22}$$

Figures 7(1a) and 7(1b) show the distribution of the sound absorption coefficients at frequencies $f=1400$ and 1590 Hz, respectively. The two frequencies correspond to the dip D_2 and the peak P_2 shown in Fig. 4(a). The dashed lines are for the sound absorption coefficients calculated by Eq. (22) directly. Overshoots and small ripples exist in the absorption curves as the particle velocity $\bar{u}(x)$ is approximated with a finite series of sine waves, cf. Eq. (3). These artifacts are smoothed out in the least square sense so that the trend of variation will not be obscured, as shown by the solid lines in Figs. 7(1a) and 7(1b). It is worthy to note that the absorption coefficients become negative when $x>0.3$ m, which implies that some acoustic energy enters the trapezoidal cavity from other portion of the micro-perforated panel and is radiated out at the region $x>0.3$ m. That is to say that acoustic energy transfer may occur within the trapezoidal cavity. By the same token, the local absorption coefficients can be larger than unity at certain frequencies, though such a phenomenon is not demonstrated at the two particular frequencies.

The local absorption characteristics of the trapezoidal CBMPPA can be attributed to its varying impedance matching conditions between the MPP absorber and the incident sound field. The second and third rows of Fig. 7 show the real and imaginary parts of the surface impedance Z_s over the micro-perforated panel, respectively. For the trapezoidal CBMPPA, the real parts of the surface impedance manifest not only the resistance of the micro-perforated panel itself but also the effect of acoustic energy transfer within the trapezoidal cavity, which differs from the conventional MPP absorber with constant air gap. As a result, the real parts of the surface impedance Z_s is not distributed uniformly, and even negative values are observed under certain situations, cf. Fig. 7(2a). The imaginary part of the surface impedance Z_s characterizes the reactance of the MPP absorber, a small value of which usually means a good impedance matching between the absorber and the incident noise field. As shown in Fig. 7(3a), the local impedance mismatch at the region $x>0.25$ m explains the poor absorption performance at the same region in Fig. 7(1a). Likewise, the good impedance matching at $x>0.25$ m in Fig. 7(3b) accounts for the high absorption performance shown in Fig. 7(1b). As already shown in Fig. 6, the sound field inside the backing cavity is dominated by the cavity modes TC-4 for both $f=1400$ Hz (D_2) and $f=1590$ Hz (P_2). The region $x>0.25$ m roughly corresponds to the portion of the sound field in which the acoustic wave is perpendicular to the micro-perforated panel, which suggests the close relationship between the varying impedance conditions and the distorted TC modes.

IV. EXPERIMENTAL STUDY

The main objective of the experimental study is to validate the basic theoretical model established. To this end, the normal incidence sound absorption coefficients of a prototype CBMPPA are measured experimentally and compared with the theoretical predictions. The experimental rig is designed based on the two-microphone transfer-function method according to ISO-10534-2 (ISO, 1998), as shown in Fig. 8(a). The test rig consists of a straight rectangular duct (15-mm-thick acrylic) so that it can be considered to be acoustically rigid. A loudspeaker is connected at the left end as the excitation source. A prototype CBMPPA to be tested is

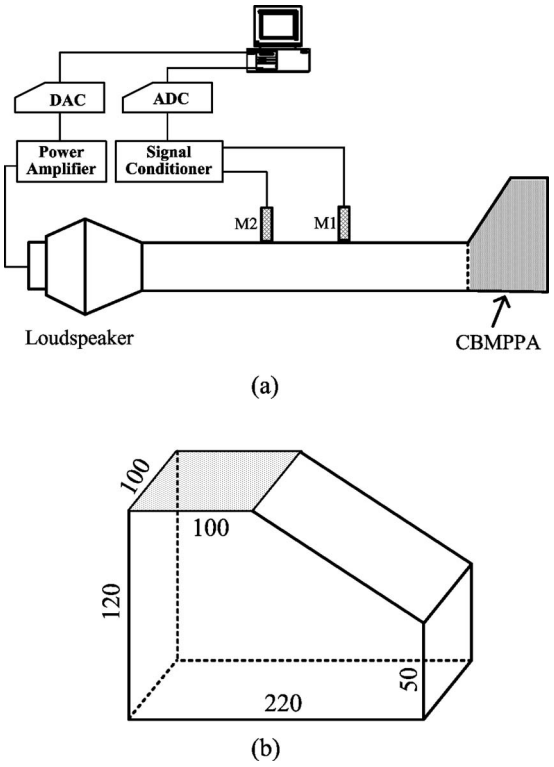


FIG. 8. Schematic of the experimental rig. (a) The two-microphone measurement system. (b) Configuration of the prototype CBMPPA (dimensions in millimeters). The MPP of the size 100×100 mm^2 is located at the top surface.

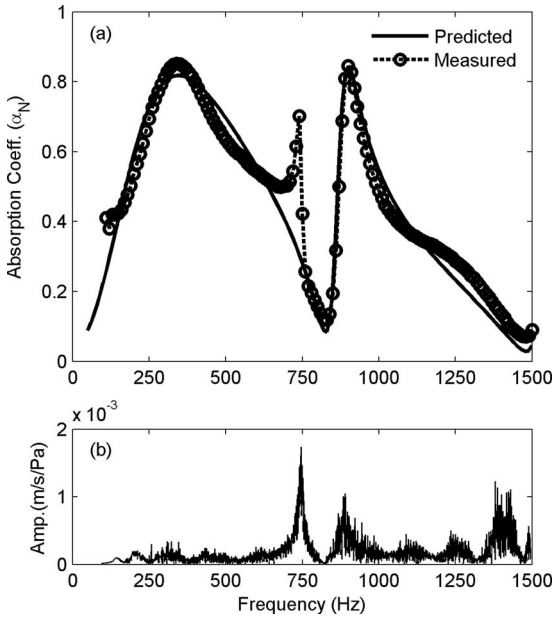


FIG. 9. Performance of the prototype CBMPPA. (a) Comparison between the measured and predicted sound absorption coefficients. (b) Amplitude of the measured frequency response of the MPP at (30 mm, 70 mm). The whole perforated area is of the size 100×100 mm^2. The reference sound pressure is measured about 3 mm in front of the panel in the duct.

installed at the other end of the duct. Two microphones M1 and M2 (B&K type 4187) measure the sound pressure inside the duct. The distance between the two microphones is properly selected according to the frequency range to be measured. From the transfer function between the two microphones, the surface impedance and the normal incidence absorption coefficient of the prototype CBMPPA can be determined. The cross section of the rectangular duct is of 100×100 mm^2. So, the first cut-on frequency in the rectangular duct is around 1700 Hz. Due to the limitation of the cross section of the test duct, it is hard to design a CBMPPA with the configuration shown in Fig. 2 to demonstrate the effect of the irregular-shaped cavity within the measurement frequency range of the test rig (say, lower than 1700 Hz). For this reason, the prototype CBMPPA has a configuration, as shown in Fig. 8(b). The cavity walls (other than the MPP) are also made of 15-mm-thick acrylic. The properties of the MPP [at the top surface in Fig. 8(b)] are

$$d = 0.415 \text{ mm}, \quad t = 0.46 \text{ mm}, \quad \sigma = 1.4, \tag{23}$$

where d is the orifice diameter, t is the thickness, and σ is the perforation ration in percentage. The MPP is made of aluminum. With the geometrical configurations shown in Fig. 8(b) and the MPP properties in Eq. (23), the absorption coefficients of the prototype CBMPPA are predicted using the theoretical model established in Sec. II.

Figure 9(a) compares the predicted and measured normal incidence absorption coefficients of the prototype CBMPPA. The solid curve shows the predicted results. Within the frequency range investigated, two peaks can be observed at frequencies $f=350$ and 900 Hz, respectively. As analyzed in Sec. III, the occurrence of the second peak is attributed to the coupling between the MPP and the distorted cavity mode that are initially decoupled with each other in a rectangular configuration. Note that, for this particular design, the absorption performance at the "dip" frequency between the two peaks is not as good as that shown in Fig. 4, but it suffices to serve the objective of model validation. The dashed curve with open circles shows the measured absorption coefficients. As depicted in Fig. 9(a), the experimental result matches the theoretical predictions very well except that an extra absorption peak appears at $f=740$ Hz. This extra peak is believed to be caused by the sound-induced panel vibration which is excluded in the current theoretical model. Figure 9(b) shows the amplitude of the measured frequency response of the MPP at the position (30 mm, 70 mm) with the reference sound pressure measured 3 mm in front of the panel in the duct. An obvious resonance peak can be observed at $f=740$ Hz, indicating that strong panel vibration occurs. The corresponding deflection shape of the MPP is measured using a scanning vibrometer (PSV-400), and the panel vibration is found to be dominated by a strong volume-displacing vibration mode at this frequency. The possible mechanism involved for the effective sound absorption at this frequency might be the increased relative air particle velocity with respect to the panel vibration at the micro-perforations. Detailed discussion on the panel vibration effect can be found in Lee *et al.* (2005). As a summary, the experimental study demonstrates that the basic theoretical model of the irregular-shaped CBMPPA is correct. Sound-induced vibration of the panel itself may influence the absorption performance significantly and should be included in the theoretical modeling in future studies, especially when light material is used to fabricate the MPP.

V. CONCLUSIONS

Owing to the altered vibro-acoustic coupling pattern, the sound absorption characteristics of a micro-perforated panel backed by an irregular-shaped cavity can be quite different from that of a conventional MPP absorber with constant air gap. From this observation, an irregular-shaped CBMPPA is proposed aiming for enhanced sound absorption performance. A 2D theoretical model is established by considering the full coupling among the MPP, the backing cavity, and the exterior sound field. Based on the verified theoretical model, the normal incidence absorption properties of a trapezoidal CBMPPA are investigated. The following conclusions can be drawn.

(1) The irregular-shaped backing cavity can significantly alter the sound absorption mechanisms and frequency distribution of overall sound absorption coefficients of the MPP absorber. Compared with the conventional MPP absorber with constant air gap, the irregular-shaped CBMPPA can produce more spectral peaks and achieve enhanced absorption performance at the troughs in the absorption curve. This property can be further explored to tune the effective absorption range of the CBMPPA by making appropriate changes to cavity shapes and dimensions. Meanwhile, it also suggests that the shape of the backing cavity could be used for future optimal design of the MPP absorber.

(2) Due to the inclination of the MPP, all TC modes that evolve from nongrazing RC modes are coupled to the air motion inside the micro-perforations, which accounts for the altered absorption pattern of the CBMPPA. Take the RC modes (m,n) of a 2D rectangular cavity for example. In a conventional MPP absorber with constant air gap, only the RC modes $(0,n)$ with $n=0,1,2,\ldots$, are coupled to the MPP. By replacing the rectangular cavity with a trapezoidal cavity, all TC modes evolving from both the RC modes $(0,n)$ with $n=0,1,2,\ldots$, and the RC modes (m,n), with m and $n>0$ are coupled to the MPP. The sound field at the spectral peaks and dips in the absorption curves are found to be dominated by a specific TC mode. The peak frequencies correspond to the resonant frequencies of the panel-cavity system, while the frequencies at the dips coincide roughly with the resonance frequencies of the rigid cavity alone.

(3) Unlike the uniform absorption performance over the MPP surface of the conventional MPP absorber, the irregular-shaped CBMPPA exhibits obvious local characteristics. Absorption coefficients less than zero and larger than unity can be observed at certain portion of the MPP surface due to the energy transfer within the backing cavity. The local absorption characteristics of the irregular-shaped CBMPPA are attributed to its varying impedance matching conditions, which in turn is closely related to the distorted acoustic mode shapes of the irregular-shaped cavity.

(4) The absorption coefficients of a prototype CBMPPA are measured using a standing wave tube and compared with the theoretical predictions. The experimental results are in good agreement with the theoretical predictions on the whole, demonstrating that the basic theoretical model is correct. Extra absorption peak is observed in the experiment, which is attributed to the sound-induced vibration of the MPP itself.

ACKNOWLEDGMENTS

The authors wish to acknowledge a grant from Research Grants Council of Hong Kong Special Administrative Region, China (Project No. PolyU 5140/09E) and supported by the Central Research Grant of The Hong Kong Polytechnic University through Grant No. G-U553.

Asdrubali, F., and Pispola, G. (**2007**). "Properties of transparent sound absorbing panels for use in noise barriers," J. Acoust. Soc. Am. **121**, 214–221.

Atalla, N., and Sgard, F. (**2007**). "Modeling of perforated plates and screens using rigid frame porous models," J. Sound Vib. **303**, 195–208.

Bies, D., and Hansen, C. (**1996**). *Engineering Noise Control: Theory and Practice*, 2nd ed. (E&FN Spon, New York).

Cheng, L., Li, Y. Y., and Gao, J. X. (**2005**). "Energy transmission in a mechanically-linked double-wall structure coupled to an acoustic enclosure," J. Acoust. Soc. Am. **117**, 2742–2751.

Doak, P. E. (**1973**). "Excitation, transmission and radiation of sound from source distributions in hard-walled ducts of finite length. I. The effects of duct cross-section geometry and source distribution space-time pattern," J. Sound Vib. **31**, 1–72.

Droltleff, H., and Zhou, X. (**2001**). "Attractive room acoustic design for multi-purpose halls," Acta Acust. **87**, 500–504.

Fuchs, H. (**2001**). "Alternative fibreless absorbers—New tools and materials for noise control and acoustic comfort," Acta Acust. **87**, 414–422.

Fuchs, H. V., and Zha, X. (**1997**). "Acrylic-glass sound absorbers in the plenum of the Deutscher Bundestag," Appl. Acoust. **51**, 211–217.

ISO 10534-2 (**1998**). "Determination of sound absorption cefficient and impedance in impedance tubes," International Organisation for Standardization.

Kang, J., and Brocklesby, M. W. (**2005**). "Feasibility of applying microperforated absorbers in acoustic window system," Appl. Acoust. **66**, 669–689.

Kang, J., and Fuchs, H. V. (**1999**). "Predicting the absorption of open weave textiles and microperforated membranes backed by an air space," J. Sound Vib. **220**, 905–920.

Lee, D. H., and Kwon, Y. P. (**2004**). "Estimation of the absorption performance of multiple layer perforated panel systems by transfer matrix method," J. Sound Vib. **278**, 847–860.

Lee, Y. Y., Lee, E. W. M., and Ng, C. F. (**2005**). "Sound absorption of a finite flexible micro-perforated panel backed by an air cavity," J. Sound Vib. **287**, 227–243.

Li, Y. Y., and Cheng, L. (**2004**). "Modifications of acoustic modes and coupling due to a leaning wall in a rectangular cavity," J. Acoust. Soc. Am. **116**, 3312–3318.

Maa, D. Y. (**1975**). "Theory and design of microperforated-panel sound-absorbing construction," Sci. Sin. **18**, 55–71.

Maa, D. Y. (**1987**). "Microperforated panel wide-band absorber," Noise Control Eng. J. **29**, 77–84.

Maa, D. Y. (**1998**). "Potential of microperforated panel absorber," J. Acoust. Soc. Am. **104**, 2861–2866.

Meirovitch, L. (**2001**). *Fundamentals of Vibrations*, McGraw-Hill, Boston, MA.

Pan, J., Ming, R., and Guo, J. (**2004**). "Wave trapping barriers," in Proceedings of the Acoustics 2004, Gold Coast, Australia, pp. 283–287.

Putra, A. (**2007**). "Sound radiation from a perforated plate," ISVR Technical Memorandum No. 973, University of Southampton, UK.

Sum, K. S., and Pan, J. (**2006**). "Effects of the inclination of a rigid wall on the free vibration characteristics of acoustic modes in a trapezoidal cavity," J. Acoust. Soc. Am. **119**, 2201–2210.

Watts, G. (**1996**). "Acoustic performance of parallel traffic noise barrier," Appl. Acoust. **24**, 95–119.

Wu, M. Q. (**1997**). "Micro-perforated panels for duct silencing," Noise Control Eng. J. **45**, 69–77.

Suitable reverberation times for halls for rock and pop music[a)]

Niels Werner Adelman-Larsen[b)]
Flex Acoustics, Diplomvej, Bygn. 377, DK-2800 Kgs. Lyngby, Denmark

Eric R. Thompson[c)]
Department of Electrical Engineering, Acoustic Technology, Technical University of Denmark, Building 352, DK-2800 Kgs. Lyngby, Denmark

Anders C. Gade
Gade & Mortensen Akustik A/S, Hans Edvard Teglers Vej 5, DK-2920 Charlottenlund, Denmark

(Received 2 December 2008; revised 9 October 2009; accepted 14 October 2009)

The existing body of literature regarding the acoustic design of concert halls has focused almost exclusively on classical music, although there are many more performances of popular music, including rock and pop. Objective measurements were made of the acoustics of 20 rock music venues in Denmark and a questionnaire was used in a subjective assessment of those venues with professional rock musicians and sound engineers as expert listeners. Correlations between the measurements show that clarity, including bass frequencies down to 63 Hz, is important for the general impression of the acoustics of the hall. The best-rated halls in the study have reverberation times that are approximately frequency independent from 0.6 to 1.2 s for hall volumes from 1000 to 6000 m^3. The worst rated halls in the study had significantly higher reverberation times in the 63 and 125 Hz bands. Since most audiences at rock concerts are standing, absorption coefficients were measured with a standing audience from 63 Hz to 4 kHz. These measurements showed that a standing audience absorbs about five times as much energy in mid-/high-frequency bands as in low-frequency bands.
 [DOI: 10.1121/1.3263611]

PACS number(s): 43.55.Fw, 43.55.Gx, 43.55.Hy, 43.55.Br [LMW] Pages: 247–255

I. INTRODUCTION

There is a long and great tradition of acoustical design for concert halls, evidenced by a large number of books written on the subject, such as from Barron,[1] Beranek,[2] and Hoffman *et al.*[3] There have been a number of studies published on the acoustics of concert halls, mainly seeking correlations between objective measurements and subjective perception of the acoustics.[4–9] This large body of literature focuses entirely on the recommended acoustics for performances of western classical music, including organ music, chamber music, orchestral music, and opera. Even when multipurpose halls are discussed, the purposes intended are usually classical music and spoken-word performances. A large portion of music performances and performance spaces has been ignored in the literature, namely, those of more popular music, such as rock, pop, jazz, country, and others.

A recent survey of performance spaces in Denmark, not including classical music halls or large outdoor arenas, estimated that there had been approximately 12 500 live concerts with about 2 million spectators within the popular music genres during 2004.[10] During the same time frame, the Danish Arts Agency (Kunststyrelsen) estimated (conservatively large) that there were 2000 classical music concerts in Denmark. This means that there were at least five times more popular music concerts than classical music concerts, and a corresponding ratio of spectators. It is perhaps strange, in light of these numbers, that there has not been more focus on the acoustics of halls for popular music concerts.

Various authors have written loosely about recommendations for acoustics for rock music. For example, in his book[1] (Chap. 10), Barron mentioned that a highly absorbent acoustic space is preferred for popular music and recommends a reverberation time below 1 s, but there does not appear to be a scientific study behind these recommendations. The present study seeks to fill the void and investigate the optimum acoustic conditions for performances of popular music, particularly for rock music. First, the relevance of the study had to be established by determining whether the acoustics of the performance space play any role in the quality of the experience perceived by the band and the sound engineers. Then, a more comprehensive subjective study was conducted and objective measurements were made on 20 Danish popular music venues using techniques similar to those from Barron.[5] For the purposes of the present study, it was assumed that the objective acoustic parameters traditionally used to characterize classical music halls can also be applied to rock concert halls. The venues studied were small to medium-sized halls from about 600–7000 m^3. There are many rock concerts held in larger spaces (e.g., sports arenas), but those are beyond the scope of the present study. The

[a)]Portions of this study were presented at the 155th Meeting of the Acoustical Society of America (Acoustics '08) and at the 123rd Audio Engineering Society Convention.
[b)]Author to whom correspondence should be addressed. Electronic mail: nwl@flexac.com
[c)]Present address: Hearing Research Center, Boston University.

subjective and objective parameters were then correlated in order to determine what is important for good acoustics for popular music performances.

A. The musical genre

The popular music genre encompasses a broad range of sub-genres, including rock, pop, jazz, punk, and country. These sub-genres have similarities which differentiate them from classical music, but also differences, including instrumentation, frequency balance, and typical sound levels at concerts, which may mean that different acoustic design is desired for the different sub-genres. In order to limit the scope of the current study, the rock and pop sub-genres were chosen as the focus. The two sub-genres are very similar and together make up a large portion of the total popular concerts performed. For the sake of brevity in the paper, the two sub-genres will be grouped together and referred to simply as rock music.

Some of the differences between rock and classical music concerts, which might lead to different requirements for the acoustics of the performance space, will be explained in Secs. I A 1–I A 5.

1. Instrumentation and sound sources

Classical music concerts usually are performed by standard orchestral instruments, from the double bass, tuba, and bass drum in the low registers to the violin, piccolo, and triangle, for instance, in the high-frequency range. The instruments are typically not amplified but rather rely on the acoustics of the space to carry the sound to the audience. Rock music, on the other hand, is typically performed on electric and/or acoustic bass and guitars, keyboards, and drum sets. The sound is almost always electrically amplified, so that the sound sources are the on-stage amplifiers and the hall's loudspeaker system (PA), which plays a mix of the instruments controlled by a sound engineer. The PA-system is used to create a desired mix of the instruments with a sufficient sound level, so that the sound carries to all audience members. Not all halls provide PA equipment, requiring bands to bring their own, and some bands only use their own PA equipment, in order to have greater control over the sound. In addition to the PA-system, many sound processors, such as EQ, compression, delay, and artificial reverberation, are used to create the final sound image in a rock concert. The delay and artificial reverberation are often used on vocals, guitars, keyboards, snare drums, tom-toms, and high-hats, but it is often recommended to high-pass filter the signals before applying artificial reverberation in order to keep the bass drum and bass guitar from becoming too smeared and indistinct.[11] The sound that arrives at the listener is a mixture of the sound from the PA and the sound of the reverberant field in the hall. Listeners that are farther from the sound source will experience more of the acoustics of the hall than those within the critical distance[12] from the sources. Of course, the perceived sound quality in the hall will be affected by the quality of the PA-system and the effects that are applied to the sound, but those factors are beyond the scope of the present study.

2. Stage acoustic requirements

A group of musicians performing any kind of music together need to be able to hear each other in order to play as a cohesive unit.[13,14] This creates challenges in the acoustic design of a classical music stage to provide support through early reflections, particularly for large orchestras. Most rock performers make use of monitors, either on-stage, in-ear or headphones, or a combination of those, which provide a mix of all the instruments and voices to each of the performers. The on-stage monitors are simply loudspeakers set up close to the musician, while the in-ear monitors are generally sound-attenuating earphones. The sound from on-stage monitors is picked up by the microphones, causing unwanted sound in the mix and sometimes feedback. In-ear monitors do not create feedback loops with the microphones, but they attenuate sound from the hall for the musicians. This can mitigate any negative effects of poorly designed acoustics on their performance, but their contact with the audience is reduced, because the audience's audible reaction to the music is also attenuated.

3. Audience

During classical concerts, the audience is usually seated in chairs with a density of about 2 people/m^2.[1] On the other hand, audiences at rock concerts are usually standing and more densely packed, with a density of about 2.7 people/m^2, estimated from a measurement performed to determine the absorption of a standing audience, described below. The difference between standing and seated audiences affects the absorption characteristics and the effective volume of the hall.

4. Listeners

For both classical and rock music, the musicians have the best perspective for judging the acoustics on the stage and have been used in prior subjective studies of classical music halls.[13,14] Rock music concerts also typically have a sound engineer, who actively controls the sound during the concert and can be considered an expert listener with the perspective of the audience. Therefore, for the present study, musicians and sound engineers were asked their subjective impression of the acoustics of the halls.

5. Balance

At a typical classical concert, most of the sound energy is in the mid-/high-frequency range, where the mean sound pressure level in the mid-frequencies may be 10 dB higher than the bass.[15] At a typical rock concert, the balance is usually quite different with 20–30 dB higher sound levels in the bass than in the mid-frequency range.[16] With so much more acoustic energy in the bass, it seemed important to consider the acoustics of the halls in the bass range for this study. Many reports of the acoustics of halls or on acoustic treatments only consider frequencies from 125 Hz and upwards.[2,3] For purposes of rock music, it may be required to specify halls and materials including the 63 Hz octave band.

This study was aimed at determining whether acoustics are important for rock concert halls and how the acoustics of

TABLE I. Details of the twenty surveyed concert halls as of 2005. Note that several halls have been renovated since this study. $T_{30,B}$ and $T_{30,M/T}$ are the reverberation times in the audience area in the 63–125 Hz octave bands and in the octave bands from 250 Hz to 2 kHz, respectively. The EDT was measured on the stage in the 63–2000 Hz octave bands. D_{50} was measured in the audience area and the 63–2000 Hz octave bands. The BR is the ratio of the average reverberation time in the 63 and 125 Hz bands to the average reverberation time in the 0.5–2 kHz octave bands. The general ratings are the ordinal positions of the mean general ratings across musicians, across sound engineers, and combined with equal weighting of the two groups.

Name	Volume (m^3)	Audience capacity	$T_{30,B}$ (s)	$T_{30,M/T}$ (s)	EDT (s)	D_{50}	BR	General ratings		
								Mus.	SEng.	Comb.
Rytmeposten	655	300	0.8	0.8	0.3	0.6	1.0	14	10	11
Lille Vega	785	500	0.5	0.7	0.4	0.7	0.7	1	6	3
Loppen	890	350	0.9	0.8	0.5	0.7	1.2	5	13	9
Skråen	1100	375	1.5	0.8	0.9	0.4	1.8	13	12	13
Paletten	1420	375	1.0	0.9	0.7	0.7	1.0	8	8	8
Stars	1440	400	0.6	0.6	0.3	0.8	0.9	1	1	4
Voxhall	1600	500	0.9	0.6	0.5	0.7	1.3	7	5	6
Sønderborghus	1600	420	1.2	1.0	0.8	0.6	1.3	20	18	19
Musikhuzet	2080	700	1.1	0.9	1.1	0.5	1.2	9	17	12
Godset	2150	700	0.7	0.8	0.5	0.6	0.8	6	4	5
Magasinet	2540	525	1.9	1.3	1.3	0.3	1.4	12	19	18
Pumpehuset	3000	600	1.2	1.1	1.0	0.6	0.9	16	15	15
Forbrændingen	3050	450	1.1	0.9	0.5	0.8	1.2	19	11	14
Train	3300	900	0.8	1.0	0.4	0.7	0.9	3	2	2
Slagelse	3800	700	1.8	1.6	1.0	0.5	1.1	17	20	20
Viften	3950	700	2.6	1.2	1.1	0.6	2.0	18	14	16
Amager Bio	4500	1000	1.2	1.0	0.8	0.6	1.1	4	9	7
Torvehallerne	5400	700	1.2	1.5	0.9	0.5	0.8	15	16	17
Store Vega	5800	1430	1.4	1.2	0.7	0.7	1.1	2	3	1
Tobakken	6500	1200	1.5	1.0	0.8	0.6	1.4	11	7	10

a hall should be designed for the best experience for the audience and the musicians. The literature on the acoustics of classical music halls has many key objective metrics and determined their correlation with perceived sound quality.[2] For this first study of the acoustics of rock/pop music halls, the main focus was to determine recommendations for the reverberation time T_{30} as a function of frequency and hall volume for small- and medium-sized halls.

II. METHODS

A survey was conducted on the perceived acoustics of 20 rock concert halls around Denmark and objective measurements were carried out in each hall. A list of the halls and some acoustic measurements from them are shown in Table I. The halls were either dedicated or multi-purpose halls, where many of the dedicated halls were converted to cinemas or industrial halls, and many of the multi-purpose halls were cultural centers, common around Scandinavia, that may also host theater productions, for example.

A. Subjective survey

There have been three primary methods for conducting subjective studies of the acoustics of concert halls for classical music, each with its advantages and disadvantages. One method has been to create a virtual concert hall in a laboratory, where either recordings from halls or simulations from a room acoustics program (e.g., ODEON and EASE) are presented to the listeners. The acoustics could either be simulated with an array of loudspeakers in an anechoic chamber[17,18] or auralized and presented over headphones.[9] With this method, listeners can quickly rate many halls without having to travel great distances. Some other benefits are that halls can be presented anonymously and blindly so that there is no bias based on a hall's reputation or visual appeal, and the exact same performance of a piece can be evaluated in all halls and positions. Despite these advantages, it can be difficult to get truly qualified listeners, e.g., professional musicians with their busy schedules, to participate in a laboratory experiment.

As an alternative, listening tests can be performed in an existing hall that has the possibility of changing the acoustics.[5,13] However, the changes that can be made to the acoustics in such a hall are often quite limited and the results may not be generalizable to all halls.

Surveys have also been done on existing halls through interviews of people who have experience with the acoustics in many halls.[19] Unfortunately, acoustic memory can be very short and is likely to be colored by many non-acoustic factors, which can make the comparison of halls with this method less precise than if the sound impression from the halls can be presented quickly in sequence in the laboratory. In spite of this challenge, this method was selected for the subjective evaluation of the halls in the present study, primarily because it was deemed to be important that the evaluation of the halls was made by people who had great experience working in and listening to the halls, namely, professional musicians and sound engineers. It was assumed that the listeners chosen for the subjective study had worked often enough in many or all of the halls to be able to remember and appropriately judge their acoustics.

Surveys for all 20 halls were sent to 50 touring musi-

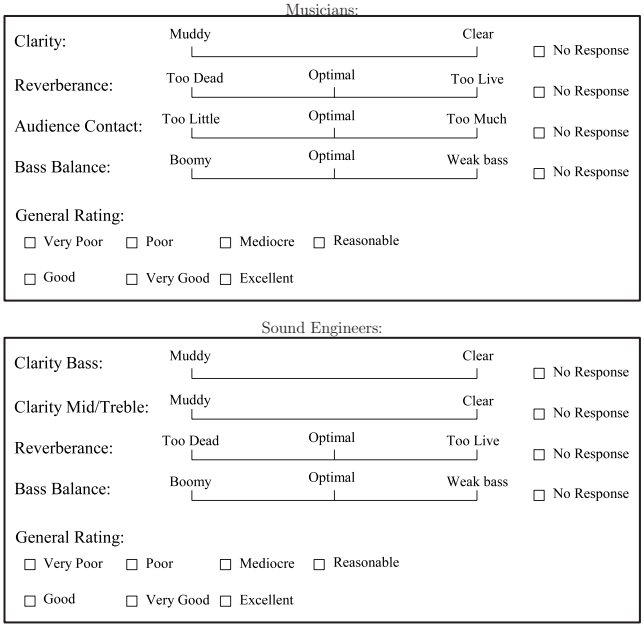


FIG. 1. Musicians' and sound engineers' questionnaire forms for the subjective rating of each of the concert halls (translated from Danish).

cians and 18 sound engineers in Denmark, of which 25 musicians and 8 sound engineers responded. The musicians and sound engineers were selected from the bands that had performed most often in Denmark and at a minimum of 8 of the 20 halls within the 3 year period before the survey was performed. There were eight bands that fulfilled the criteria, as well as several freelance musicians and sound engineers.

The first page introduced the survey by saying (translated from Danish): "As a musician, one evaluates venues—consciously or unconsciously—based on factors, such as: how good is the visual contact with the audience, is the temperature appropriate, is the service good, etc. In this anonymous survey, the focus is on the acoustics of the venue for the performers. This means: how does the hall respond to the music that is played—judged independently (as far as possible) of the PA-system, the monitor technicians, etc." Then the survey asked what kind of monitors the band uses (in-ear, headphone, stage, other), whether the respondent discusses the acoustics of halls with their colleagues (yes/no), how important acoustics are for the respondent (very, somewhat, a little, not important—translated from Danish), whether the respondent had chosen not to play in certain halls because of the acoustics (yes/no), and whether the respondent thought that possible negative effects of acoustics could be mitigated through the use of in-ear monitors (very, somewhat, a little, no). Then, the respondents were asked to complete a questionnaire for each hall, asking for ratings of the halls on several acoustic aspects.

The design of the questionnaire for each hall was based on the questionnaire used by Barron.[5] The musicians' form is shown in Fig. 1, translated from Danish to English. The first two scales, "Clarity" and "Reverberance," are the same as on Barron's questionnaire. For the present study, Barron's ratings of "Envelopment," "Intimacy," and "Loudness" were dropped because these are expected to be more influenced by the PA-system configuration in a rock concert than by the room acoustics. Instead, the scales "Audience Contact" and "Bass Balance" were added. The sound engineers' form differed from the musicians' form on two questions. The Clarity rating was divided into "Clarity Mid-Treble" and "Clarity Bass," and the Audience Contact rating was removed. Each group of respondents had four attributes to rate with a continuous scale and a general rating with discrete values for each hall. The instructions were that the respondents should complete as many forms as possible, but should at least provide a general rating for all of the halls with which they were familiar.

It was expected that the three subjective ratings Clarity, Reverberance, and Bass Balance would be correlated with the objective measurements D_{50} ("Deutlichkeit" or "Definition"), T_{30} or EDT (reverberation time or early decay time), and BR (bass ratio). The Audience Contact rating came from the first author's own experience as a professional rock and jazz musician that the acoustics of the stage and hall can have a strong effect on the feeling of contact with an audience.

The respondents were free to set a mark anywhere on the continuous lines. The lines were 10.8 cm long in the original format with an "optimal" mark at the center point for all but the Clarity ratings. There are advantages and disadvantages to including the optimal mark on the scales. By including the optimal mark, the respondents were given a reference point for where they thought the best halls should lie on the scale. However, an optimal mark can make the scales non-linear, because the optimal point of the scale may not be exactly half-way between the end-points and may be at a different position for each respondent. The positions of the respondents' marks on the line were measured, assuming a linear scale, and the data were assembled for statistical and correlational analysis of the data.

B. Objective measurements

The objective measurements were performed in accordance with ISO 3382:1997.[20] Measurements were made on the stage and in the audience area, using an omni-directional, dodecahedral speaker array with a subwoofer as the sound source, and a second set of measurements was made using the halls' PA-systems. The DIRAC software package (Acoustics Engineering, v3.0) was used on a laptop computer to generate a frequency sweep signal and to calculate the room impulse response. An AKG C34 condenser microphone in omni-directional mode was used to measure the impulse response. At least three measurements were conducted on the stage and another three (minimum) measurements in the audience area, including one point at the sound engineer's position. In each hall with a balcony, measurement positions were added on and below the balcony. The same microphone positions were used with both the dodecahedral array and the PA-system. According to ISO 3382, at least two source positions should be used, which was possible when measuring with the dodecahedral array, with the source placed at two different positions on the stage. However, when the PA-system was used as the source, it could not be moved to a different position, so only one source position was used for

these measurements. The PA-systems' room equalizers were bypassed for the measurements.

The objective measurements were conducted in unoccupied halls and included reverberation time (T_{30}), EDT, and Deutlichkeit/definition (D_{50}). In a previous study,[21] a correlation coefficient of 0.96 was found between D_{50} and Clarity (C_{80}) in similar halls. Also, since the rhythmic information and lyrics are so important for this genre, it was assumed that the intelligibility/clarity parameter with the shortest integration time was the most relevant. Therefore, D_{50} was chosen over C_{80}. The correlation of these metrics with the general ratings of the halls give a starting point for the design of acoustics for rock concert halls.

Sound absorption coefficients can be found for a seated audience[22] and absorption areas for single standing persons.[23] Simply multiplying the absorption area of a single standing person by the number of people in attendance will overestimate the sound absorption, because a large portion of the absorption area will be rendered ineffective by the presence of other tightly packed members of the audience. Therefore, the sound absorption coefficient of a standing rock audience was measured by comparing the reverberation times of a concert hall with and without audience present. The measurement was made using the DIRAC software in the Pumpehuset concert hall in Copenhagen during a break (with audience present) and after the concert (empty hall). The concert hall is a rectangular hall ($32.3 \times 11.6 \times 8$ m^3) with an audience area of 225 m^2 and capacity of 600 people, or an approximate density of 2.7 people/m^2. The floor in the hall was concrete with a rubber coating.

III. RESULTS AND DISCUSSION

A. First page of the questionnaire

The 25 musicians who responded to the questionnaire came from 11 different groups. There were eight drummers, seven bass players, five guitar players, three keyboardists, and two singers. In response to the question "How important are the acoustics of a venue for you?" seven out of eight sound engineers and 17 out of 25 musicians said that acoustics are "very important" to them. The remaining sound engineer and seven musicians said that acoustics are "important" and the remaining musician said that acoustics are only "slightly important." Two of the eight sound engineers surveyed and 8 of the 25 musicians said that they have chosen not to play in certain halls because of the acoustics. All sound engineers and all musicians said that they discuss the acoustics of specific halls with colleagues. Five sound engineers reported that their bands used in-ear monitors, seven reported using on-stage monitors, and one reported using headphones as monitors (note that the respondents could choose more than one monitor type). Fourteen musicians reported using in-ear monitors, 19 use on-stage monitors, and 3 musicians (all drummers) reported using headphones. On the question of whether in-ear monitors can help to mitigate the possible bad effects of a hall's acoustics, four sound engineers and nine musicians responded "very much," three sound engineers and eight musicians responded "somewhat," and one sound engineer and three musicians responded "a little." The remaining five musicians either did not respond or responded "don't know."

These results on the importance of acoustics and whether the respondents discuss acoustics with their colleagues are probably biased, in that those who find acoustics important and discuss acoustics are more likely to respond to a survey on acoustics. The responses to the question on whether the respondent ever chose not to play in certain halls because of the acoustics may also be skewed. Some of the musicians surveyed may not be involved in booking concerts and may not be aware that their band has declined jobs because of the acoustics of the hall. For example, two musicians from the same band were surveyed. One responded "yes" that they choose not to play certain halls, and the other responded "no" that they do not. Presumably, the one who responded yes is involved in booking concerts, while the one who responded no just shows up to play where he or she is told. Overall, these results showed that acoustics are important for rock musicians and sound engineers, and this provided further encouragement to pursue the investigation.

B. General ratings of the halls

The general ratings were analyzed by assigning a number from 1 to 7 to the general ratings options, where a 1 corresponded to "Excellent" and 7 corresponded to "Very Poor." The mean general rating was then calculated for the musicians and for the sound engineers, and the combined rating was calculated as the mean of the two groups. The ordinal rank of the halls' ratings from 1 (best) to 20 (worst) for each group and the combined rating are shown in Table I. The halls are sorted in order from smallest to largest by volume, and it is interesting to note that there is no correlation between the size of the hall and the overall rating.

Interestingly, the driest hall, Stars, is in tenth place in the musicians' ratings but is the favorite of the sound engineers, which moves it to the fourth best rating overall. Stars was also rated driest on the "Reverberance" scale (the only hall rated by the musicians as "too dry"), significantly drier ($p < 0.05$) than all but one other hall. So even though the sound engineers appreciate the recording studio quality of the hall, it is a good example that a hall can be too dry for musicians. The four lowest rated halls have relatively high T_{30} and typically longer reverberation in the low frequencies. Viften has a very long reverberation time at 63 Hz (over 3 s) and much shorter reverberation (around 1 s) for frequencies above 500 Hz. This is also the hall that the sound engineers rated the lowest on Clarity Bass.

C. Correlations between subjective ratings

There were significant differences between halls seen in the analysis [analysis of variance (ANOVA)] of the responses on every rating scale except for Audience Contact. The ratings of Audience Contact were about the same for all halls for each respondent, indicating that the respondents may not have really understood the category. However, even within the small variance, there were significant differences between respondents, so the ratings of Audience Contact

TABLE II. Correlation coefficients for the musicians' subjective ratings (Clarity, Reverb, Audience Contact, and Bass Balance). Significant correlations ($|r|>0.5$) are shown in bold.

	Clarity	Reverb	Aud. Cont.	Bass Bal.
Reverb	**–0.58**			
Aud Cont.	0.02	0.00		
Bass Bal.	**0.67**	–0.49	0.06	
Gen. Rat.	**0.75**	–0.42	0.21	**0.70**

may be based more on non-acoustic factors, such as the role of the respondent in the band or position on the stage (i.e., lead singers may feel more contact with the audience than drummers).

Correlation coefficients were calculated pairwise between the five subjective ratings of the musicians and the sound engineers (see Tables II and III, respectively). The musicians' "General Rating" was strongly correlated with Clarity and Bass Balance, indicating a preference for crisp, not boomy halls. Clarity and Reverberation also have a strong inverse correlation, as opposed to some other studies[4,5] that found no correlation between the two parameters.

There were also strong correlations between the sound engineers' subjective ratings (Table III) of General Rating and the two Clarity' ratings (bass and mid/treble), indicating a preference for crisp sound. The Clarity ratings from the two frequency ranges may reflect quite different acoustic profiles in the different frequency ranges, as seen in the data in Table I. A look at the Clarity data set showed that if there was a difference in the ratings, then "Clarity Mid/Treble" was generally rated as more clear than Clarity Bass.

D. Objective measurements

1. Standing audience absorption

Figure 2 shows absorption coefficients for a standing audience, measured as described in Sec. II B, along with coefficients for a seated audience from Meyer *et al.*[19,22] The data show that the absorption coefficients of a standing audience is five to six times higher in the mid-high-frequency bands than in the low-frequency bands, and that there is very little absorption in the low frequencies. Note that the absorption coefficients that are greater than 1 can be attributed to an audience not being a simple two-dimensional surface. Therefore, the effective absorption area of an audience is larger than the area they cover on the floor. In addition, there will be complex diffraction effects between and around audience members, and a reduction in the effective acoustic volume of the hall. Note that the edge diffraction effects mean that the exact absorption coefficients may depend on the perimeter of the audience area. The imbalance in absorption coefficients across frequencies means that the presence of an audience in a hall will lead to lower mid-high-frequency reverberation times, as compared to the empty hall, but will have little effect in the bass-frequencies. In addition, PA-systems are generally designed to be highly directive in the high frequencies, radiating most of their energy directly at the audience, but can be assumed to be nearly omni-directional in the low-frequency ranges.[24] Therefore, a hall that has approximately equal reverberation times across frequencies when empty will have disproportionately long low-frequency reverberation times when the hall is full. In order to have a balanced hall with a full audience, the reverberation times in the low-frequency bands would have to be lower than in the mid-frequency bands when measured in an empty hall.

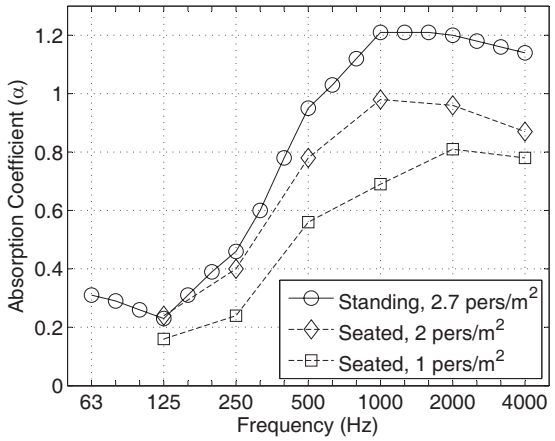


FIG. 2. Absorption coefficients for an audience. Standing audience with a density of about 2.7 people/m^2 (circles) and a seated audience on wooden chairs with a density of 2 people/m^2 (diamonds) (Refs. 19 and 22) and 1 person/m^2 (squares) (Refs. 19 and 22).

TABLE III. Correlation coefficients for the sound engineers' subjective ratings (Clarity Bass, Clarity Mid/Treble, Reverb, and Bass Balance). Significant correlations ($|r|>0.5$) are shown in bold.

	Clarity B	Clarity M/T	Reverb	Bass Bal.
Clarity M/T	**0.66**			
Reverb	**–0.50**	–0.46		
Bass Bal.	**0.51**	0.25	–0.38	
Gen. Rat.	**0.74**	**0.72**	**–0.64**	**0.50**

2. PA vs omni-directional source measurements

Objective measurements were made in the halls using both the omni-directional (dodecahedral) speaker and the halls' PA-systems as the sound source. This allowed a comparison between the results obtained with the different sources. Three-way ANOVAs were conducted on the T_{30}, EDT, and D_{50} data with main factors of source (omni, PA), receiver area (audience, stage), and frequency with a threshold of significance at $p=0.05$. The T_{30} analysis showed no significant main effects of source or receiver area, and no source-receiver interaction. There was a significant main effect of source on the EDT measurements, a significant main effect of location, and a significant source-location interaction with $p<0.01$ for all three analyses. *Post hoc* analysis showed that the omni-stage EDT measurement was significantly lower than the omni-audience and the PA measurements. The PA-stage EDT was significantly lower than the PA-audience EDT, but there was no significant difference between the two audience measurements. The main effects of source, location, and frequency were all significant at the $p<0.001$ level for the D_{50} measurements, as was the source-location interaction. The source-location/frequency interac-

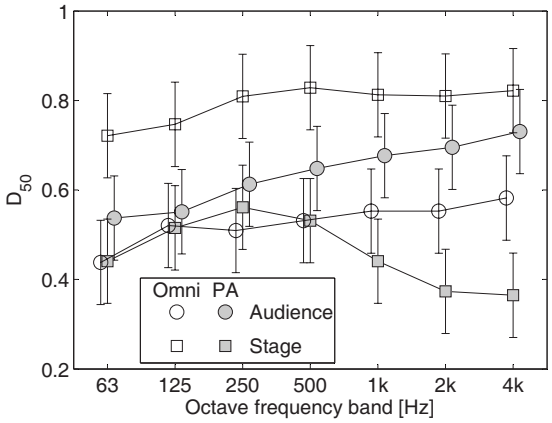


FIG. 3. Mean D_{50} values across halls measured with an omni-directional source (open symbols) and with the hall's PA-system (gray filled symbols), in the audience area (circles) and on-stage (squares). The error bars show the 95% confidence intervals for the differences.

tion (see Fig. 3) was also significant with $p<0.01$.

There were no significant differences between the omni- and PA measurements in the audience. Of course, there will be positions in a hall, for which there will be a difference in the D_{50} measurements, but for a spatial average over many positions in the hall, no significant differences were found. The D_{50} measured on-stage with the omni-source are higher than the other measurements, because the measurement positions were relatively close to the source, so the direct-to-reverberant energy ratio was high. The D_{50} measured on-stage with the PA are much lower than those measured with the omni-source. In addition, the PA-stage measurement shows much lower D_{50} with the higher frequency bands. This is because the PA-systems are generally in front of the stage and are pointed toward the audience. This means that the PA speakers do not radiate much high-frequency energy onto the stage, due to their directivity, so most of the high-frequency energy on-stage is from the reverberant sound field, resulting in a low D_{50}.

These data show that the PA-system can be used for acoustic measurements in the audience area of a hall but should not be used for measurements on the stage. It should be noted that measurements made with the PA-system as a source will not be compliant with ISO 3382.[20]

E. Correlation between subjective and objective parameters

Correlation coefficients were calculated between the mean subjective parameters and the objective parameters $T_{30,W}$(63–2000 Hz), $T_{30,B}$(63–125 Hz), $T_{30,MT}$(250–2000 Hz), EDT(63–2000 Hz), D_{50}(63–2000 Hz), and BR (the ratio of T_{30}(63–250 Hz) to T_{30}(500–2000 Hz)) measured with the omni-source for each hall and are shown in Table IV, with significant correlations ($p<0.05$) marked in bold. The musicians' subjective parameter Audience Contact was not correlated with any of the objective parameters, reinforcing the idea that this parameter may be based more on non-acoustic effects. All of the other subjective parameters for the musicians and the sound engineers were correlated with $T_{30,W}$ and with EDT. The musicians' General Rating was significantly correlated with $T_{30,W}$, but the correlations with $T_{30,B}$ and $T_{30,MT}$ were not significant. The sound engineers' "Clarity B" and "Clarity M/T" ratings were correlated with D_{50} from the audience area, as had been expected, and the musicians' Clarity rating was correlated with D_{50} on the stage. Also, the sound engineers Clarity M/T rating was significantly correlated with $T_{30,MT}$, but not with $T_{30,B}$. The musicians' Bass balance rating was correlated with the measured BR, but the sound engineers' was not.

F. Correlation between T_{30}, hall volume, and the general rating

The reverberation time of a hall generally increases with hall volume, and listeners' expectations of hall quality also require longer reverberation times from larger halls. For ex-

TABLE IV. Correlation coefficients between the musicians' and sound engineers' mean subjective ratings and the measured objective parameters. T_{30} is the mean reverberation time measured in the audience area with the omni-directional source, with subscripts W for wideband from 63–2000 Hz, B for bass from 63–125 Hz, and MT for mid/treble from 250–2000 Hz. EDT is the mean early decay time from 63–2000 Hz measured on the stage with the omni-directional source. The D_{50} values are the mean from 63–2000 Hz measured in the audience area with the PA source (subscript A) and with the omni-source on the stage (subscript S). BR is the bass ratio, i.e., ratio of T_{30}(63–250 Hz) and T_{30}(500–2000 Hz). Significant correlations ($p<0.05$) are marked in bold.

Subjective ratings		Objective measurements						
		$T_{30,W}$	$T_{30,B}$	$T_{30,MT}$	EDT	$D_{50,A}$	$D_{50,S}$	BR
Musicians	General rating	**0.52**	0.50	0.46	**0.51**	–0.33	**–0.50**	0.37
	Clarity	**–0.63**	**–0.55**	**–0.61**	**–0.71**	0.47	**0.66**	–0.31
	Reverberance	**0.69**	**0.60**	**0.66**	**0.77**	–0.35	**–0.56**	0.26
	Bass balance	**–0.65**	**–0.69**	**–0.52**	**–0.71**	0.41	**0.52**	**–0.53**
	Audience contact	–0.20	–0.15	–0.21	–0.13	0.32	0.31	–0.01
Sound Eng.	General rating	**0.69**	**0.58**	**0.68**	**0.77**	**–0.65**	**–0.51**	0.32
	Clarity B	**–0.80**	**–0.73**	**–0.74**	**–0.81**	**0.55**	**0.58**	–0.42
	Clarity M/T	**–0.60**	–0.49	**–0.60**	**–0.75**	**0.70**	0.36	–0.25
	Reverberance	**0.74**	**0.60**	**0.76**	**0.80**	**–0.58**	**–0.61**	0.26
	Bass balance	**–0.70**	**–0.65**	**–0.64**	**–0.57**	0.20	**0.58**	–0.33

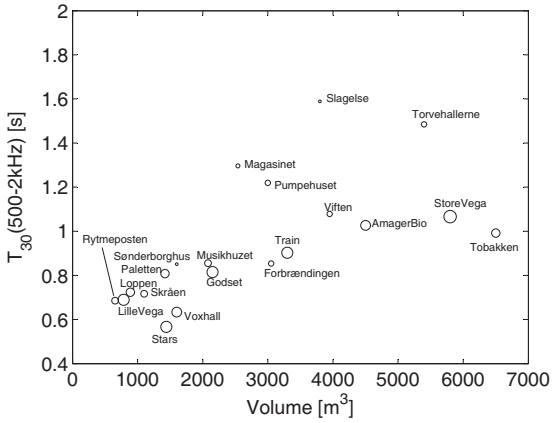


FIG. 4. Mean measured T_{30} on the stage and in the audience area from 500 Hz to 2 kHz vs hall volume for the 20 halls in the study (empty halls). The size of the marker indicates the combined general rating of the hall with better ratings having larger markers.

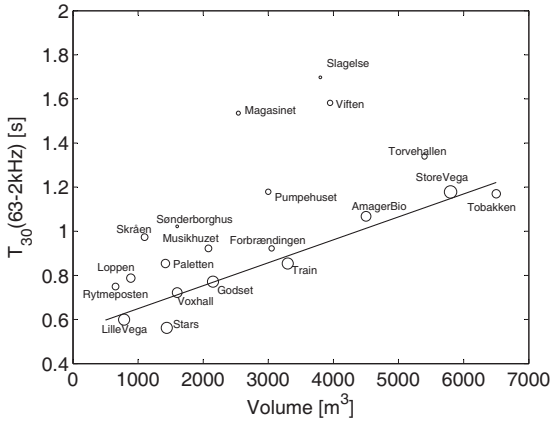


FIG. 5. Mean measured T_{30} on the stage and in the audience area from 63 Hz to 2 kHz vs hall volume for the 20 halls in the study (empty halls). A larger marker indicates a better combined general rating. The line shows the recommended reverberation time as a function of hall volume.

ample, Train and Sønderborghus have very similar reverberation times across frequency; however, Train was rated much higher than Sønderborghus. Adding the dimension of Volume can help clarify the ratings. Figure 4 shows a plot of the mean reverberation time of the stage and audience areas measured with the omni-directional source in the 500 Hz to 2 kHz octave bands for each hall as a function of its volume. The halls with a higher combined rating (see Table I) are marked with a larger circle in Fig. 4 than the lower rated halls. From this, it can be seen that Train has about twice the volume of Sønderborghus, so should be expected to have a longer reverberation time. Most of the halls in the plot fall in an area that increases in T_{30} with volume. Those that are significantly beyond this area are rated the lowest (Slagelse, Magasinet); however, there is an overlap of highly rated and mediocre, or even poorly rated, halls. For example, Viften and Sønderborghus are close to the main cluster, even though they are two of the five lowest rated halls overall. This means that T_{30} in the 500–2000 Hz bands is a poor predictor of the general hall rating.

Greater separation between the highly rated and mediocre halls can be achieved by including the bass-frequencies in the model. Figure 5 shows the same data as Fig. 4, except that all frequency bands from 63 Hz to 2 kHz are included. Viften now lies well away from the main cluster because it has a disproportionately long reverberation time in the bass-frequencies. This indicates the importance of including the bass-frequencies in the mean hall reverberation time as a means of predicting the general impression of the acoustics. A weighted regression line was fitted to the data shown in Fig. 5 for the ten top-rated halls, with double weight for the five top-rated halls. The equation for this line is

$$T_{30} = 0.55\ \text{s} + 1.04 \times 10^{-4}\ \text{V s/m}^3, \quad (1)$$

where V is the volume of the hall. (Note that recommended reverberation times in the literature are usually proportional to the logarithm of hall volume. Further studies extending this investigation to larger hall volumes may reveal a similar trend for rock music halls.) This model explained 80% of the variance (R^2) of the best halls' data.

By separating the halls into groups of the best- and worst-rated halls, some trends can be seen that can be used to help design the acoustics of a rock concert hall. Figure 6 shows the estimates of the mean and standard errors of the reverberation time as a function of hall volume by frequency band. The model shown in Eq. (1) was used to normalize the T_{30} data for an analysis of the differences between the two groups. A two-way ANOVA on the normalized T_{30} with main effects of group (best/worst) and frequency band showed a significant effect of group [$F(1,112)=33.7$, $p \ll 0.001$] and of frequency [$F(6,112)=4.5$, $p<0.001$], as well as a significant interaction [$F(6,112)=2.32$, $p<0.05$]. Similar differences were also found using only the musicians' ratings and the stage acoustics, as well as for the engineers' ratings with the audience acoustics. Therefore, the results are presented for the combined ratings with the overall hall mean reverberation times. The estimated mean normalized reverberation times for the two groups and 95% confidence intervals are shown as a function of frequency band in Fig. 6. This shows that the normalized reverberation times are significantly lower in the low-frequency bands for the best-rated halls, and that there is little difference in the high-frequency bands. The

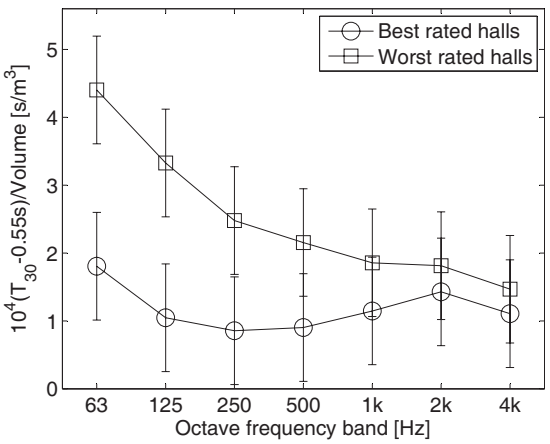


FIG. 6. Linear transform of T_{30} by volume as a function of frequency band, grouped by best- and worst-rated halls. Shown are the estimates of the means of ($T_{30}-0.55$ s)/volume and 95% confidence intervals for the two groups.

worst-rated halls in this study tend to have sloping reverberation-time profiles with higher T_{30} in the bass than in the higher frequency bands, while the best halls have, on average, approximately equal reverberation times across frequency. These data suggest that the best halls should have flat reverberation time profiles across frequencies and that the reverberation times should be close to the line drawn in Fig. 5, i.e., 0.6–1.2 s for halls from 1000–7000 m^3, respectively. Of course, there are many other acoustic and nonacoustic factors that can influence the general acoustic impression of a concert hall, but those are beyond the scope of the present study.

IV. CONCLUSIONS

In the subjective survey, both musicians and sound engineers find the acoustics of the concert hall very important for their performance. More than one in three musicians responding reported choosing not to play in a hall on account of bad acoustics. Therefore, it is important to properly design the acoustics for a rock or multipurpose venues.

A standing audience at a rock concert will absorb five to six times as much energy in the mid-high-frequency bands than in the low-frequency bands. If an empty hall has a disproportionately long reverberation time in the bass, then the addition of an audience will only make the bass ratio even more unbalanced. A modern rock or pop performance has very high sound levels in the bass-frequencies, so it is important to consider the bass acoustics in the hall.

The general acoustic impression of a rock concert hall is strongly correlated with the perceived clarity of the hall, also in the bass-frequencies. Therefore, it is critical to consider the bass acoustics, including the 63 Hz octave band, in the acoustic design of the hall. This means that manufacturers of acoustic treatments should also measure and report the acoustic properties down to 63 Hz.

Objective acoustic measurements were made with both an omni-directional source and with the halls' PA-system. There were no significant differences in the measured acoustic parameters between the sources when measured in the audience area, but there were large differences seen on the stage. Therefore, the PA-system may be usable as a sound source for measurements in the audience area but should definitely not be used to characterize the stage area.

The best-rated halls in the study have a flat reverberation time profile across frequencies and have a reverberation time as a function of volume close to Eq. (1), while the worst-rated halls had significantly higher reverberation times in the bass-frequencies. Therefore, it is recommended that rock venues should be designed with reverberation times close to Eq. (1). This recommendation is valid for small to medium-sized halls and needs to be verified for scalability to larger halls. In order to help predict the quality of a hall, when publishing data or writing a report on a hall used for rock music, the reverberation times for the hall should be reported in frequency bands from 63 Hz and up. Of course, T_{30} is only one aspect of acoustic design. Further studies should investigate the placement of reflective and absorptive surfaces in the hall and stage area to optimize other acoustic parameters to meet the desires of the musicians, sound engineers, and audience.

Note that several of the halls listed here have been renovated since the measurements for this study were performed in 2005, so the acoustic data listed here and the subjective impressions may not be up-to-date with the current state of the halls.

[1] M. Barron, *Auditorium Acoustics and Architectural Design* (E & FN Spon, London, 1993).
[2] L. Beranek, *Concert Halls and Opera Houses: Music, Acoustics, and Architecture*, 2nd ed. (Springer, New York, 2003).
[3] *Halls for Music Performance: Another Two Decades of Experience 1982-2002*, edited by I. Hoffman, C. Storch, and T. J. Foulkes (Acoustical Society of America, Melville, NY, 2003).
[4] R. J. Hawkes and H. Douglas, "Subjective acoustic experience in concert auditoria," Acustica **24**, 235–250 (1971).
[5] M. Barron, "Subjective study of British symphony concert halls," Acustica **66**, 1–14 (1988).
[6] T. J. Cox and B. M. Shield, "Audience questionnaire survey of the acoustics of the Royal Festival Hall, London, England," Acustica **85**, 547–559 (1999).
[7] T. Hidaka and L. L. Beranek, "Objective and subjective evaluations of twenty-three opera houses in Europe, Japan, and the Americas," J. Acoust. Soc. Am. **107**, 368–383 (2000).
[8] T. Hidaka and N. Nishihara, "Objective evaluation of chamber-music halls in Europe and Japan," J. Acoust. Soc. Am. **116**, 357–372 (2004).
[9] Y. J. Choi and F. R. Fricke, "Evaluation of the relative acoustic performance of two auditoria using measurements and auralization," Acta. Acust. Acust. **91**, 1051–1062 (2005).
[10] Foreningen Spillesteder.dk, "Spillesteds-undersøgelse, 2005 [Survey of music venues, 2005]," available online at: http://www.spillesteder.dk/temperatur_2005.htm (Last viewed August, 8, 2008; in Danish).
[11] P. White, "Where to use processors and why: Part 3," Sound on Sound **13**, 242–246 (1997).
[12] The critical distance is the distance from a sound source at which the reverberant sound energy is equal to the energy of the sound radiating directly from the sound source. At greater distances from the source, the reverberant field dominates the sound received. The critical distance decreases with increasing reverberation time.
[13] A. C. Gade, "Investigations of musicians room acoustic conditions in concert halls. I. Methods and laboratory experiments," Acustica **69**, 193–203 (1989).
[14] A. C. Gade, "Investigations of musicians room acoustic conditions in concert halls. II. Field experiments and synthesis of results," Acustica **69**, 249–262 (1989).
[15] R. C. Cabot, C. R. Genter, and T. Lucke, "Sound levels and spectra of rock-music," J. Audio Eng. Soc. **27**, 267–284 (1979).
[16] K. Dibble, "Hearing loss & music," J. Audio Eng. Soc. **43**, 251–266 (1995).
[17] M. Barron, "The subjective effects of first reflections in concert halls—The need for lateral reflections," J. Sound Vib. **15**, 475–494 (1971).
[18] A. C. Gade, "Subjective room acoustic experiments with musicians, Report 32, Acoustics Laboratory," Ph.D. thesis, Technical University of Denmark, Lyngby, Denmark (1982).
[19] H. Kuttruff, *Room Acoustics*, 4th ed. (Spon, London, 2000).
[20] ISO 3382, *Acoustics—Measurements of the Reverberation Time of Rooms with Reference to Other Acoustical Parameters*, 2nd ed. (International Organization for Standardization, Geneva, Switzerland, 1997).
[21] N. W. Larsen, E. Olmos, and A. C. Gade, "Acoustics in halls for rock music," in Joint Baltic-Nordic Acoustics Meeting (BNAM), Mariehamn, Finland, (2004).
[22] E. Meyer, D. Kunstmann, and H. Kuttruff, "Über einige Messungen zur Schallabsorption von Publikum (On some measurements of sound absorption of the audience)," Acustica **14**, 119–124 (1964).
[23] U. Kath and W. Kuhl, "Messungen zur Schallabsorption von Personen auf ungepolsterten Stühlen (Measurements of sound absorption of people on unupholstered seats)," Acustica **14**, 50–55 (1964).
[24] G. Adams, "The room environment," in *Loudspeaker and Headphone Handbook*, edited by J. Borwick, (Butterworths, London, 1988), pp. 264–318.

Evaluation of a method for the measurement of subwoofers in usual rooms

Manuel Melon[a] and Christophe Langrenne
Conservatoire National des Arts et Métiers, 292 rue Saint Martin, 75141 Paris cedex 3, France

Philippe Herzog
LMA-CNRS, 31 chemin Joseph Aiguier, 13402 Marseille cedex, France

Alexandre Garcia
Conservatoire National des Arts et Métiers, 292 rue Saint Martin, 75141 Paris cedex 3, France

(Received 6 May 2009; revised 10 November 2009; accepted 10 November 2009)

This paper evaluates the potential of the field separation method (FSM) for performing subwoofer measurements in a small test room with poor absorbing properties, as is commonly available. The FSM requires the knowledge of both acoustic pressure and velocity fields on a closed surface surrounding the tested source. Pressures and velocities, measured using a p-p probe on a half-sphere mesh, are collected under various conditions: in a room with variable reverberation time (6.4–0.6 s) and with four measurement half-sphere radii. The measured data are expanded on spherical harmonics, separating outward and inward propagation. The pressure field reflected by walls of the surrounding room is then subtracted from the measured field to estimate the pressure field that would have been radiated under free-field conditions. Theoretical frequency response of the subwoofer is computed using an analytical formulation derived from an extended Thiele and Small model of the membrane motion, coupled to a boundary element model for computing the radiated pressure while taking into account the actual subwoofer geometry. Measurement and simulation results show a good agreement. The effects of the measurement distance, the measurement point number, and the room reverberation time on the separation process are then discussed.
 [DOI: 10.1121/1.3270392]

PACS number(s): 43.58.Vb, 43.20.Ye, 43.60.Sx, 43.38.Ja [EGW] Pages: 256–263

I. INTRODUCTION

Sound sources dedicated to the lower audio frequency band are more and more widely used, e.g., as subwoofer for HiFi or home theater systems, or for active noise control. Both technology of the driver and design of enclosures have known significant improvements, but the measurement of such sources is still a tricky problem.

The standard measurement methods[1] involve free-field pressure response, which is almost impossible to achieve at frequencies from 20 to 100–200 Hz because such free-field conditions cannot be easily reached. The simplest method is to perform measurements outdoor, but really quiet places with buildings far enough are not so easy to find nowadays. Moreover, bad weather conditions, i.e., excessive wind or rain, can postpone the measurement process for several days.

Conversely, indoor measurements are not realistic, as no test room is truly anechoic at such low frequencies. As an example, a good 400 m^3 anechoic room generally has a cut-off frequency around 70 Hz. Due to the modal behavior of such a room below its cut-off frequency (or to the reflections from nearest walls), the acoustic field radiated by a loudspeaker system is strongly modified. Moreover, large anechoic chambers are expensive and many institutions or companies cannot afford to own or rent them for routine measurements.

To overcome this problem, many scientists and engineers have explored alternative estimation methods allowing to use more common equipments. For instance, the impulse response of a sound source may be evaluated in very large rooms, using a time window to remove sound reflected from the walls. This windowing process can be done either in the time domain[2,3] or in the cepstral domain.[4] For simple loudspeakers, the far field response may also be estimated from near field[5,6] or internal pressure measurements,[7] provided that all vibrating surfaces are identified and well characterized. The loudspeaker velocity may also be monitored with an accelerometer; then, using rigid piston assumption, low-frequency response can be calculated.[8] More generally speaking, parametric identification of the electroacoustic model of the tested loudspeaker system may be used to compute its virtual response, and several of these methods may be combined.[9]

Nevertheless, the previously cited methods rely on assumptions which are not always found realistic in the field: low-frequency speakers often involve reflex or bandpass systems whose impulse response can last more than 100 ms, while reflections from walls can occur 10 ms after the arrival of the direct sound. Even with a pre-equalization to shorten the measured impulse,[10] windowing cannot separate direct and reflected sound, leading to severe estimation errors

[a]Author to whom correspondence should be addressed. Electronic mail: manuel.melon@cnam.fr

around the low cut-off frequency. On the other hand, model identification assumes that all the radiating surfaces are known and physically well described. However, cheap loudspeaker boxes exhibit wall vibrations and sometimes leakages. Moreover, a loudspeaker cone is far from a circular piston or any of the academic radiators usually considered.

So, even if real free-field measurements are barely feasible, a general method should avoid simplistic assumptions about the device under test. On the other hand, source size can always be considered as small compared to low-frequency wavelengths, and measurements at several locations around the source can be combined to get information about wavefronts instead of local pressure. Thus, nearfield acoustic holography can be a useful tool for characterizing a complicated pressure field radiated by a sound source. For instance, accurate results have been obtained by Lu *et al.*[11] when applying Helmholtz equation least-squares method[12] to the measurement of the directivity patterns of a baffled plate. However, these results have been obtained at high frequencies where anechoic conditions are easily achieved. When dealing with nonanechoic rooms, a solution, first proposed by Weinreich *et al.*,[13] consists in using spherical harmonics expansions of a double layer pressure field, separating outward and inward terms to distinguish the field radiated by a tested source from the field reflected by walls. Later, Tsukernikov[14] used an integral formulation to solve this problem, while Langrenne[15] provided a hybrid method connecting integral formulation to spherical harmonic expansion. Williams, in his book,[16] reformulated Weinreich approach for intensity probe measurements. Note that all these formulations require either double layer pressure or pressure-velocity measured data on a mesh enclosing the tested source. These techniques will be afterward referred as field separation method (FSM).

The FSM approach has already been used to measure closed-box and active/passive subwoofer systems in a semi-anechoic room:[17] the on-axis frequency response has been compared to the one obtained outdoor in nearly free-field conditions, showing a fairly good agreement. These encouraging results do not mean that FSM is a "magic trick," and the present paper attempts to establish some of the limitations of this method. It has therefore been tested under rather difficult conditions (standard room to reverberant room) to check its robustness. Moreover, as this method requires the knowledge of both pressure and velocity field on a closed surface surrounding the source, we try to give indications about the lowest number of measurement points which may give satisfactory results.

In this paper, the field separation method is used to recover the free-field response of a closed-box system from measurements performed in a reverberant room. In the first section, the theoretical basis of the field separation method is recalled. Then, the parameters of an accurate model of the loudspeaker are identified from electrical measurements in order to compute the theoretical on-axis acoustic pressure field radiated by the subwoofer. In Sec. II, pressure measurements are performed in a reverberant room with increasing additional damping materials on its walls. The separation algorithm is then used to estimate the free-field response and compare it to the theoretical response. Section V discusses these results to determine an efficient configuration for practical use of the method.

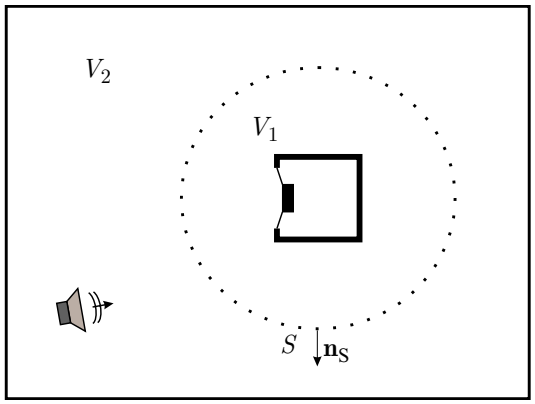


FIG. 1. (Color online) Geometry of interest.

II. THEORY

The aim of the method is to recover the free-field response $p^o(\mathbf{r})$ radiated by a sound source from pressure and velocity field measurements performed in bounded space. As the theoretical basis of FSM have already been described in Refs. 15 and 16, only a brief summary will be given here.

The geometry of the problem is illustrated by Fig. 1: the measurement surface S is a closed one, surrounding volume V_1 which includes the tested source. Volume V_2 can contain secondary sources and is bounded by reflective room walls.

The pressure field $p(\mathbf{r})$ is then given by $p(\mathbf{r})=p^o(\mathbf{r})+p^i(\mathbf{r})+p^s(\mathbf{r})$, where $p^i(\mathbf{r})$ is the incident field reflected by walls or radiated by other sources in V_2 and $p^s(\mathbf{r})$ is the scattered/radiated field created by the incident field on the body of the studied source. At very low frequencies or for small sources, scattering on the body of the source can be neglected: $p^s(\mathbf{r})\simeq 0$. With this assumption, $p(\mathbf{r})=p^o(\mathbf{r})+p^i(\mathbf{r})$. Note that, if needed, scattering effects at higher frequencies could be subtracted by using an integral formulation.[18]

If measurements are performed on a spherical surface of radius s, $p(\mathbf{s})$ and $v(\mathbf{s})$ can be conveniently expanded on normalized spherical harmonics:

$$p(\mathbf{s}) = \sum_{n=0}^{N} \sum_{m=-n}^{n} \alpha_{nm} Y_{nm}(\theta_s, \varphi_s),$$

$$v(\mathbf{s}) = \sum_{n=0}^{N} \sum_{m=-n}^{n} \beta_{nm} Y_{nm}(\theta_s, \varphi_s), \tag{1}$$

where θ_s is the colatitude, φ_s is the longitude, and $Y_{nm}(\theta_s, \varphi_s)$ are the normalized spherical harmonic functions.[16] Note that the $e^{-i\omega t}$ time dependence has been suppressed. The α_{nm} and β_{nm} coefficients are complex constants related to the measurement and are the unknown quantities of the problem.

The pressure and velocity fields can be theoretically expressed in terms of incoming and outgoing waves:

$$p(\mathbf{r}) = \sum_{n=0}^{N} \sum_{m=-n}^{n} (a_{nm} h_n^{(1)}(kr) + b_{nm} j_n(kr)) Y_{nm}(\theta_r, \varphi_r),$$

$$v(\mathbf{r}) = \frac{1}{i\rho_0 c} \sum_{n=0}^{N} \sum_{m=-n}^{n} (a_{nm} h_n'^{(1)}(kr) + b_{nm} j_n'(kr)) Y_{nm}(\theta_r, \varphi_r), \tag{2}$$

where $\rho_0 c$ is the characteristic impedance of air and k is the wave number. The first kind spherical Hankel function $h_n^{(1)}$ represents the outgoing field, while the spherical Bessel function j_n represents the incoming field.

By using the Wronskian relation for spherical harmonic functions, the a_{nm} coefficient are obtained from Eqs. (1) and (2):

$$a_{nm} = ik^2 s^2 [\alpha_{nm} j_n'(ks) - i\rho_0 c \beta_{nm} j_n(ks)]. \tag{3}$$

Finally, the outgoing field is given by

$$p^o(\mathbf{r}) = \sum_{n=0}^{N} \sum_{m=-n}^{n} a_{nm} h_n^{(1)}(kr) Y_{nm}(\theta_r, \varphi_r). \tag{4}$$

In practice, spherical harmonic expansions are finite, consisting of terms up to order N and then giving a maximum frequency for propagative waves. The usual criterion is given by $ka \simeq N$,[19] where a is the radius of the smallest sphere surrounding the source. However, when performing measurements in reverberant space, the incoming field can be seen as a superposition of plane waves which can be expanded on spherical harmonics. This expansion will describe the physical field with a good accuracy if the criterion $ks \simeq N$ is fulfilled. As the field p° is calculated from expansions of the total fields $p(s)$ and $v(s)$, which also contain the incident field, this last criterion will prevail.

Expansions of the measured data are performed using a Gaussian sampling:[20] the azimuth angle is sampled at $2(N+1)$ equiangle samples while the elevation angle requires only $(N+1)$ samples. The method chosen in this paper to compute the α_{nm} and β_{nm} coefficients is a matrix inversion using singular value decomposition. With a Gaussian sampling, condition number is generally near unity.

The outgoing field can be calculated if both acoustic pressure and velocity are known on S. In practice, this can be achieved by using a standard p-p probe or a p-u probe.[21] In this paper, since we used p-p probes with 10 cm spacing, we could have directly applied Weinreich formulation based on pressure measurements on two concentric surfaces.[13] Another solution consists in calculating pressure and velocities on the middle surface S [mean values for $p(s)$ and approximation of Euler's equation by finite difference for $v(s)$] and then use the formulation described here. Simulations performed on a monopole in free space showed that this last method gives the smallest reconstruction error rates for p^o. These results were obtained in the 10–300 Hz frequency band with a 10 cm microphone spacing. We therefore chose this implementation. Nevertheless, when dealing with p-p probe, estimation of the velocity by Euler's equation can be affected in the higher frequency range. A future comparison with results obtained from a p-u probe would be interesting to perform.

For practical purposes, measurements are here performed on a half-sphere, which the ground of the test room is base plane. Assuming a perfectly rigid behavior of this ground (painted concrete), measured data can be considered as symmetric with respect to the ground, and the expansions can be performed only on spherical harmonics which exhibit the same symmetric properties, i.e., even values of $(n+m)$. An alternative procedure would be to duplicate the measured data from the upper half-sphere to the lower half-sphere symmetrically to the ground, and use the full expansion.

III. SUBWOOFER RESPONSE SIMULATION

For the purpose of validating the measurements, we need reference data. We chose to consider a standard loudspeaker system and build its model by combining two parts: the mechanical behavior of the loudspeaker is described by an electroacoustic model, and the acoustic radiation is computed from a boundary element method (BEM) numerical integral. Considering a small subwoofer, both models are used within their best validity ranges, and this combination allows to take in account simultaneously the actual losses and the actual geometry of the experimental situation.

The subwoofer used in this paper is a closed-box system with a Peerless 269 SWR 51 XLS loudspeaker. The box is a cube with edge length of 0.395 m and walls consisting of 3 cm medium density Fiberboard wood. The driver has a free resonance frequency f_s of about 20 Hz. The closed box has an interior volume of 37 l, so its compliance raises the resonance frequency to about 37 Hz. As the driver electrical damping is very high, a series resistance of 4.7 Ω has been added to the loudspeaker giving a total quality factor Q_{tc} close to 0.82. This system is typical of a good quality commercial subwoofer. However, the enclosure is not filled with any damping material in order to facilitate the modeling of the system.

Determination of the actual speaker parameters is performed by measuring its electrical impedance under several operating conditions, as described below. For each measurement of the electrical impedance Z_e, the parameters of three functions are identified by a least-squares optimization method. The first part of Z_e is the electrical resistance R_e, which is assumed constant here, as only low-level measurements are considered. The second part is the electrical inductance, which is frequency-dependent, and is described by a slightly hysteretic model as proposed by Leach.[22] The third part of Z_e is the motional impedance, corresponding to the electromotive force generated by the coil motion. This part reflects the mass-spring resonance of the moving part of the driver and includes the creep effect in the suspensions. This latter effect is described by the exponential model proposed by Knudsen,[23] as it requires a single parameter in addition to the usual Thiele and Small model.

As all three parts of Z_e have a significant influence at all frequencies, they are determined in a global iterative identification process. However, the influence of their parameters

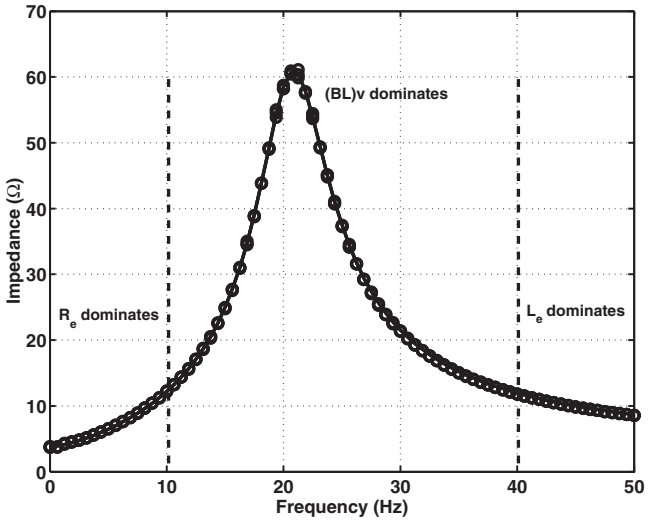


FIG. 2. Comparison of three measurements of the same loudspeaker-box combination (circles) and the corresponding estimation of Z_e (solid line). The black vertical dashed lines show examples of the two limits between the three frequency zones where specific parts of the impedance are estimated. Actual transition frequencies are automatically adjusted during the iterative process.

are dominant in specific frequency zones: very low frequencies for R_e, around resonance for the motional part, and at medium frequencies for the inductance. The identification process therefore involves a single loop separated into three consecutive steps, focusing the estimation of each part of Z_e on the corresponding frequencies. This allows to use simple analytic identification formulas for each aspect, and the resulting algorithm has been found more robust than with a single global minimization. Identification details are beyond the scope of this paper, which only requires its results; nevertheless, the principle of the method can be found in Ref. 24. Figure 2 shows a comparison of three measurements of the loudspeaker-box combination (circles) and the corresponding estimation of Z_e (solid line). They agree very well with the measurements over the whole frequency range. As the frequency range of interest is centered around the loudspeaker resonance, the electrical impedance is almost dominated by its motional part, proportional to the diaphragm velocity. The agreement between measured and simulated electric impedances results therefore from the agreement between velocities, showing that the estimated parameter set provides a good description of the loudspeaker volume velocity.

TABLE I. Loudspeaker's parameters.

Parameter	Value
Resonance frequency	20.8 Hz
Mechanical quality factor	3.54
Electrical quality factor	0.238
Moving mass	0.113 kg
Suspension compliance	4.41×10^{-4} m N^{-1}
Voice coil resistance	3.81 Ω
Diaphragm piston area	0.0353 m^2
Force factor	16.7 T m
Lossy inductor parameters[a]	$L_e=0$, $K=0.0266$, $n=0.683$
Creep parameter (EXP model)[b]	$\beta=0.0522$

[a]Reference 22.
[b]Reference 23.

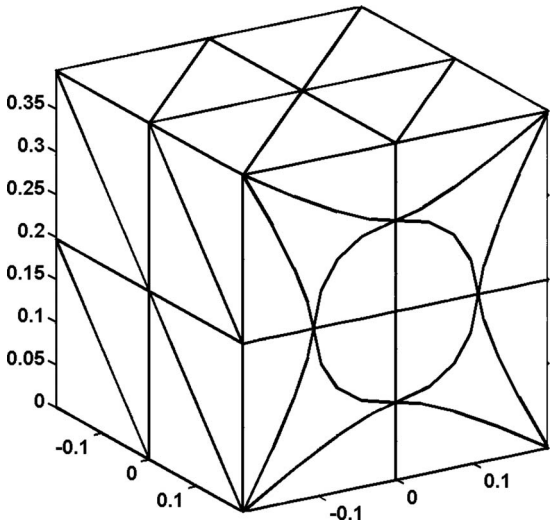


FIG. 3. Simplified mesh of the subwoofer.

The electrical impedance is identified as explained above for three load cases: free standing driver, driver with an added mass M_a ($M_a=0.1803$ kg), or driver mounted in the subwoofer's box. Each load (mass or box) induces a shift in resonance and damping compared with the stand-alone speaker, and this shift is related to the relative mass or stiffness of the speaker compared to the current load. All measurements are therefore combined in a global least-squares identification for estimating the actual speaker parameters from the resonance shifts caused by different loads. Table I summarizes the identified values of the loudspeaker's parameters, which lead to an accurate description of the electromechanical behavior of the considered subwoofer.

The on-axis pressure is then computed by using an analytical method (the baffled planar piston model) or a BEM. Note that, for the numerical method, the effect of the assumed perfectly reflective ground is taken into account by using a half-space Green function. Two meshes have been

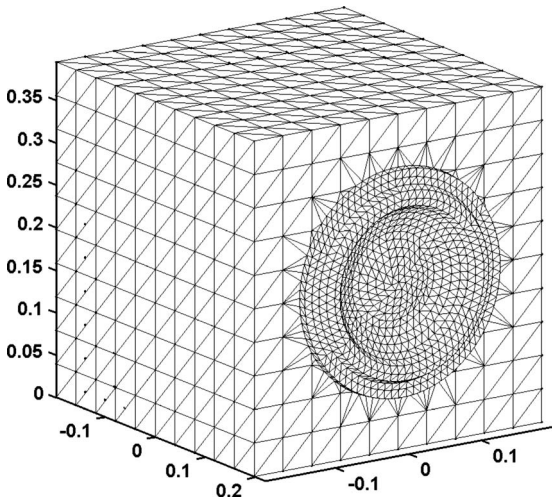


FIG. 4. Detailed mesh of the subwoofer.

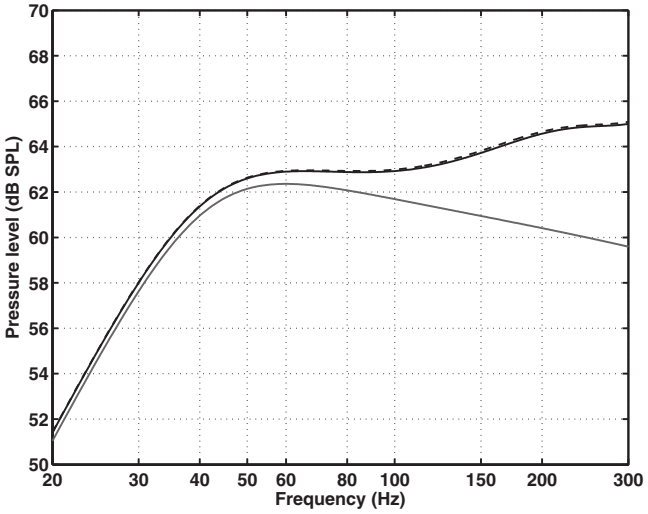


FIG. 5. On-axis pressure level at 4 m: analytical model (gray line), BEM using simple mesh (black line), and BEM using detailed mesh (dashed line).

tested: a very simplified one made of 130 points decomposed onto 64 curvilinear triangular elements with six nodes (Fig. 3) and a much more detailed one made of 2814 points decomposed onto 1409 triangular elements with three nodes (Fig. 4). For the numerical simulations, the membrane velocity has been adjusted to give the same volume velocity as the one resulting from the electroacoustic model. The velocity is considered as uniform over the membrane for the simplified mesh. For the more elaborated mesh, the velocity is uniform on the membrane and inner half of the suspension, and has a linear decrease in the outer half of the suspension.

Results are shown in Fig. 5. As expected, the analytical and numerical models nearly give the same results at very low frequencies (below 50 Hz). At somewhat higher frequencies, the difference in terms of radiation conditions, i.e., a baffled source and a source located close to the edge formed by the ground and the box, is no longer negligible giving a 5 dB deviation at 300 Hz. This first result shows that, even for such low frequencies, the effect of the actual geometry of the system on the acoustic radiation has to be taken into account.

Conversely, the resolution of the mesh has much lower importance: the two numerical simulations are very close, with differences lower than 0.1 dB at 20 Hz and 0.2 dB at 300 Hz. This greatly facilitates the discretization of subwoofers and speeds up calculation (3.5 s per frequency for the simplified mesh compared to 50 min per frequency for the detailed mesh on the same computer).

TABLE II. Measured reverberation times (s) in octave 125 and 250 Hz for the four configurations shown in Fig. 6. Results are averaged from six measurement positions (three microphone positions and two source positions).

Configuration	125 Hz	250 Hz
a	6.4	6.5
b	2.7	2.5
c	1.4	1.7
d	0.5	0.7

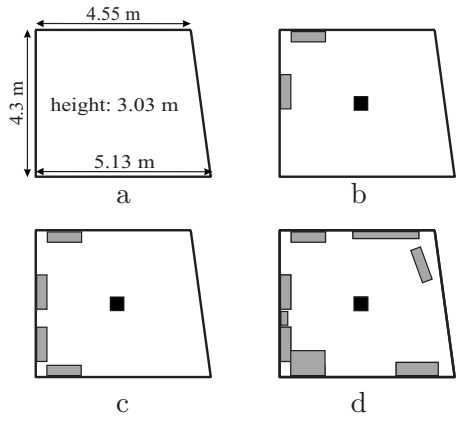


FIG. 6. Schematic drawing of the reverberant room with or without damping materials. The black square shows the location of the tested subwoofer. The gray rectangles show damping materials locations in the four studied configurations (a–d). Empty room layout gives room's dimensions.

IV. MEASUREMENTS

A. Experimental set-up

The subwoofer is driven by a band limited white noise (10–300 Hz) test signal. The amplifier used is a RANE MA 6S model. The subwoofer is put on the rigid ground of a reverberant room which has a volume of about 63 m^3. The reverberation time of this test room can be modified by adding or removing blocks of damping materials. Four configurations have been chosen for the measurements. Table II gives reverberation times in octave bands centered at 125 and 250 Hz for the four configurations sketched in Fig. 6. Note that the values for case d correspond to the ones generally observed in usual rooms.

Automated five layer pressure measurements are performed by a semi-spherical robot (Fig. 7). The radii of the five measurement half-spheres range from 0.6 to 1 m by increasing steps of 0.1 m. The five microphones are calibrated in amplitude and phase up to 1200 Hz. Acoustic data on the medium surface S can then be computed from two adjacent measurement half-spheres using their mean value for the pressure and approximation of Euler's equation by finite difference for velocity. This configuration leads to four measurement surface S of radii 0.65, 0.75, 0.85, and 0.95 m.

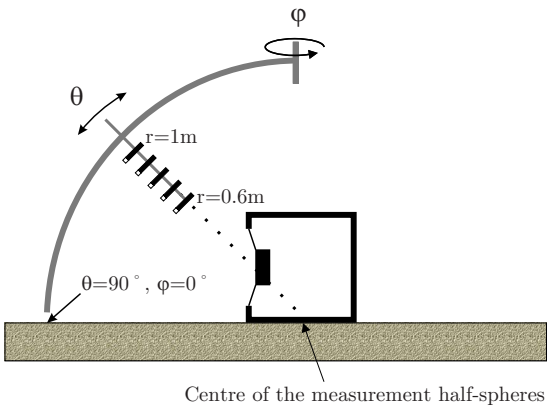


FIG. 7. (Color online) Geometry of the measurement set-up.

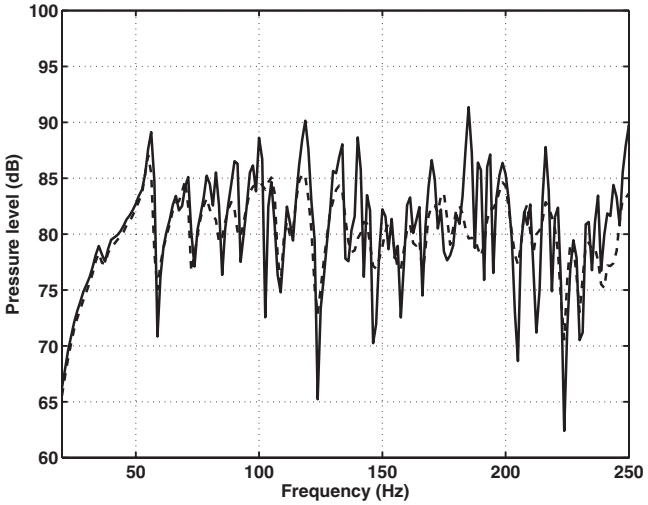


FIG. 8. Measured pressure level in dB SPL at position (0.6 m, 81°, 0°) for configurations b (solid line) and d (dashed line).

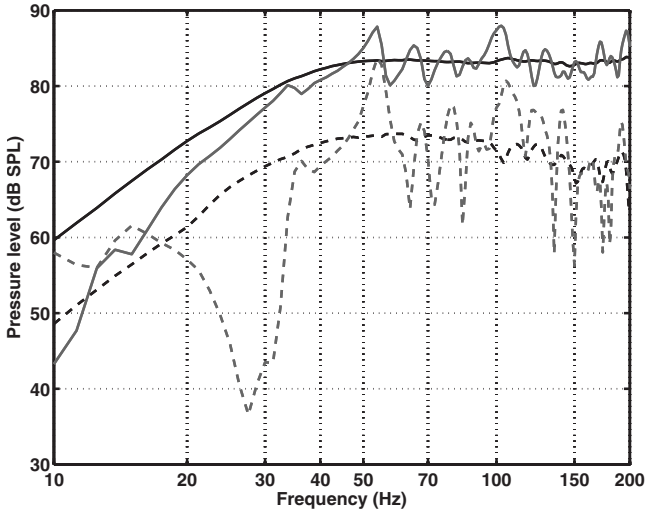


FIG. 9. Mean measured pressure level in dB SPL for configuration d at positions (0.65 m, 81°, 0°): gray line and (0.65 m, 81°, 180°): gray dashed line. Outgoing pressure level in dB SPL at positions (0.65 m, 81°, 0°): black line and (0.65 m, 81°, 180°): black dashed line.

These four different measurement meshes have been tested for each reverberation time, fulfilling the condition that the discretization angle $\Delta\psi$ is the same for azimuth and zenith. Then, the four meshes have, respectively, the following number of points: 16 points ($\Delta\psi=45°$), 36 points ($\Delta\psi=30°$), 64 points ($\Delta\psi=22.5°$), and 100 points ($\Delta\psi=15°$).

B. Results

All frequency curves shown in this paper are transfer functions thus giving pressure level for 1 rms V electrical input. The measured pressure level in front of the subwoofer at position (0.6 m, 81°, 0°) is plotted on Fig. 8 for configurations b and d. The frequency curves clearly show the modal behavior of the testing room. For instance, maximum pressure variations are in the range of 25 dB for case b and 15 dB for case d. The first mode along the vertical direction is in evidence around 56 Hz. Note that the first mode ($\simeq$40 Hz) along the 4.3 m distant walls is not visible in Fig. 8 because the subwoofer radiates near the nodal line. Upper modes are also distinguishable and can be easily related to the room's dimensions.

Once the measurements are completed, the separation process can be applied to the data. First, velocities and pressures on the medium half-sphere S between two measurement meshes are calculated. Then, the spherical harmonic expansion is performed and the outgoing pressure field is calculated. Except where otherwise stated, all expansions have been truncated to $N=5$.

An interesting result highlighting the effect of the separation process is shown in Fig. 9 for configuration d. The modal influence of the testing room seems to be removed from the frequency response curve. The recovered frequency curve (thick line) in front of the loudspeaker is very smooth and has the typical shape of a closed-box frequency response. The recovered outgoing pressure at the rear of the closed-box system (dashed line), however, exhibits oscillations of 1–2 dB. This may result from the increased distance between this point and the membrane of the loudspeaker compared to the same distance for the front measurement point. The higher direct-to-reverberated signal ratio in front of the subwoofer presumably allows a better field separation.

V. DISCUSSION

In order to evaluate the validity of the method, the on-axis pressure p^o, computed from Eq. (4), is compared to the theoretical one presented in Fig. 10 for configuration d. The recovered outgoing on-axis pressure at 4 m from the loudspeaker's membrane is calculated from the spherical harmonic expansion. This distance allows to reduce the uncertainty on the location of the acoustic center. The theoretical curve is the one processed by BEM using the detailed mesh. One can see that the two curves agree well: they are within 0.6 dB of each other for frequencies lower than 100 Hz, within 1 dB up to 200 Hz, and within 2 dB up to 300 Hz.

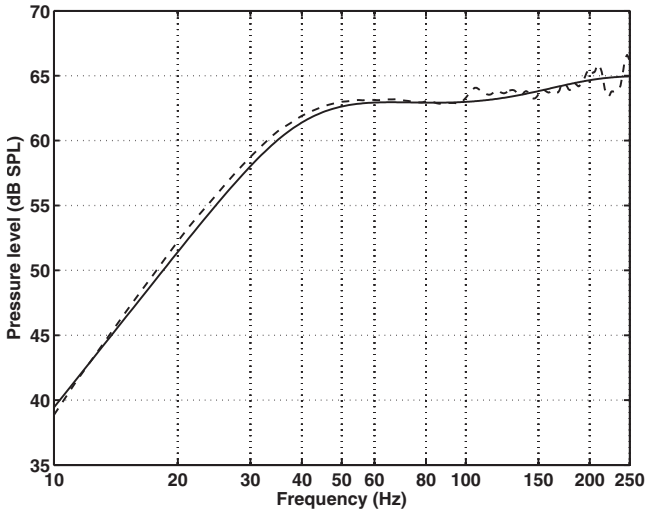


FIG. 10. On-axis pressure level at 4 m: field separation method computed from measurements at 0.9 and 1 m for configuration d (dashed line), BEM using detailed mesh (solid line).

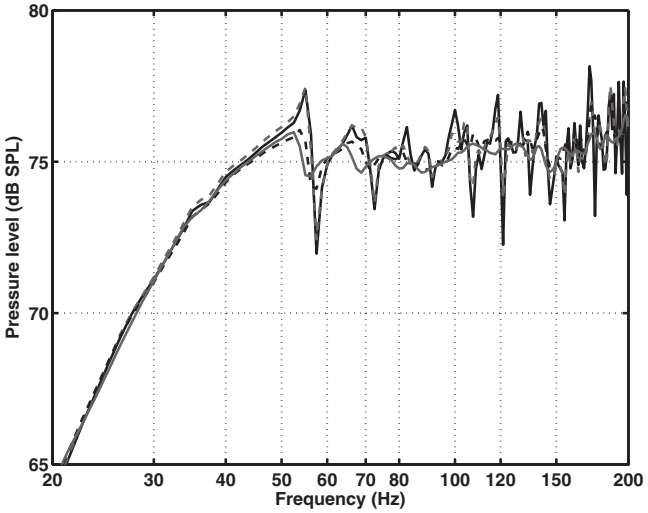


FIG. 11. Outgoing pressure level in dB SPL at position (1.22 m, 80.4°, 0°) computed from measurements at 0.9 and 1 m for configurations: a (gray dashed line), b (black line), c (black dashed line), and d (gray line).

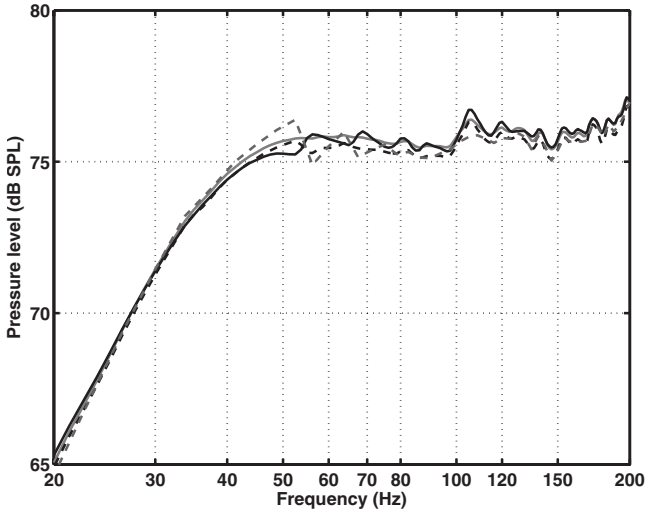


FIG. 12. Outgoing pressure level in dB SPL at position (1.22 m, 80.4°, 0°) for configuration d with 100 point mesh computed from measurements: at 0.6 and 0.7 m (gray line), at 0.7 and 0.8 m (black line), at 0.8 and 0.9 m (dashed line), and at 0.9 and 1 m (gray dashed line).

Remaining oscillations on the recovered response correspond to frequency zones where the incident pressure field on the subwoofer has the largest values. Consequently, small errors on the pressure and velocity estimations can produce larger errors at these frequencies. Moreover, with high incident field magnitudes, the scattered field may not be negligible anymore, especially at higher frequencies. This latter effect has to be added to the measurement errors. It could be removed by using an integral formulation as proposed by Langrenne.[18] However, using BEM instead of spherical harmonics expansion greatly increases the measurement processing and may sometimes be tricky, limiting its practical interest at the current state of the technique.

The effect of the room reverberation is shown on Fig. 11. The separation process has been applied to the data measured at $r=0.9$ and 1 m. For the highest reverberation times (cases a and b), the modal behavior of the testing room is still clearly visible although the separation process has been applied. For the lowest reverberation times (cases c and d), remaining effect of the room is acceptable (oscillations are in the range of 1 dB up to 150 Hz). The method is therefore compatible with many usual rooms and requires only a slight wall treatment in most available facilities.

The influence on the separation process of the measurement distance from the source is shown in Fig. 12 for configuration b. When moving away from the source, the outgoing pressure to ingoing pressure ratio decreases. Then, the efficiency of the proposed method decreases. This drawback is corroborated by curves plotted in Fig. 12: effect of the testing room becomes significant when measurement distance increases.

For the practical implementation of the field separation method, an important parameter is the measurement point number required to obtain reliable results. Figure 13 shows outgoing pressure level for several meshes: 16, 36, 64, and 100 points for configuration d. For this particular comparison, spherical harmonic expansions have been performed with a maximum order $N=3$ to be compatible with 16 point mesh processing. This configuration leads to a maximum frequency of 250 Hz when satisfying the criteria $ks=N=3$ with $s=0.65$ m. For the tested source, the results obtained from the four data meshes are comparable. In the 20–200 Hz frequency band, the maximum difference between the four curves is lower than 1 dB. A quite small number of measurement points may therefore be used with good separation results.

Effect of the maximum expansion order N has also been investigated for configuration d with measurements at 0.6 and 0.7 m. Maximum deviation between expansions at $N=9$ and $N=3$ does not exceed 0.23 dB (0.003 dB for $N=5$ and 9). This result is, however, related to the behavior of the device under test: a closed box has a quasi-monopolar radiation pattern and therefore requires a shorter expansion than a more complicated source. We had to stick to such a simple source to be able to build a reliable model for assessing the method.

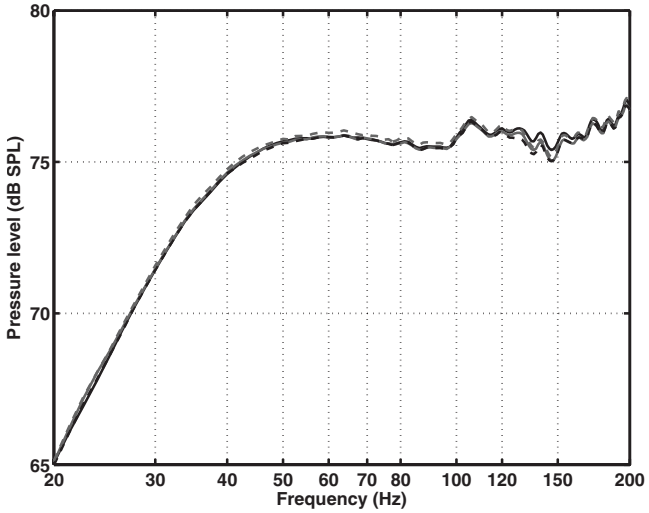


FIG. 13. Outgoing pressure level in dB SPL at position (1.22 m, 80.4°, 0°) for configuration d computed from measurements at 0.6 and 0.7 m for: 100 point mesh (black line), 64 point mesh (gray line), 36 points (dashed line), and 16 points (gray dashed line).

Anyway, the present results emphasize the practical interest of the proposed method considering that the measurements have been performed in a room with common size and reverberation time. For simple sources, a reduced number of measurement points suffice to isolate the outgoing field with an accuracy which seems better than what can be obtained with even a large anechoic room, and post-processing is fast with today's computers. On the negative side, this method requires carefully calibrated sensors and a robot to move them. A mesh of fixed microphones could be an alternative, but this would increase the calibration burden.

This work is still restricted to small subwoofers, as they have a simple radiation pattern, and their size allows to reach a satisfying compromise between good separation (requiring a short measurement distance) and almost far-field conditions. Further work should involve different sources, including several vibrating surfaces (e.g., vented box, active +passive radiators, or small loudspeakers arrays). The influence of the position of the source within the measurement surface should probably also be investigated with such a variety of sources.

ACKNOWLEDGMENTS

This work has been supported by the French National Research Agency [Agence nationale de la recherché (ANR)] as part of the PARABAS project (Grant No. ANR-06-BLAN-0081). The authors wish to thank Nicolas Steinmetz for his enthusiastic efforts in performing the measurements. The authors also wish to thank one of the reviewer for constructive remarks which helped to improve the final paper.

[1]International Electrotechnical Commission, Sound system equipment—Part 5: Loudspeakers, IEC 268-5, 2nd ed. (1989).
[2]R. C. Heyser, "Acoustical measurements by time delay spectrometry," J. Audio Eng. Soc. **15**, 370–382 (1967).
[3]J. M. Berman and L. M. Fincham, "The application of digital techniques to measurements of loudspeakers," J. Audio Eng. Soc. **25**, 370–384 (1977).
[4]S. I. Soterios, "A method for extending the measured response of a loudspeaker towards the low frequencies, with the use of cepstral analysis," in AES 92nd Convention (1992), Preprint No. 3233.
[5]D. B. Keele, "Low-frequency loudspeaker assessment by nearfield sound-pressure measurement," J. Audio Eng. Soc. **22**, 154–162 (1974).
[6]R. T. Kessel, "Predicting farfield pressures from nearfield loudspeaker measurements," in 85th Convention of the Audio Engineering Society (1988), Preprint No. 2729.
[7]R. H. Small, "Simplified loudspeaker measurements at low frequencies," J. Audio Eng. Soc. **20**, 28–33 (1972).
[8]J. Christophorou, "Low-frequency loudspeaker measurements with an accelerometer," J. Audio Eng. Soc. **28**, 809–816 (1980).
[9]C. J. Struck and S. F. Temme, "Simulated free field measurements," J. Audio Eng. Soc. **42**, 467–482 (1994).
[10]J. Vanderkooy and S. P. Lipshitz, "Can one perform quasi-anechoic measurements in normal rooms?," 125th Convention of the Audio Engineering Society (2008), Preprint No. 7525.
[11]H. Lu, S. Wu, and D. B. Keele, "High-accuracy full-sphere electro acoustic polar measurements at high frequencies using the HELS method," 121st Convention of the Audio Engineering Society (2006), Preprint No. 6981.
[12]Z. X. Wang and S. F. Wu, "Helmholtz equation least-squares method for reconstructing the acoustic pressure field," J. Acoust. Soc. Am. **102**, 2020–2032 (1997).
[13]G. Weinreich and E. B. Arnold, "Method for measuring acoustic radiation fields," J. Acoust. Soc. Am. **68**, 404–411 (1980).
[14]I. E. Tsukernikov, "Calculation of the field of a sound source in a bounded space," Sov. Phys. Acoust. **35**, 304–306 (1989).
[15]C. Langrenne, "Méthodes de résolution du problème inverse pour l'identintification de sources en milieu confiné (Regularization methods of the acoustic inverse problem for the characterization of sources in bounded space)," Ph.D. thesis, Maine University, Le Mans, France (1997), in French.
[16]E. G. Williams, *Fourier Acoustics* (Academic, London, 1999).
[17]M. Melon, C. Langrenne, D. Rousseau, and P. Herzog, "Comparison of four subwoofer measurement techniques," J. Audio Eng. Soc. **55**, 1077–1091 (2007).
[18]C. Langrenne, M. Melon, and A. Garcia, "Boundary element method for the acoustic characterization of a machine in bounded noisy environment," J. Acoust. Soc. Am. **121**, 2750–2757 (2007).
[19]T. Hansen, "Spherical expansions of time domain acoustic fields: Application to near field scanning," J. Acoust. Soc. Am. **98**, 1204–1215 (1995).
[20]B. Rafaely, "Analysis and design of spherical microphone arrays," IEEE Trans. Speech Audio Process. **13**, 135–143 (2005).
[21]F. Jacobsen and H.-E. de Bree, "A comparison of two different sound intensity measurement principles," J. Acoust. Soc. Am. **118**, 1510–1517 (2005).
[22]W. Leach, "Loudspeaker voice-coil inductance losses: Circuit models, parameter estimation and effect on frequency response," J. Audio Eng. Soc. **50**, 442–450 (2002).
[23]J. G. Knudsen and M. H. Jensen, "Low-frequency loudspeaker models that include suspension creep," J. Audio Eng. Soc. **41**, 3–18 (1993).
[24]J. C. Le Roux, "Le haut-parleur électrodynamique: Estimation des paramètres électroacoustiques aux basses fréquences et modélisation de la suspension (The electrodynamic loudspeaker: Determination of the low-frequency electroacoustic parameters and modelization of the suspension)," Ph.D. thesis, Maine University, Le Mans, France (1994), in French.

Full bandwidth calibration procedure for acoustic probes containing a pressure and particle velocity sensor

Tom G. H. Basten[a)]
TNO Science and Industry, P.O. Box 155, 2600 AD Delft, The Netherlands

Hans-Elias de Bree
Department of Vehicle Acoustics, HAN University, Ruitenberglaan 26, 6826 CC, Arnhem, The Netherlands

(Received 13 October 2008; revised 6 October 2009; accepted 4 November 2009)

Calibration of acoustic particle velocity sensors is still difficult due to the lack of standardized sensors to compare with. Recently it is shown by Jacobsen and Jaud [J. Acoust. Soc. Am. **120**, 830–837 (2006)] that it is possible to calibrate a sound pressure and particle velocity sensor in free field conditions at higher frequencies. This is done by using the known acoustic impedance at a certain distance of a spherical loudspeaker. When the sound pressure is measured with a calibrated reference microphone, the particle velocity can be calculated from the known impedance and the measured pressure. At lower frequencies, this approach gives unreliable results. The method is now extended to lower frequencies by measuring the acoustic pressure inside the spherical source. At lower frequencies, the sound pressure inside the sphere is proportional to the movement of the loudspeaker membrane. If the movement is known, the particle velocity in front of the loudspeaker can be derived. This low frequency approach is combined with the high frequency approach giving a full bandwidth calibration procedure which can be used in free field conditions using a single calibration setup. The calibration results are compared with results obtained with a standing wave tube. [DOI: 10.1121/1.3268608]

PACS number(s): 43.58.Vb, 43.58.Fm [AJZ] Pages: 264–270

I. INTRODUCTION

For several years direct measurement of the acoustic particle velocity in air is possible by a particle velocity transducer called the Microflown.[1,2] In combination with a small pressure microphone, a very compact pressure-velocity sound probe is available which can be used for direct measurement of impedance,[3] sound intensity,[4] or sound energy.[5,6] To make proper measurements, one has to know how the output voltage of the sensor relates to the value of the acoustic particle velocity at the measurement location. Also the phase between pressure and velocity sensor has to be accurately known, especially for intensity measurements. Until recently, the calibration of this pressure-velocity probe for the full acoustic bandwidth was quite cumbersome. No reference particle velocity sensor exists, so it is not possible to calibrate a particle velocity sensor by comparing it with a standard reference sensor as is done with microphones. The solution is to create an environment where the acoustic impedance is known. The calibration technique used so far was based on a standing wave tube,[7,8] which could only be used for low frequencies. Recently a high frequency approach was introduced by Jacobsen and Jaud based on a spherical source.[9] Although best calibration results were achieved in an anechoic room of good quality, acceptable results were obtained in an ordinary room, which makes this approach very attractive for a standard calibration procedure. However, for low frequencies the approach described by Jacobsen and Jaud gives unreliable results. Therefore, the method is extended to lower frequencies. In the current paper, a full bandwidth calibration technique for the pressure-velocity sound probe is described. The technique is composed of two parts: one for the lower frequency range and one for the higher frequencies. The results of both calibration steps are combined in the final step. The complete procedure can be performed with a single calibration setup. First, some background of the high and low frequency calibration steps will be given.

II. HIGH FREQUENCY CALIBRATION

For the calibration of a pressure-velocity probe, a special loudspeaker is designed that has a known acoustic impedance,[9] see Fig. 1. The pressure velocity probe and a reference pressure microphone with a known sensitivity (in the present case a G.R.A.S. 40AC with G.R.A.S. 26AF preamplifier is used) will be positioned at certain distance in front of the speaker. The microphone and the pressure-velocity probe are positioned at nearly the same position, see Fig. 1. With the pressure measured with the reference microphone and the known normalized impedance at the measurement location, the pressure-velocity probe can be characterized.

The loudspeaker consists of a hard plastic sphere in which a loudspeaker is placed. This spherical loudspeaker can be modeled as a sphere with radius a and a moving piston with radius b. The relation between sound pressure and particle velocity (the acoustic impedance) on the axis of the piston is given by[10]

[a)]Author to whom correspondence should be addressed. Electronic mail: tom.basten@tno.nl

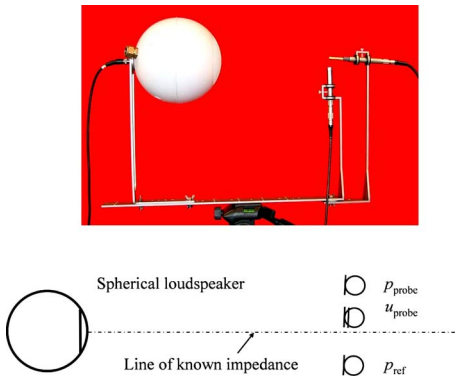


FIG. 1. (Color online) Measurement setup for the piston in a sphere for high frequency calibration.

$$Z_{\text{sphere}}(r) = -j\rho c \frac{\sum_{m=0}^{\infty} (P_{m-1}(\cos\alpha) - P_{m+1}(\cos\alpha)) \dfrac{h_m(kr)}{h'_m(ka)}}{\sum_{m=0}^{\infty} (P_{m-1}(\cos\alpha) - P_{m+1}(\cos\alpha)) \dfrac{h'_m(kr)}{h'_m(ka)}}, \tag{1}$$

where r is the distance from the centre of the sphere, $\alpha = \arcsin(b/a)$, a is the radius of the sphere, b is the radius of the loudspeaker, P_m is the Legendre function of the order m, h_m is the spherical Hankel function of the second kind and order m, h'_m is its derivative, ρ is the density of the air, and c is the speed of sound in air. The calibration results are normalized with the specific impedance ρc. The sensitivity of the particle velocity sensor will therefore be given in mV/Pa*, where 1 Pa* corresponds to 1 Pa/$\rho c \approx 2.4$ mm/s. The specific impedance depends on environmental conditions, such as temperature, atmospheric pressure, and relative humidity.

Although the (normalized) impedance on the axis of the spherical source is described by the complex expression (1), the resulting impedance is quite similar to the acoustic impedance at a certain distance r of a monopole source, given by

$$Z(r) = \rho c \frac{jkr}{1 + jkr}. \tag{2}$$

The ratio of the impedance of the piston (diameter piston 6.5 cm) in a sphere (diameter 20.5 cm), described by Eq. (1) and the acoustic impedance of a monopole source [Eq. (2)], is given in Fig. 2 for varying distance from the front of the sphere. As can be expected, the difference between both impedance descriptions is largest at a short distance from the source and for low frequencies.

For the calibration procedure in the current paper, the standard pressure-velocity probe is used. This probe has a 0.5 in. packaging for protection and increases the sensitivity of the particle velocity sensor. However, this packaging means that the PU probe can be used under 10 kHz. Above these frequencies, the packaging has too much influence on the measured sound field.

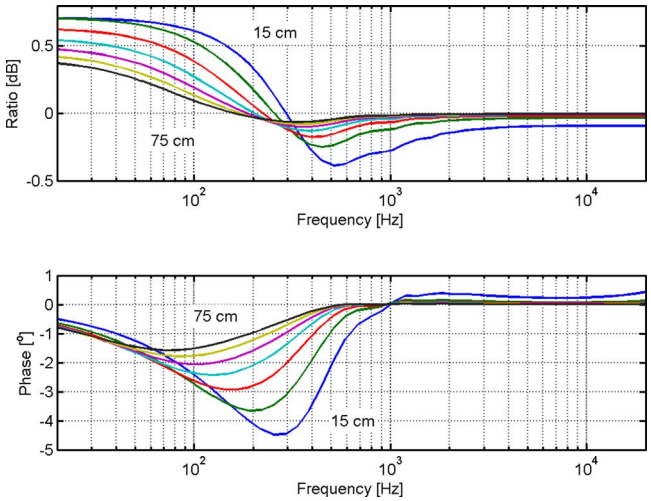


FIG. 2. (Color online) The ratio between the impedance at a certain distance in front of a piston in a sphere and the impedance in front of a monopole (diameter piston 6.5 cm, diameter sphere 20.5 cm). The distance from the front of the sphere is increased from 15 to 75 cm in steps of 10 cm. The dB scale in the upper figure is defined by 20 log (amplitude ratio).

III. LOW FREQUENCY CALIBRATION

At lower frequencies, the calibration procedure described in Sec. II has some drawbacks. In an ordinary room, at lower frequencies the background noise has higher pressure levels than the noise that is generated by the source. This is shown in Fig. 3 where some measurement results are given. The output of the microphone with the source on and the output when the source is switched off are given in the upper figure (all measurements are done with a Siglab 20–42 signal analyzer with a frequency resolution of 1.5625 Hz). As can be seen, the background noise is dominant for frequencies below 50 Hz. For the particle velocity, this effect is not observed, see the lower figure. There, the difference between background noise and the emitted signal from the loudspeaker is much higher. The main reason for this observation is that in the near field of a sound source the ratio between particle velocity and pressure increases. For a monopole this can be easily observed by applying Eq. (2). So

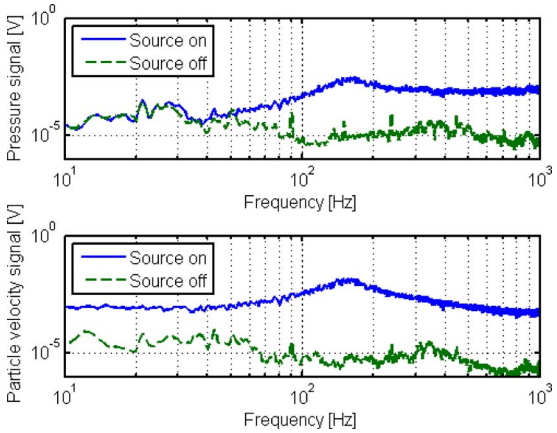


FIG. 3. (Color online) Pressure and velocity signals just in front of the moving piston with source switched on and off. At low frequencies, the microphone signal is clearly more affected by background noise than the particle velocity signal.

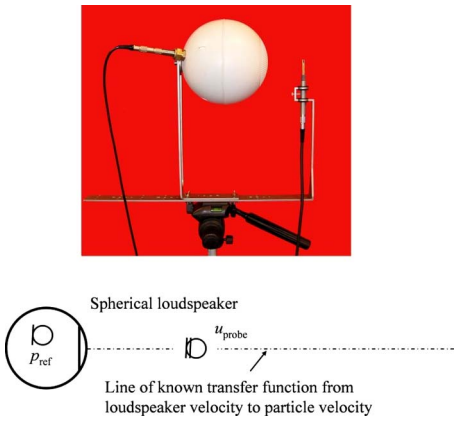


FIG. 4. (Color online) Measurement setup for the low frequency calibration. The pressure microphone is now put inside the spherical loudspeaker.

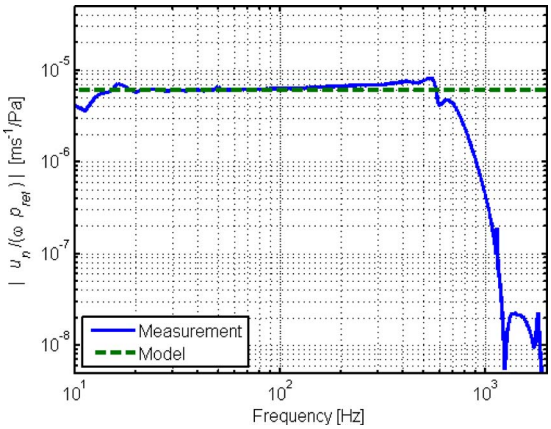


FIG. 5. (Color online) Transfer function between pressure in the sphere and the displacement of the membrane. The displacement is calculated by $|u_n/\omega|$.

for low frequencies, when the probe is in the near field of the source, the ratio between particle velocity and acoustic pressure is significant. For the background noise, for which the probe is in the far field, the ratio between pressure and particle velocity is similar as for high frequencies.

Therefore, in an anechoic room with low background noise, the method described in Sec. II will work down to 50 Hz.[9] In an ordinary room, the background noise is higher and the method starts to work properly from 100 to 200 Hz. The sound pressure microphone, however, can be calibrated in the frequency range 20–10 kHz, because its calibration is based on the comparison with the output of the reference microphone which is at the same position. Both sensors are omnidirectional and it is of no consequence if the output is caused by background noise or by the spherical loudspeaker. For the particle velocity sensor, this does not hold because its calibration is based on the known impedance due to the loudspeaker. When the background noise becomes dominant, Eq. (1) cannot be applied anymore. Therefore, depending on the amount of background noise, the velocity sensor can be calibrated in the frequency range from several hundreds of Hertz to about 10 kHz. For the low frequencies, a different approach will be followed, see Fig. 4.

Now the reference microphone is put into a hole in the sphere and tightened with rubber rings. The reference microphone measures the interior pressure variations in the sphere and the relation between the interior pressure and the particle velocity at a distance in front of the sphere will be used for calibration. The advantage is that the pressure inside the sphere is sufficiently high down to lower frequencies and there is a simple relation between the interior pressure and the particle velocity at the probe position. For low frequencies (well below the first internal acoustic resonance of the sphere), the frequency response between the interior pressure and velocity of the membrane of the loudspeaker is inversely proportional to the frequency. Due to the continuity condition, the particle velocity just in front of the membrane is similar to the velocity of the membrane itself. Therefore, the relation between the sound pressure in the sphere and the particle velocity u_n just in front of the membrane is given by

$$u_n = -\frac{j\omega V_0}{\gamma A_0 p_0} p_{\mathrm{ref}}. \tag{3}$$

where ω is the angular frequency, V_0 is the interior volume of the sphere, A_0 is the surface area of the moving piston, p_0 is the ambient pressure, and γ is the ratio of specific heats (1.4 for normal air). It is assumed that the compression and rarefaction of the air in the sphere is an adiabatic process.

To verify that this relation is valid, a simple experiment is performed, see Fig. 5. Here the transfer function between the sound pressure in the sphere and the displacement of the membrane is given. The displacement is measured by integrating twice the output of an accelerometer, which for this measurement is glued on the membrane of the loudspeaker. According to Eq. (3), the transfer function between pressure and displacement is frequency independent. In Fig. 5, it is clear that this independency is found from 10 Hz up to about 300 Hz. In this figure also the theoretical model is given, where it is assumed that the internal volume of the sphere is 2.8×10^{-3} m^3 (corresponding to 17.5 cm internal diameter) and that the surface area is 3.3×10^{-3} m^2 (corresponding to a piston with diameter of 6.5 cm).

For higher frequencies, the method fails because of internal acoustic modes in the sphere. If the wavelength is smaller than the dimensions of the sphere, the sound pressure is not uniform in the sphere and the simple relation given by Eq. (3) is not valid anymore.

Because the characteristics of an acoustic field around a sphere with a moving piston are known, also the particle velocity at a certain distance in front of the moving piston can be derived when the particle velocity just in front of the piston is known. Therefore, the particle velocity at the measurement position can be related to the interior pressure of the sphere.

The relation between the particle velocity just in front of the piston, which equals the normal velocity of this piston (u_n) and the particle velocity at a distance r from the centre of the sphere, is given by[10]

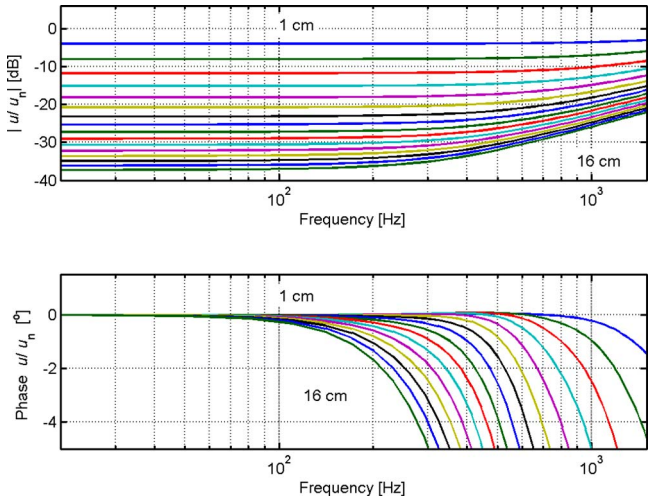


FIG. 6. (Color online) Transfer function between particle velocities at several distances from the moving piston and the particle velocity just in front of the piston as given by Eq. (4). The diameter of the sphere is 20.5 cm and the piston diameter is 6.5 cm. The dB scale in the upper figure is defined by 20 log (amplitude ratio).

$$u(r) = -\frac{u_n}{2}\sum_{m=0}^{\infty}\left(P_{m-1}(\cos\alpha) - P_{m+1}(\cos\alpha)\right)\frac{h_m'(kr)}{h_m'(ka)}. \quad (4)$$

The ratio of the surface velocity of the loudspeaker (u_n) to the particle velocity $u(r)$ measured from 1–16 cm in front of a piston (6.5 cm) in a sphere (20.5 cm) calculated with Eq. (4) is given in Fig. 6. It is clear that for lower frequencies (<200 Hz), the amplitude is almost independent of the frequency and the phase difference is almost zero. This means that the phase difference between the reference pressure microphone in the sphere and the particle velocity sensor has to be 90° [see Eq. (3)]. This fact can be used for calibration.

As a check for this behavior, a measurement is performed. The transfer function is measured between the particle velocity 1 cm in front of the piston and the particle velocity at 14 cm, see Fig. 7. As can be clearly seen, the phase difference at low frequencies (<200 Hz) is almost zero, as is also predicted by the model, for which the results are also plotted in this figure. The measurement results are in very good agreement with the model.

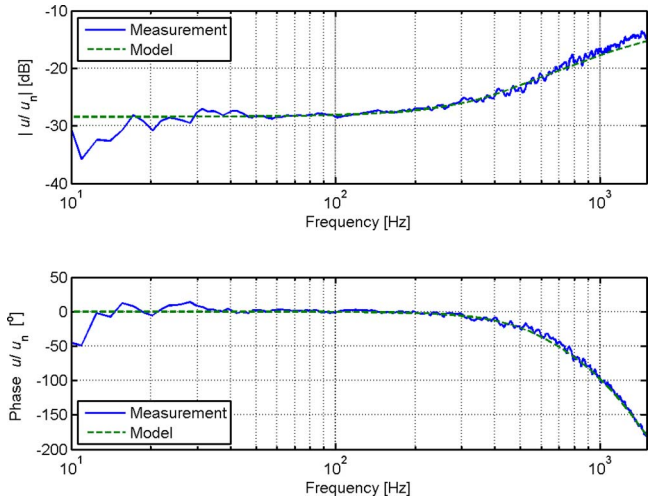


FIG. 7. (Color online) Measured and modeled transfer functions of particle velocities measured at 1 cm and at 14 cm in front of the large spherical loudspeaker. (Upper) Modulus. (Lower) Phase.

Combining Eqs. (3) and (4) gives a relation between the particle velocity at position r and the pressure in the piston. This means that when the interior acoustic pressure of the sphere is measured, the (normalized) particle velocity at distance r is exactly known. This known particle velocity value is used for calibration. This can only be done for frequencies, well below the first internal resonance frequency of the sphere. For higher frequencies, the pressure inside the sphere is not uniform anymore. With a correction, taking into account the effect of the first internal resonance frequency of the sphere, the low frequency approach can be applied up to higher frequencies.

IV. COMBINING THE LOW AND HIGH FREQUENCY APPROACHES

Two steps to determine the full bandwidth sensitivity and phase characteristics of the pressure-velocity probe are described, one for the high frequency range and one for the low frequency range. These two results are combined in the final step. It will be shown that the high frequency approach and low frequency approach give similar results in the medium frequency range around 300 Hz.

The complete procedure is now as follows. First, a measurement is performed with the high frequency configuration. That means that the reference pressure sensor and the pressure-velocity probe are placed at a known position on the axis of the spherical source, while white noise is emitted. The transfer functions between all sensors are measured. The ratio of the reference microphone output (in volts) and the output of the pressure sensor (in volts) is directly used to determine the sensitivity S_p (mV/Pa) of the pressure sensor:

$$S_p\left[\frac{\text{mV}}{\text{Pa}}\right] = \frac{p}{p_{\text{ref}}}\left[\frac{\text{V}}{\text{V}}\right]\cdot S_{\text{ref}}\left[\frac{\text{mV}}{\text{Pa}}\right], \quad (5)$$

where S_{ref} is the known sensitivity of the reference microphone (14 mV/Pa), which is assumed to be independent of frequency in the frequency range of interest and independent of the static pressure.

The particle velocity cannot directly related to the reference sensor output. Here the model of the impedance of the piston in a sphere, Eq. (1), has to be used. The sensitivity of the particle velocity sensor S_u [mV/Pa*] is calculated by

$$S_u\left[\frac{\text{mV}}{\text{Pa*}}\right] = \frac{u}{p_{\text{ref}}}\left[\frac{\text{V}}{\text{V}}\right]\cdot Z_{\text{sphere}}\left[\frac{\text{Pa}}{\text{Pa*}}\right]\cdot S_{\text{ref}}\left[\frac{\text{mV}}{\text{Pa}}\right]. \quad (6)$$

Now the sensitivity of the pressure sensor of the probe is known for the complete frequency range and the sensitivity of the particle velocity sensor only for the high frequency range.

Next, the reference pressure sensor is put in the hole of sphere and tightened with rubber rings so that the sphere is leakage-free. The pressure-velocity probe is placed at a known position in front of the moving membrane, but closer than in the previous step (about 1 cm in front of the moving

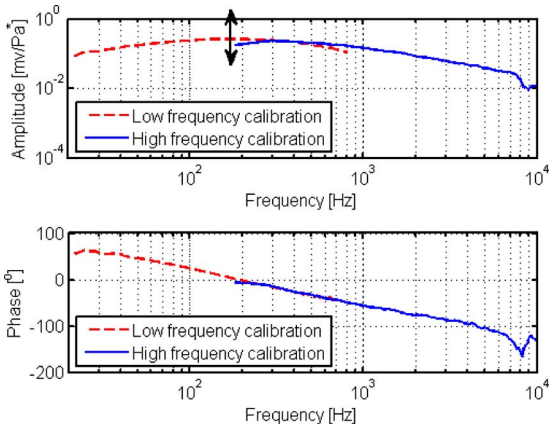


FIG. 8. (Color online) Low and high frequency calibrations, combined.

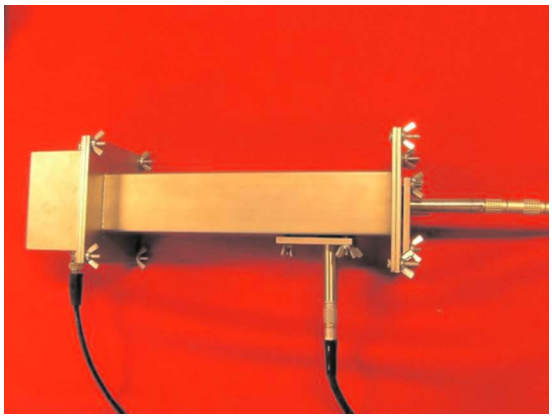


FIG. 9. (Color online) Small standing wave tube for low frequency calibration of the pressure-velocity probe.

piston). Again all transfer functions are measured while white noise is emitted by the source. The particle velocity is now calculated based on Eqs. (3) and (4):

$$S_u\left[\frac{\mathrm{mV}}{\mathrm{Pa}^*}\right] = \frac{u}{p_{\mathrm{ref}}}\left[\frac{\mathrm{V}}{\mathrm{V}}\right] \cdot \frac{u_n}{u}\left[\frac{\mathrm{Pa}^*}{\mathrm{Pa}^*}\right] \cdot \frac{p_{\mathrm{ref}}}{u_n}\left[\frac{\mathrm{Pa}}{\mathrm{Pa}^*}\right] \cdot S_{\mathrm{ref}}\left[\frac{\mathrm{mV}}{\mathrm{Pa}}\right]. \tag{7}$$

In the next step, the low and high frequency curves are connected. For the amplitude, see Fig. 8, upper, the low frequency calibration curve (<350 Hz) is tuned such that it connects with the high frequency curve. The parameters in Eq. (7) to make this connection are $\rho=1.2\ \mathrm{kg/m^3}$, $c=340\ \mathrm{m/s}$, $V_0=2.8\times10^{-3}\ \mathrm{m^3}$ and $A_0=3.3\times10^{-3}\ \mathrm{m^2}$. The phase is continuous over the frequency. For the low frequency approach around 350 Hz, the phase is exactly known. For the high frequency part, a discontinuity can be introduced by the correction term [Eq. (1)]. Within this term, the distance between the probe and the source is the most important parameter which can be used to tune the phase, such that it connects with the low frequency part. The low and high frequency calibration curves have a small overlap around 350 Hz. Around this frequency, both curves can be combined to create a full bandwidth calibration curve.

Before connecting the low and high frequency calibration curves, the complex sensitivities for the high frequency approach are smoothed. Because of the reflections in the ordinary room, the determined sensitivities seem rather noisy. This is avoided by smoothing the results by a moving average filter.[11] Another technique to smooth the curves is a time selective technique. In this way room reflections can be removed by canceling the reflected signals in the impulse response. However, this method is not very accurate for low frequencies.[12]

V. COMPARISON WITH STANDING WAVE TUBE CALIBRATION

Another environment where the acoustic impedance is exactly known is in the interior of a rigidly terminated tube where a sound wave is generated by a speaker,[8] see Fig. 9 and schematically in Fig. 10.

This tube is also known as the standing wave tube and can be used as an alternative for calibration at low frequencies. With the reference pressure sensor at the rigid end of the tube, the pressure and particle velocity are known at each location in the tube. This tube can therefore be used for calibration of the particle velocity sensor. However, it only works down to low frequencies below the cut-off frequency f_c of the tube, which is determined by the diameter of the tube d and the speed of sound:

$$f_c = \frac{c}{1.71d}. \tag{8}$$

Above the cut-off frequency (which is about 4.5 kHz for the current tube), the sound field is not one-dimensional anymore, i.e., the sound pressure across the cross section of the tube is not uniform and the impedance cannot be described anymore with simple relations. Non-ideal behavior, such as damping due to viscous and thermal effects and non-ideal reflection of the rigid termination, can be neglected.[8]

The relation between pressure $p(x)$ and particle velocity $u(x)$ at an arbitrary location in the tube is given by

$$\frac{u(x)}{p(x)} = \frac{j}{\rho c}\tan(k(l-x)). \tag{9}$$

The sensitivity of the pressure microphone S_p and the particle velocity S_u of the probe can be calculated by

$$S_p\left[\frac{\mathrm{mV}}{\mathrm{Pa}}\right] = \frac{p}{p_{\mathrm{ref}}}\left[\frac{\mathrm{V}}{\mathrm{V}}\right]\frac{1}{\cos(k(l-x))}S_{\mathrm{ref}}\left[\frac{\mathrm{mV}}{\mathrm{Pa}}\right], \tag{10}$$

$$S_u\left[\frac{\mathrm{mV}}{\mathrm{Pa}^*}\right] = \frac{u}{p_{\mathrm{ref}}}\left[\frac{\mathrm{V}}{\mathrm{V}}\right]\frac{1}{j\sin(k(l-x))}S_{\mathrm{ref}}\left[\frac{\mathrm{mV}}{\mathrm{Pa}}\right]. \tag{11}$$

A short standing wave tube as given in Fig. 9 is used for calibration for low frequencies. The inner dimensions of the tube are $47\times47\ \mathrm{mm^2}$ and the probe is positioned such that $l\text{-}x=55$ mm. While white noise is emitted by the loudspeaker, the transfer functions p/p_{ref} and u/p_{ref} are measured. The sensitivities are calculated by Eqs. (10) and (11),

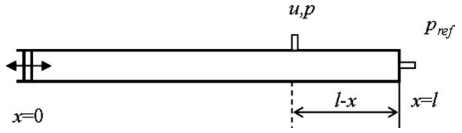


FIG. 10. Schematic representation of the standing wave tube.

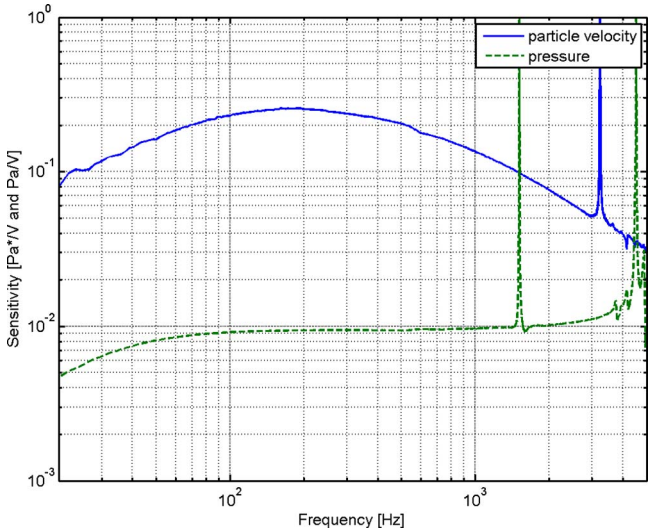


FIG. 11. (Color online) Calculated sensitivities of the particle velocity sensor and the pressure sensor.

where the reference microphone has a sensitivity of 14 mV/Pa. The calculated sensitivities are given in Fig. 11.

The calculated phase response of the pressure and particle velocity sensor relative to the reference pressure sensor is given in Fig. 12.

The calibration results of the particle velocity sensor measured with the sphere setup and standing wave tube setup are compared in Fig. 13. Both the absolute amplitude and phase responses are given. In Fig. 14, the difference between the standing waving tube and the sphere calibration for low frequencies is given. Up to 400 Hz, the error in magnitude is less than 0.5 dB and the phase error is smaller than 4°. The deviations are probably caused by experimental inaccuracies. Especially the offset in the phase error is probably caused by a small leakage during the standing wave tube experiments.

VI. DISCUSSION

The results given in the current paper are all obtained in an ordinary room with background noise due to all kind of

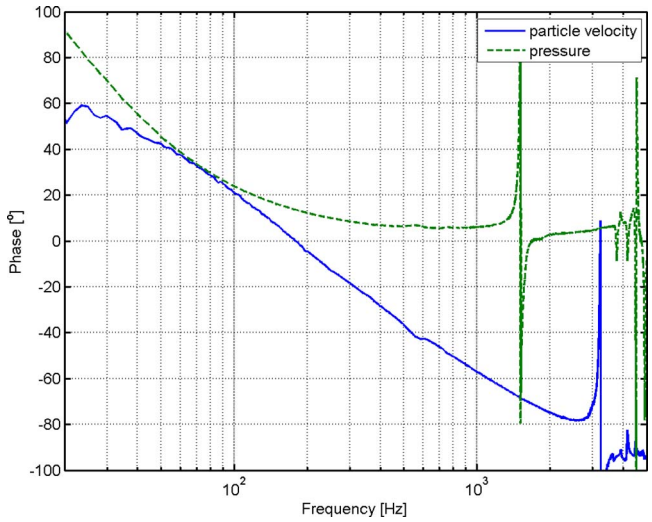


FIG. 12. (Color online) Phase response of particle velocity and pressure sensors of the pressure-velocity probe relative to the reference pressure sensor.

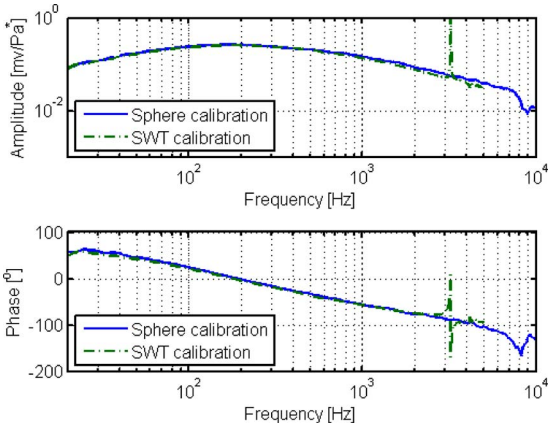


FIG. 13. (Color online) Results of calibration of the particle velocity probe. The curves for the sphere calibration, and the standing wave tube results are given. (Upper) Amplitude. (Lower) Phase.

activities in the neighborhood. Because the measurements are performed very close or even inside the sphere, the calibration results seem to be still very good. The influence of reflections due to the floor, wall, and furniture inside the room are canceled by smoothing the results with a moving average filter. There is a good indication that the proposed method is valid both for low and high frequencies. The comparison with a standing wave tube shows the accuracy for low frequencies and in the paper by Jacobsen and Jaud,[9] the accuracy for higher frequencies is demonstrated. However, because there are many potential errors, a complete sensitivity study for the different parameters involved should be performed. In the paper by Jacobsen and Jaud,[9] it is, for instance, shown that the distance between the probe and the sphere has no significant influence and the calibration performed at several source probe distances agreed within ±0.3 dB and ±1° above 100 Hz. However, the exact distance between probe and source is difficult to determine, but can be found because the phase relation for low frequencies is accurately known. Using this fact, the distance between probe and source can be used as the parameter to tune the phase for high frequencies such that it fits to the low fre-

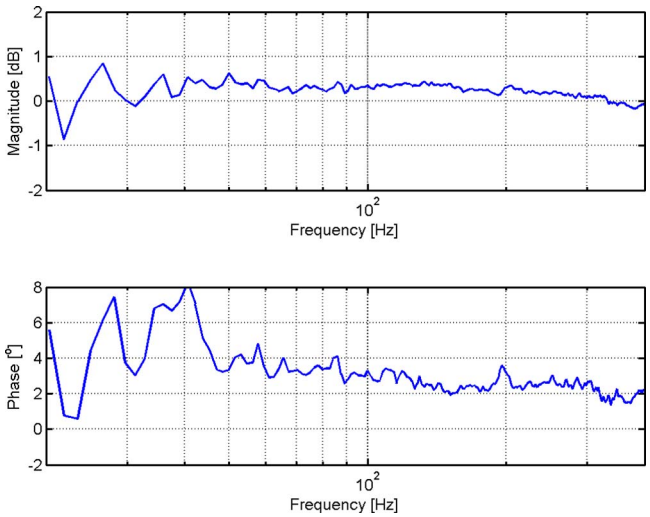


FIG. 14. (Color online) Difference between standing standing wave calibration and the sphere calibration.

quency phase. Therefore, the combination of high and low frequency results make the approach described in this paper very strong.

VII. CONCLUSIONS

A full bandwidth calibration procedure for pressure-velocity probes is described, based on the known sound field around a spherical source, which is modeled as a sphere with a moving piston. For lower frequencies (<1 kHz), the velocity sensor of the probe is calibrated close in front of the spherical loudspeaker and the reference sound pressure microphone is measuring the interior sound pressure of the sphere which is directly related to the piston motion. For higher frequencies (>100 Hz) the pressure-velocity probe is calibrated in front of a spherical loudspeaker in combination with a reference sound pressure microphone. Room reflections are canceled by a moving average technique in the frequency domain. Low and high frequency results, obtained with the same calibration setup, are combined at intermediate frequencies yielding the amplitude and phase response of the pressure velocity probe in the full acoustic bandwidth.

ACKNOWLEDGMENTS

This research is financially supported by the Netherlands Organization for Scientific Research NWO within the Casimir program (stimulating knowledge exchange between research institutes and industry).

[1]H. E. de Bree, "The Microflown, an acoustic particle velocity sensor," Acoust. Aust. **31**, 91–94 (2003).
[2]H. E. de Bree, "An overview of Microflown Technologies," Acta. Acust. Acust. **89**, 163–172 (2003).
[3]R. Lanoye, G. Vermeir, W. Lauriks, R. Kruse, and V. Mellert, "Measuring the free field acoustic impedance and absorption coefficient of sound absorbing materials with a combined particle velocity-pressure sensor," J. Acoust. Soc. Am. **119**, 2826–2831 (2006).
[4]F. Jacobsen and H.-E. de Bree, "A comparison of two different sound intensity measurement principles," J. Acoust. Soc. Am. **118**, 1510–1517 (2005).
[5]D. B. Nutter, T. W. Leishman, S. D. Sommerfeld, and J. D. Blotter, "Measurement of sound power and absorption in reverberation chambers using energy density," J. Acoust. Soc. Am. **121**, 2700–2710 (2007).
[6]F. Jacobsen, "Measurement of total sound energy density in enclosures at low frequencies," in Acoustics08, Paris, France (2008).
[7]D. Stanzial and D. Bonsi, "Calibration of the p-v Microflown® probe and some considerations on the physical nature of sound impedance," in Euronoise, Naples, Italy (2003).
[8]R. Raangs, T. Schlicke, and R. Barham, "Calibration of a micromachined particle velocity microphone in a standing wave tube using a LDA photon-correlation technique," Meas. Sci. Technol. **16**, 1099–1108 (2005).
[9]F. Jacobsen and V. Jaud, "A note on the calibration of pressure velocity sound intensity probes," J. Acoust. Soc. Am. **120**, 830–837 (2006).
[10]E. G. Williams, *Fourier Acoustics: Sound Radiation and Nearfield Acoustical Holography* (Academic, New York, 1999).
[11]H. E. de Bree, M. Nosko, and E. Tijs, "A handheld device to measure the acoustic absorption in situ," in Fifth International Styrian Noise, Vibration and Harshness Congress, Graz, Austria (2008).
[12]H. E. de Bree, R. Lanoye, S. de Cock, and J. van Heck, "In situ, broad band method to determine the normal and oblique reflection coefficient of acoustic materials," in SAE Noise and Vibration Conference and Exhibition, Grand Traverse, MI (2005), SAE Technical Paper No. 2005-01-2443.

Design of an efficient music-speech discriminator

Lorenzo J. Tardón,[a] Simone Sammartino, and Isabel Barbancho
Departamento de Ingeniería de Comunicaciones, E.T.S. Ingeniería de Telecomunicación, Universidad de Málaga, Campus Universitario de Teatinos s/n, E29071 Málaga, Spain

(Received 21 April 2009; revised 23 September 2009; accepted 26 September 2009)

In this paper, the problem of the design of a simple and efficient music-speech discriminator for large audio data sets in which advanced music playing techniques are taught and voice and music are intrinsically interleaved is addressed. In the process, a number of features used in speech-music discrimination are defined and evaluated over the available data set. Specifically, the data set contains pieces of classical music played with different and unspecified instruments (or even lyrics) and the voice of a teacher (a top music performer) or even the overlapped voice of the translator and other persons. After an initial test of the performance of the features implemented, a selection process is started, which takes into account the type of classifier selected beforehand, to achieve good discrimination performance and computational efficiency, as shown in the experiments. The discrimination application has been defined and tested on a large data set supplied by Fundación Albéniz, containing a large variety of classical music pieces played with different instrument, which include comments and speeches of famous performers.
 [DOI: 10.1121/1.3257204]

PACS number(s): 43.60.Dh, 43.60.Lq, 43.75.Xz [WMC] Pages: 271–279

I. INTRODUCTION

For many years, the problem of discrimination between music and speech has been an important task for audio/video processing research groups due to the growing need of supplying classified audio archives over large data sets of radio broadcasts, musical recordings, etc. Nowadays, there is a strong source of demanding applications, some of them due to the increasing web access to audio-video contents, based on tagged audio-video retrieval; this is the specific context for which we develop our system. The search and definition of efficient descriptors for the discrimination of audio obtained by different sources are handled since 1960s (Ref. 1) and since then, many authors dealt with the problem from different points of view. In Ref. 2, a simple electronic setup for audio signal pre-processing and features extraction is shown, which employs a fuzzy combiner based on two specific features for real time discrimination of speech and music (SMD); in the paper by Saunders,[3] a method for SMD in radio broadcast audio is shown; it is based on the definition of a threshold on a measure of the skew of the distribution of the zero-crossing rate (ZCR). In the paper by Carey *et al.*,[4] a more general view of features for SMD is presented; the performance of some of them is evaluated using a simple classifier based on the extracted Gaussian mixture model (GMM) parameters,[5] although the work is not intended to a specific application. In Ref. 6, another point of view is considered: how to deal with encoded audio; in this case, MPEG-1 audio is considered and the volume of segments, directly extracted from coded data, is used in the discrimination. In Ref. 7, a multilayer perceptron is used to perform discrimination, using, as features, the mean and variance of the samples of the Mel-filtered spectrum. Goodwin and Laroche[8] argued on the importance of the features selected for SMD and proposed to incorporate a wide variety of features to the discrimination system and, also, linear discriminant analysis is said to be applied to reduce dimensionality of feature vectors, and this paper is focused on the design of the cost functions for a dynamic programming algorithm to find data clusters. In Ref. 9, the combination of features extracted, related to root mean square (rms) measures and ZCR, is the core of the study, where the behavior of the features is, in this case, analyzed in detail. In this paper, means and variances of several features are considered, and a classification based on some specific tests is described.

In this paper, we consider some of the ideas of these authors and others, so we will pay attention to a number of features which can be useful for discrimination and we make an analysis of their behavior before using them blindly in our system. We also consider the choice of a simple and efficient classification strategy and we search for an optimal combination of the features for the classification according to a certain criterion.

This paper is organized as follows: in Sec. II, the overall set of features, mostly extracted from the selected bibliography, is shown and an analysis of the behavior of the most relevant ones is done. In Sec. III, we present the type of classifier selected for the target application, and in Sec. IV, the discrimination performances of the implemented features and the selection of features to optimize the performance of the system with the classification strategy chosen are shown. Finally, in Sec. V, some conclusions are drawn.

II. FEATURES FOR SPEECH-MUSIC DISCRIMINATION

As mentioned before, a main stage of a SMD process consists of the selection and extraction of numerical features from the different audio sources. A wide range of features

[a]Author to whom correspondence should be addressed. Electronic mail: lorenzo@ic.uma.es

has been used and proposed for the purpose of SMD, and some of them are extracted from the time domain and others in the frequency domain and their usage reaches other areas of audio processing such as speech recognition[10,11] and music classification.[12]

A common stage previous to feature extraction is the division of the audio data into small pieces in order to classify each of them as speech or music. As this granularity has certain influence on the performance of the features selected, we present the relevant parameters before the description of the features and their analysis. In our implementation, the audio files are sampled at 44 100 Hz and they are divided into segments of 1 s (T_s) of duration, and each descriptor is calculated on a number of portions of these segments; we will call them frames and their duration will be 20 ms (T_f). Using these global parameters, a number of features are extracted to create a feature vector with audio data divided into segments and frames, as described, and then, sample means and variances are obtained for time and frequency domain features. Although high order sampled statistics are more sensitive to noise, the analysis of the sampled variance is important due to the different behavior of voiced and unvoiced parts of speech with respect to different features.[13]

In the following, all the implemented features are discussed, while the ones that were finally selected for the optimal subset are exposed with more detail. Although more features could have been considered, even non-linear combinations of some of them, as in Ref. 9, we have limited our work to the ones shown in the Secs. II A and II B, which lead to adequate discrimination performance. In the following, the input data will be denoted as $\bar{x}$ (we drop the segment index to simplify the notation), the 20 ms length vector will be denoted as $\bar{x}_j=[x_j(1),x_j(2),\ldots,x_j(n)]^T$, where j indexes the frame under analysis, with $j=1,2,\ldots,50$, where 50 is the number of frames per segment in our implementation, and T_s/T_f, in a general case. Hence, for each segment, the mean and standard deviation of each descriptor (generally denoted as d) will be computed and stored in the vector $\bar{d}=[d_1,d_2,\ldots,d_{50}]^T$. Note that, for simplicity, we have omitted the segment index.

A. Selected features for the optimal subset

In this section, we expose a brief description and illustrate the specific behavior of the selected features on the training data set available for the project. Descriptions are explicitly provided to ensure the reproducibility of the experiments and results obtained in this work; thus, specific parameters that may be involved in the definition of the feature extraction process will be defined. The frames and the segments defined to obtain the features will be disjoint and multiplied by a rectangular window function, unless otherwise indicated for selected features.

1. Root mean square

The rms is related to the signal power[14] and it is estimated as

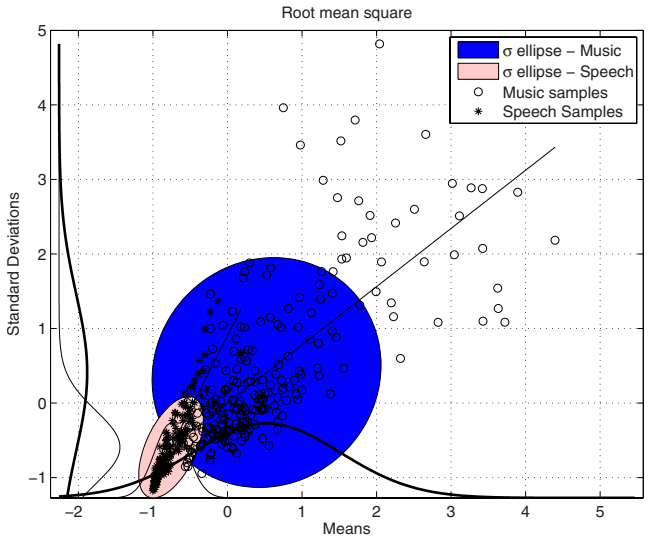


FIG. 1. (Color online) Scatter plot of speech and music root mean square means and standard deviations of a large 1 s length sample set. Linear fits, 1.5σ ellipses, and normal pdfs for means and standard deviations are shown too.

$$\mathrm{rms}_j=\sqrt{\frac{1}{n}\sum_{i=1}^{n}x_j^2(i)}, \tag{1}$$

where n is the number of samples in the audio frame under analysis.

Generally, the feature reflects the higher occurrence of lower power frames in speech and the more homogeneous distribution of sound power in music. In Fig. 1, the scatter plot of means and standard deviations of the rms of a set of hundreds of 1 s length audio segments is shown. Illustrative ellipses have been drawn to facilitate the observation of the different behavior of the features extracted from music and speech samples. The ellipses have their center at the mean of each feature, the semi-axes are 1.5 times the standard deviations of the features (σ_{mean} and σ_{std}), and the rotation angle is obtained using $\mathrm{atan}(\sigma_{\mathrm{std}}/\sigma_{\mathrm{mean}})$; it has been observed that this choice is not heavily affected by the outliers.

Note that the presence of silence or quasi-silence instants, more frequently in speech than in music, leads to a higher ratio between the spreads of means and standard deviations in speech than in music.

2. Zero-crossing rate

The ZCR measures the rate of zero amplitude crossings of a sound signal with respect to its length. It is a good indication of the dominant frequency,[15] and it is highly correlated with the spectral centroid.[16] This feature is computed by counting the number of zero-crossing of the audio signal per second of sample (for a given sampling rate). In our implementation, we use

$$\mathrm{ZCR}_j=\frac{1}{2n}\sum_{i=1}^{n}\left|\mathrm{sign}(x_j(i))-\mathrm{sign}(x_j(i-1))\right|, \tag{2}$$

where $\mathrm{sign}(x_j(i))$ is 1 for positive amplitudes and -1 for negative ones and n is the number of samples in the audio frame.

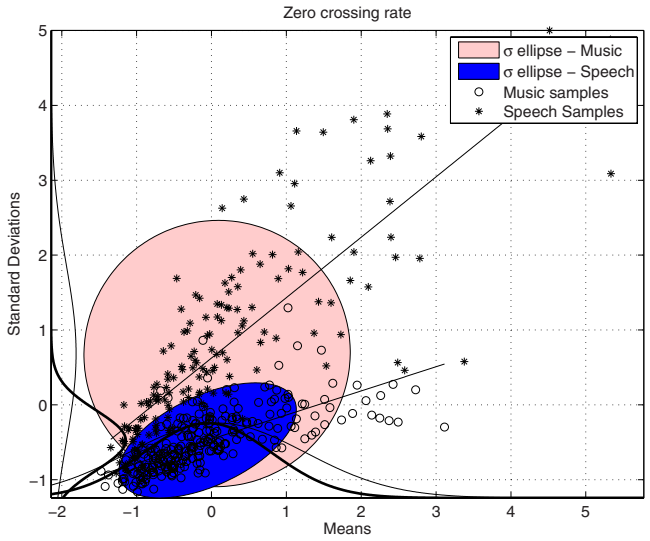


FIG. 2. (Color online) Scatter plot of speech and music zero-crossing rate means and standard deviations of a large 1 s length sample set. Linear fits, 1.5σ ellipses, and normal pdfs for means and standard deviations are shown too.

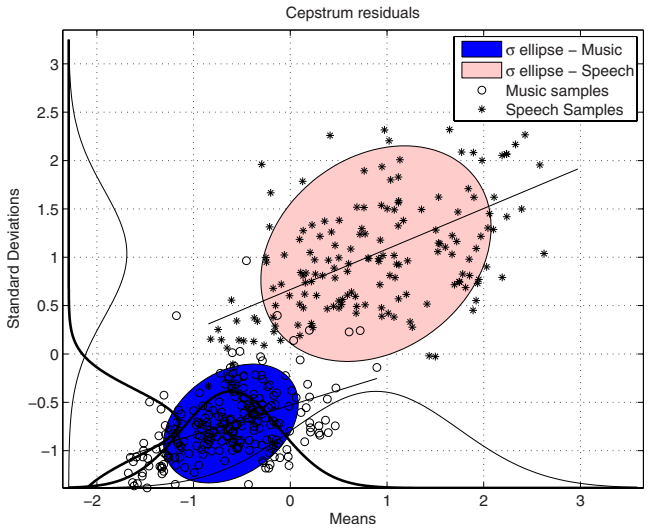


FIG. 3. (Color online) Scatter plot of speech and music cepstrum residual means and standard deviations of a large 1 s length sample set. Linear fits, 1.5σ ellipses, and normal pdfs for means and standard deviations are shown too.

Generally, in speech samples, voiced and unvoiced segments exhibit characteristic low and high zero-crossing rates.[13]

Thus, a relevant difference in variance between the standard deviations of speech and music is easily observed in the scatter plot shown in Fig. 2.

3. Cepstrum residuals

The term cepstrum is a form of assessing the spectrum shape of an audio signal[16] in the so called cepstrum domain.[1] The cepstrum residuals give a measure of the rate of change in the bands of the spectrum.

In this work, the cepstrum residuals (CERs) are computed as the Euclidean distance between the cepstrum of signal and its smoothed form;[17] specifically, we use the following expression:

$$\mathrm{CER}_j = \sqrt{\sum_{k=1}^{m} (C_j(k) - \mathrm{SC}_j(k))^2}, \tag{3}$$

where $C_j(k)$ is the kth element of the raw cepstrum obtained with a discrete Fourier transform (DFT) of length 1024 of the frame j and $\mathrm{SC}_j(k)$ is the kth element of the jth smoothed cepstrum using a moving average method with a five sample window. The cepstrum $\bar{C}_j = [C_j(1), C_j(2), \ldots]^T$ are obtained as follows:

$$\bar{C}_j = \mathrm{DFT}^{-1}\{\ln(|\mathrm{DFT}(\bar{x}_j)|)\}, \tag{4}$$

where DFT( ) represents the function to obtain the discrete Fourier transform and $\bar{x}_j$ is padded with zeros before the calculation of the DFT.

The behavior of CER is shown in Fig. 3. Cepstrum residuals show higher mean and variance for speech samples with respect to music ones, where good class-separability characteristics are exhibited since speech contains higher variations in the frequency contents than music, and their smoothing process causes an increase of the residuals.

4. Spectral flux

The spectral flux (SPF) is a measure of the variation rate between short fractions of audio vectors. From a mathematical point of view, such feature has been computed as the Euclidean norm of the difference between the magnitude vectors of two adjacent audio fragments.[17] In this work, such residuals are extracted for each pair of frames in each segment using a DFT of length 1024.

Basically, the spectral flux measures how fast the signal spectrum changes. In the literature, different interpretations of the behavior of the feature are found,[16,17] but concerning our work, we consider that the type of audio source is a main aspect that influences the outcomes of the feature and its behavior.

Specifically, in the samples supplied by Fundación Albéniz for the development of the discrimination application, the music played or the lyrics sung are short fragments interrupted by teacher's or student's speech with mean and variance of the spectral flux larger in music segments than in speech ones, as shown in Fig. 4.

5. Magnitude

Magnitude (MAG) refers to the behavior of the magnitude of the DFT calculated for each segment (not sectioning it into frames). The mean and standard deviation of the magnitude samples of the DFT are calculated. A detailed description of the algorithm used follows. Let $\bar{X} = \mathrm{DFT}(\bar{x})$, where the DFT is of length m, with k the smallest integer such that $m = 2^k \geq N$, with N the length of the segment, then the mean values and the standard deviation of the bins of $|\bar{X}|$ are obtained. Figure 5 shows, visually, that discrimination based on both mean and variance can be performed; we have found that this feature is one of the simplest features with usable discrimination capabilities between speech and music.

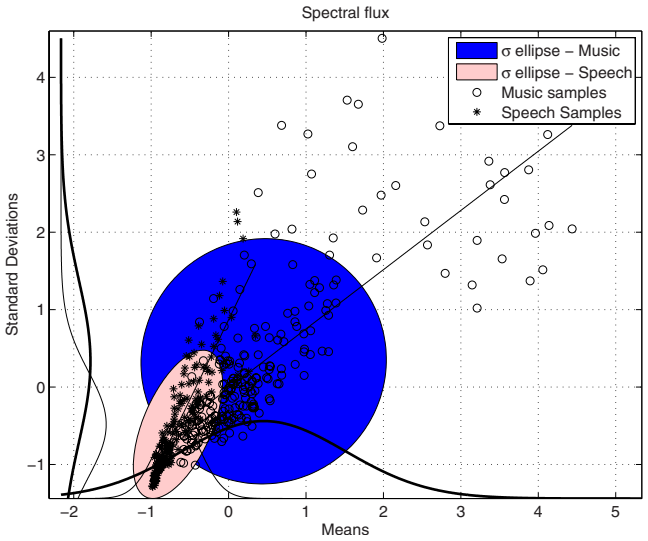


FIG. 4. (Color online) Scatter plot of speech and music spectral flux means and standard deviations of a large 1 s length sample set. Linear fits, 1.5σ ellipses, and normal pdfs for means and standard deviations are shown too.

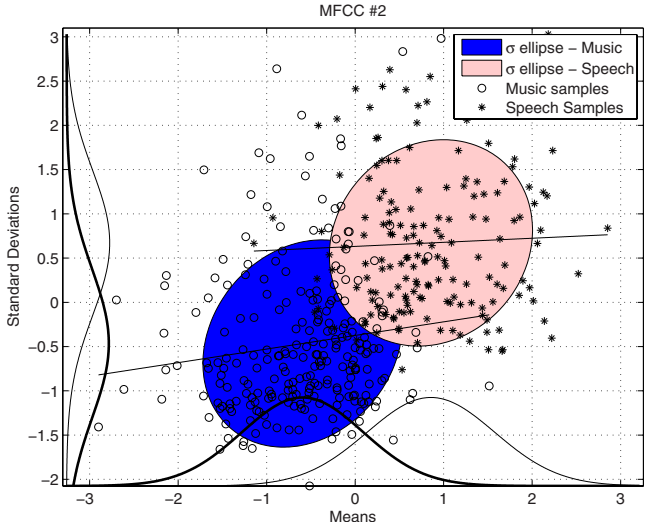


FIG. 6. (Color online) Scatter plot of speech and music Mel frequency cepstral coefficient 2 means and standard deviations of a large 1 s length sample set. Linear fits, 1.5σ ellipses, and normal pdfs for means and standard deviations are shown too.

6. Mel frequency cepstral coefficients

The analysis of the human perception of frequency revealed that logarithmic arrangement of frequency for audio samples is perceived as linearly stepped by a group of listeners in standard conditions. This study gave rise to the so called Mel-scale.[18] One of the main usages of the Mel-scale in audio signal processing is the derivation of the Mel frequency cepstral coefficients (MFCCs) regarded as one of the most efficient tools for speech compression and retrieval.[19]

The computation process of the MFCCs[11,20] includes a pre-emphasis filter applied before the 512 point DFT. After that, a Mel filter-bank, with 40 filters, is applied to the amplitude of the DFT. Pre-emphasis of the signal, $\bar{x}$, consists of the application of a finite impulse response (FIR) filter defined by the following expression: $x'(i)=x(i)-ax(i-1)$, with $a=0.97$.[20] A Hamming window is applied before the DFT to reduce the discontinuity between frames in a segment and the magnitude of the side lobes in the DFT due to the windowing.[20] A 50% of overlap is employed when analyzing the signal. After the Mel filtering, the discrete cosine transform of the logarithm of the output is calculated. Note that there may be slight differences in the calculation of the MFCCs due to the differences in the pre-filtering or even in the specific Mel filter-bank used;[21] in our case, the integral of each filter, in the bank of Mel filters, is normalized to 1. In this work, the first five coefficients have been selected.[12] Each coefficient is considered a single feature, whose mean and standard deviation are extracted.

In Fig. 6, a scatter plot of the second coefficient is shown, where its discriminating capability is evident.

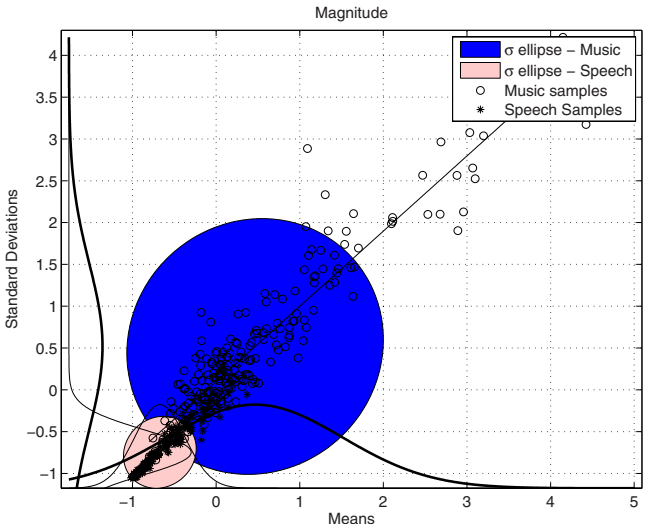


FIG. 5. (Color online) Scatter plot of speech and music magnitude means and standard deviations of a large 1 s length sample set. Linear fits, 1.5σ ellipses, and normal pdfs for means and standard deviations are shown too.

7. Volume dynamic ratio

The spread of rms over the frames in a segment, normalized with respect to its maximum value, depends on the type of sound source[22] and it can also discriminate efficiently music from speech. This feature is referred to as volume dynamic ratio (VDR) and in our implementation we use

$$\mathrm{VDR} = 1 - \frac{\min_j(\mathrm{rms}_j)}{\max_j(\mathrm{rms}_j)}. \tag{5}$$

In each segment, speech presents higher mean values and lower variance than music (Fig. 7).

B. Other features evaluated

In this section, the audio features evaluated that were not selected for the final discrimination subset are presented. The specific implementation used in our study is shown but no figures illustrate the behavior in this case.

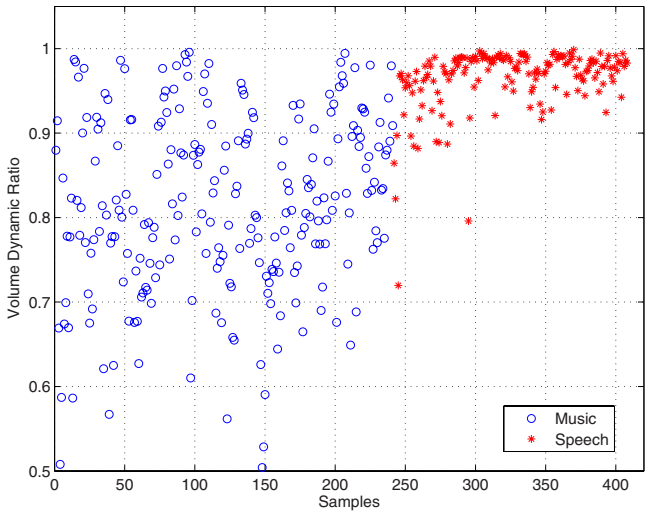


FIG. 7. (Color online) Scatter plot of speech and music volume dynamic ratio outcomes.

1. Silence ratio

One segment is considered "silence" if its amplitude is lower than a fixed threshold.[14] In this work, the silence ratio is calculated by tagging the frames with rms_j lower than 10% of the rms of the entire segment:

$$\mathrm{SIR} = \frac{\mathrm{card}\{j{:}\mathrm{rms}_j \leq 0.1 \cdot \mathrm{rms}\}}{50}, \tag{6}$$

where the function card{ } returns the number of elements that accomplish the specified condition.

2. Spectral centroid

This is the center of gravity of the spectrum itself,[12] and it is estimated in each frame as follows:[23]

$$\mathrm{SPC}_j = \frac{\sum_{k=1}^{m} F_j(k) \cdot M_j(k)}{\sum_{k=1}^{m} M_j(k)}, \tag{7}$$

where $F_j(k)$ and $M_j(k)$ represent the kth frequency and DFT magnitude samples of the jth frame, respectively. The DFT was calculated with length 1024.

3. Spectral rolloff point

It is the frequency value up to which 95% of signal energy resides.[16] We estimate this value for each frame using

$$\mathrm{SRP}_j = F_j(t){:}\left[\sum_{k=1}^{t<m} M_j(k) = 0.95 \cdot M_j\right], \tag{8}$$

with $M_j = \Sigma_{k=1}^{m} M_j(k)$, where $M_j(k)$ is the magnitude of kth sample of the DFT of the jth frame. $F_j(t)$, with $t < m$, is the frequency that corresponds to the 95th percentile of the energy of the DFT.

4. Bandwidth (spectral centroid range)

The bandwidth (BDW) represents here a frequency range around the spectral centroid, where the signal energy is concentrated,[14] it is estimated as follows:

$$\mathrm{BDW}_j = \frac{1}{m}\sum_{k=1}^{m} M_j(k) \cdot (F_j(k) - \mathrm{SPC}_j), \tag{9}$$

where SPC_j is the spectral centroid of the jth frame.

5. Frame energy and segment energy

The frame energy represents the energy of each frame and, correspondingly, the segment energy. We use the following measures obtained in the frequency domain.

Frame energy:

$$\mathrm{FRE}_j = \sum_{k=2}^{m} (M_j(k))^2, \tag{10}$$

where $(M_j(k))^2$ represents the squared magnitude of the DFT (with length m) of each frame.

Segment energy:

$$\mathrm{SGE} = \sum_{k=2}^{m} (M(k))^2, \tag{11}$$

where $(M(k))^2$ represents the squared magnitude of the DFT (with length m) of each segment. In both equations, m is the smallest integer such that $m = 2^k \geq N$, with N the number of samples of frame and segment, respectively. Note that, now, the magnitude of the DFT sample at frequency 0 (offset) is not included in the measures. Segment energy represents a scalar magnitude and it is integrated in the vector of descriptors as it is, while the mean and variance are computed for the frame energy.

6. Fundamental frequency

The fundamental frequency, in our context, is related to pitch, although these can be considered different views of the frequency perception by humans and computers.[24] A way to robustly estimate the fundamental frequency is based on the autocorrelation function (ACF);[25] in this work, we proceed as follows. In each frame,

- compute the ACF $y_j(i) = \Sigma_{k=0}^{n-1} x_j(k) \cdot x_j(k+i)$,
- find the largest local maxima of $y_j(i)$, with $i > 0$, and
- obtain the frequency corresponding to the index selected:

$$F_j = F_j(k) \leftrightarrow \max_{i, i>0}(y_j(i)). \tag{12}$$

Note that although the method is not completely robust against errors in the identification of the fundamental frequency (other harmonics can be taken as fundamental frequency using this technique), the method is valid for us since we do not really need the fundamental frequency but a dominant frequency.

7. Salience of pitch

This feature quantifies the prominence of the fundamental frequency. We use the following expression:

$$\mathrm{SOP}_j = \frac{\underset{i\geq 1}{\text{local max}}\left(\sum_{k=0}^{n-1} x_j(k)\cdot x_j(k+i)\right)}{\sum_{i=1}^{n} |x_j(i)|^2}, \tag{13}$$

where the function local $\max_{i\geq 1}(y(i))$ returns the largest $y(j)$, with $j\geq 1$, such that $y(j)\geq y(j-1)$ and $y(j)\geq y(j+1)$.

III. CLASSIFICATION OF THE VECTORS OF FEATURES

Although the discrimination between music and speech sometimes is done using heuristic tests with a few features at hand[9] or by direct thresholding,[6] a more common practice is the usage of several features as inputs to a more elaborate classifier. Carey[4] used a Gaussian mixture model (GMM) to model the classes of audio signal, and the difference in the log-likelihood is used for classification. Saunders used a multivariate Gaussian model[3] (the tests performed are not specified in this paper). A set of fuzzy rules is used in Ref. 2, a neural network is used in Ref. 7 using as inputs the first order statistics of the spectra of the audio segments employed, and the k-nearest neighbor classifier is also considered. Goodwin and Laroche[8] turned their attention to the identification of cost functions for a dynamic programming procedure that uses linear discriminant analysis for dimensionality reduction. In our case, we initially deal with a large number of features, as shown, and we expect to use them in an efficient way in a multi-feature classifier. We choose Fisher linear discriminator to deal with multiple features as a technique to reduce the problem of dimensionality[26] and a Gaussian model for the projections of the vectors of features;[27] then, the complementary error function is used to measure the probability that a certain sample is at least that distance away from the mode of the distribution, and the comparison of the probabilities for the different classes is used for the classification.

Now, we turn our attention to Fisher linear discriminant function.[26] Consider an n-dimensional vector of observations (audio features in our context), Fisher linear discriminant function is aimed to maximize the clustering of the two classes,[28] maximizing the separation between them, i.e., maximizing the inter-class variance and minimizing the intra-class variance.[26] From a geometrical point of view, this is accomplished by means of a projection in a favorable direction $\bar{v}$. To this end, the function

$$J(\bar{v}) = \frac{(\mu_m - \mu_s)^2}{S_m^2 + S_s^2} \tag{14}$$

must be maximized.[26] In our specific context,

- $\bar{v}$ is the projecting vector,
- $\mu_{m,s} = 1/n_{m,s}\sum_{i=1}^{n_{m,s}} p_i$ are the means of projected points p_i for music (m) and speech (s) samples, and
- $\tilde{S}_{m,s}^2 = \sum_{i=1}^{n_{m,s}}(p_i-\mu_i)^2$ are the so called scatters of projected points p_i for music and speech samples.

In order to solve Eq. (14), let

- $p=\bar{v}^t\bar{d}$ denote the projected form p of the vector of descriptors $\bar{d}$;
- $S_W = \sum_{i=1}^{n_m}(p_i-\mu_i)+\sum_{i=1}^{n_s}(p_i-\mu_i) = S_m + S_s$ is the sum of the music and speech scatter matrices, also known as *intra-class scatter matrix*; and
- $S_B = (\mu_m-\mu_s)\cdot(\mu_m-\mu_s)^T$ represents the *inter-class scatter matrix*.

Then, Eq. (14) can be rewritten as

$$J(\bar{v}) = \frac{v_t S_B v}{v_t S_W v}. \tag{15}$$

Then, the problem can be solved by converting it into a well known generalized eigenvalue problem whose solution is given by[27]

$$v = S_W^{-1}(\mu_m - \mu_s). \tag{16}$$

Ideally, two separated clouds of projected feature vectors should be found. However, the features themselves behave as random-variables (rvs) with unspecified probability density function (pdf) and, so, the projection of the vector of rvs, as a sum of rvs, can be taken to be a normal rv according to the central limit theorem.[27,29] Then, the class membership of each projection is decided by comparing the probability that such projection is at least that distance away from the mode of each class distribution. Thus, a sample p will be assigned to the class C_x with $x=m$ or s for the music class and the speech class, respectively, using

$$C_x = C_c : c \left| \max_{c=\{s,m\}} \left(\mathrm{erfc}\left\| \frac{p-\mu_c}{\sigma_c} \right\| \right) \right. . \tag{17}$$

Note that this is a monotonic function,[30] so, from a computational point of view, we decide to use the argument for the classification. With this strategy, the computational load of the classifier is mainly due to the process of extraction of the selected features and, also, the number of parameters that the classifier needs to store is small.

Now, a number of features, which should attain the best possible classification behavior, with respect to the selected classifier should be found. In Sec. IV, we analyze the behavior of the features described previously, in relation to the selected classifier, to make the final decision about the features that will be finally used.

IV. PERFORMANCE ANALYSIS AND FINAL CLASSIFIER DESIGN

In order to check the performance of the features and to select the features that should be finally used in the classification, the samples given by Fundación Albéniz were manually cut into fragments of a few seconds of length, such that each of them contained a single type of audio, either speech or music. Then, a large set of 1 s audio segments randomly selected was extracted for training and testing the classifier and the features. Initially, the classification performance of the isolated features was observed, then an optimal subset of features for the training data set and classification strategy selected was found.

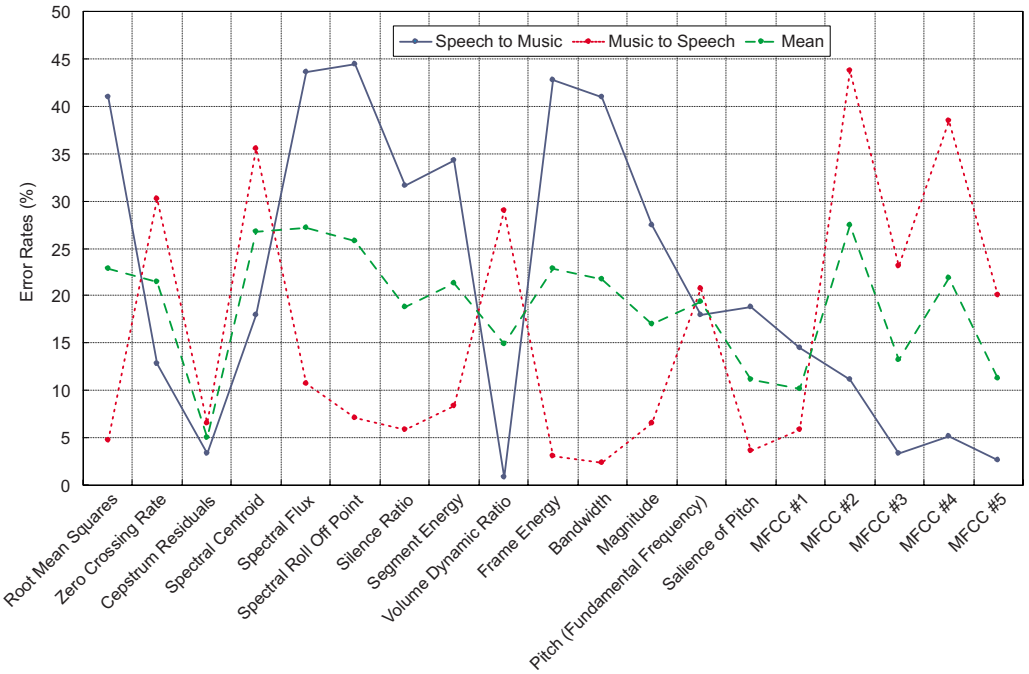


FIG. 8. (Color online) Classification error rates. Speech-to-music (solid), music-to-speech (dotted), and mean values (dashed).

A. Performance test of single features: Simple grouping

The classification performance of each feature is evaluated. In Fig. 8 the music-to-speech and speech-to-music error rates are shown for an implementation of the classifier based on a single feature. Training and test audio samples are different, and the complete set is about 1000 audio segments.

This figure shows that cepstrum residuals seem to be the best features for classification, achieving a mean error rate of 5% (assuming that speech and music are equally likely). Also, in the figure, it can be observed that the volume dynamic ratio shows the lowest speech-to-music error. With the data at hand, we turn to the main target of the work which is to find a subset of features that, combined in the selected classifier, attain the best classification performance with the classification strategy selected based on the usage of the Fisher linear discriminant function. This search is motivated by the fact that a proper combination of favorable features should attain better classification performance than single features. For example, consider the "time domain" features (root mean squares, zero-crossing rate, silence ratio, and volume dynamic ratio); this subset attains 5.1% and 12.4% for speech-to-music and music-to-speech error rates, respectively, when combined using the ideas of Fisher linear discriminant analysis (LDA). Those features attain a global error rate of 8.8% versus a 14.5% of averaged global error rate of that subset of features. The remaining features achieve a better classification error rate, see Table I. So, we will search for a subset of features that attain the best possible performance with the classifier selected.

B. Selection of features and definition of the classifier

In order to select a subset of features that attain the best performance with the selected classification strategy, a recursive search has been used. As 19 features are initially available, one could try to find the parameters of the classifier and evaluate its performance for all the possible combinations of features. Unfortunately, the number of subsets to evaluate[31] makes the full search computationally intractable. With the computational power at hand, we find that we can carefully evaluate subsets of up to 12 features, so, we test the performance of all the subsets of 12 features or less; with this, more than 1.8×10^9 combinations have been evaluated. Figure 9 shows the evolution of the error rates of the best subset for a maximum number of features in the subset of $k=1,2,3,\ldots,12$.

Observe that Fig. 9 indicates that the performance of the best subset for $k>8$ is almost the same as for $k=8$, discouraging the implementation of a costly recursive procedure for $k\gg8$. The subset that requires the lowest computation time with the lowest number of features, which achieve the lowest error rate with the selected classification procedure, is composed by the features shown in Table II. This subset of features, used in the Fisher classifier, with the Gaussian approximation of the projections of the classes, reaches a mean error rate of 0.3% (see Table II).

After finding this subset, we continue analyzing the performance of the classifier over the subsets evaluated, paying

TABLE I. Error rates for time and frequency domain subsets. Comparison among combined and averaged rates is shown.

Subset	Speech-to-music (%)	Music-to-speech (%)	Average (%)
Time domain	5.1	12.4	8.8 Combined
	21.6	17.5	19.5 Single (mean)
Frequency domain	6.4	3.0	4.7 Combined
	21.9	15.7	18.8 Single (mean)

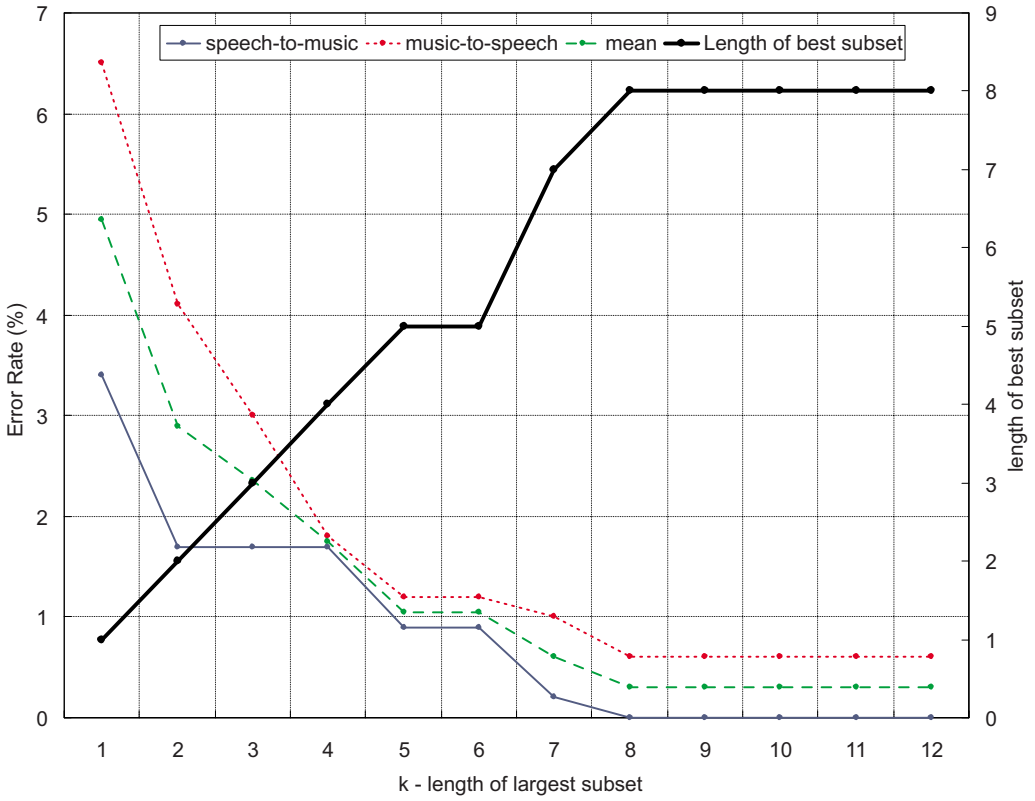


FIG. 9. (Color online) Results of recursive classification using k varying lengths of subset. Speech-to-music, music-to-speech, and mean error rates as shown with bold, dotted, and dashed lines, respectively. The thick line shows the number of features of the best subset.

attention to the subsets that attain the lowest error rates. Monitoring the best hundred subsets, we discover that the first five combinations attain the same mean error rate, about 0.3%.

In Fig. 10, the appearance rate of all the features in this subset is shown. Note that no more than ten features are involved in these subsets, which constitute a computationally tractable subset, so, we select all these features for the final classification tool. This subset attains a mean error rate of 0.3% over the test data set. Table III shows the best five subsets of features found by our iterative search; the features are represented by the acronyms defined in the section devoted to the description of each feature. In this table, subset 5 contains the features finally selected for the final discrimination application.

TABLE II. Error rates for each feature contributing to the optimal subset.

	Error rate		
Feature	Speech as music (%)	Music as speech (%)	Average (%)
Cepstrum residuals	3.4	6.5	5.0
Spectral flux	43.6	10.7	27.2
Magnitude	27.4	6.5	17.0
Mel frequency coefficient 2	11.1	43.8	27.5
Mel frequency coefficient 3	3.4	23.1	13.3
Mel frequency coefficient 4	5.1	38.5	21.8
Mel frequency coefficient 5	2.6	20.1	11.4
Volume dynamic ratio	0.9	29.0	15.0
Grouped feature set	0.0	0.6	0.3

V. CONCLUSIONS AND DISCUSSION

In this paper, we have dealt with the problem of music-speech discrimination. We have evaluated the performance of a number of features, most of them taken from literature, using specific parameters driven from the requirements and

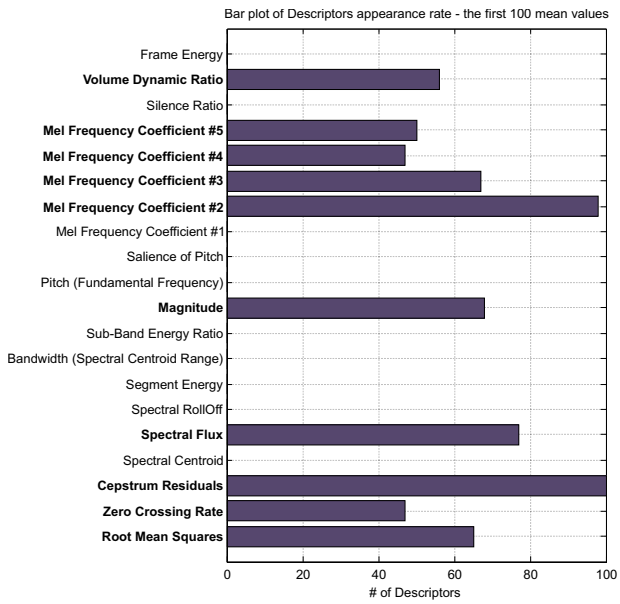


FIG. 10. (Color online) Feature appearance on the best subsets for the Fisher linear classifier.

TABLE III. Best subsets of features found for the classification.

Subset 1	Subset 2	Subset 3	Subset 4	Subset 5
CER	rms	rms	ZCR	rms
SPF	ZCR	ZCR	CER	ZCR
MAG	CER	CER	SPF	CER
MFCC 2	SPF	SPF	MAG	SPF
MFCC 3	MAG	MFCC 2	MFCC 2	MAG
MFCC 4	MFCC 2	MFCC 3	MFCC 3	MFCC 2
MFCC 5	MFCC 3	MFCC 4	MFCC 4	MFCC 3
VDR	MFCC 4	MFCC 5	MFCC 5	MFCC 4
⋯	MFCC 5	VDR	VDR	MFCC 5
⋯	⋯	⋯	⋯	VDR

the conditions of the data available, and, then, we have chosen a classification strategy based on Fisher linear discriminant analysis to define a classifier which is computationally tractable, in which the computational load is due to the process of extraction of features. Also, the classifier selected does not require the storage of large sets of data or features to perform the classification, but only a small number of parameters.

With the selected classification strategy, an iterative search has been conducted to find a subset of features that attain the best possible classification performance with a tractable computational load. 10 out of 19 features were selected on the basis of the analysis of the behavior of the features and their combinations in the selected classification strategy; thus, a fast classifier was developed that attained, on the training and evaluation data sets, mean error rates well below 1%.

ACKNOWLEDGMENTS

This work was funded by the Spanish Ministerio de Industria, Turismo y Comercio under Project No. FIT-350201-2007-8 and by the Spanish Ministerio de Educación y Ciencia under Project No. TSI-2007-61181. The whole audio dataset used in this work was kindly provided by Fundación Albéniz.

[1]B. P. Bogert, M. J. R. Healy, and J. W. Tukey, *The Frequency Analysis of Time Series for Echoes: Cepstrum, Pseudo-Autocovariance, Cross-Cepstrum, and Saphe Cracking* (Wiley, New York, 1963).
[2]R. M. Aarts and R. T. Dekkers, "A real-time speech-music discriminator," J. Audio Eng. Soc. **47**, 720–725 (1990).
[3]J. Saunders, "Real-time discrimination of broadcast speech/music," in Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP '96 (1996), Vol. **2**, pp. 993–996.
[4]M. J. Carey, E. S. Parris, and H. Lloyd-Thomas, "A comparison of features for speech, music discrimination," in Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP '99 (1999), Vol. **1**, pp. 149–152.
[5]E. Pampalk, "Computational models of music similarity and their application in music information retrieval," Ph.D. thesis, Vienna University of Technology, Vienna, Austria (2006).
[6]R. Jarina, N. Murphy, N. O'Connor, and S. Marlow, "Speech-music discrimination from MPEG-1 bitstream," in *Advances in Signal Processing, Robotics and Communications*, SSIP'01—WSES International Conference on Speech, Signal and Image Processing, edited by V. V. Kluev and N. E. Mastorakis (WSES, Athens, Greece, 2001), pp. 174–178.
[7]H. Harb and L. Chen, "Robust speech music discrimination using spectrum's first order statistics and neural networks," in Proceedings of the 7th International Symposium on Signal Processing and Its Applications (2003), Vol. **2**, pp. 125–128.
[8]M. M. Goodwin and J. Laroche, "A dynamic programming approach to audio segmentation and speech/music discrimination," in Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP '04 (2004), Vol. **4**, pp. iv-309–iv-312.
[9]C. Panagiotakis and G. Tziritas, "A speech/music discriminator based on RMS and zero-crossings,"IEEE Trans. Multimedia **7**, 155–166 (2005).
[10]S. B. Davis and P. Mermelstein, "Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences," IEEE Trans. Acoust., Speech, Signal Process. **28**, 357–366 (1980).
[11]C. Lee, C. Chou, C. Han, and R. Huang, "Automatic recognition of animal vocalizations using averaged MFCC and linear discriminant analysis," Pattern Recogn. Lett. **27**, 93–101 (2006).
[12]G. Tzanetakis and P. Cook, "Musical genre classification of audio signals," IEEE Trans. Speech Audio Process. **10**, 293–302 (2002).
[13]C. Shahnaz, W.-P. Zhu, and M. O. Ahmad, "A multifeature voiced/unvoiced decision algorithm for noisy speech," in Proceedings of the 2006 IEEE International Symposium on Circuits and Systems, ISCAS 2006 (2006), pp. 2525–2528.
[14]M. Liu and C. Wan, "A study on content-based classification and retrieval of audio database," in 2001 International Database Engineering and Applications Symposium (IDEAS '01) (2001), pp. 339–345.
[15]B. Kedem, "Spectral analysis and discrimination by zero-crossings," Proc. IEEE **74**, 1477–1493 (1986).
[16]P. Arnaud, "Speech/music discriminator—Project report," Tampere University of Technology, Finland, http://www.cs.tut.fi/sgn/arg/arno (Last viewed 4/21/2009).
[17]E. Scheirer and M. Slaney, "Construction and evaluation of a robust multifeature speech/music discriminator," in Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP '97 (1997), Vol. **2**, pp. 1331–1334.
[18]S. S. Stevens, J. Volkman, and E. B. Newman, "A scale for the measurement of the psychological magnitude pitch," J. Acoust. Soc. Am. **8**, 185–190 (1937).
[19]H. Zhou, A. Sadka, and R. M. Jiang, "Feature extraction for speech and music discrimination," in International Workshop on Content-Based Multimedia Indexing, CBMI 2008 (2008), pp. 170–173.
[20]W. Han, C. Chan, C. Choy, and K. Pun, "An efficient MFCC extraction method in speech recognition," in Proceedings of the IEEE International Symposium on Circuits and Systems (2006) pp. 145–148.
[21]S. Sigurdsson, K. B. Petersen, and T. Lehn-Schiøler, "Mel frequency cepstral coefficients: An evaluation of robustness of MP3 encoded music," in Proceedings of the 7th International Conference on Music Information Retrieval (ISMIR) (2006).
[22]M. Liu, C. Wan, and L. Wang, "Content-based audio classification and retrieval using fuzzy logic system: Towards multimedia search engines," Soft Comput. **6**, 357–364 (2002).
[23]D. Hosseinzadeh and S. Krishnan, "Combining vocal source and MFCC features for enhanced speaker recognition performance using GMMs," in IEEE 9th Workshop on Multimedia Signal Processing, MMSP 2007 (2007), pp. 365–368.
[24]S. Suk, H. Chung, and H. Kojima, "Voice/non-voice classification using reliable fundamental frequency estimator for voice activated powered wheelchair control," in Lecture Notes in Computer Science, International Conference on Embedded Software and Systems (2007), Vol. **4523**, pp. 347–357.
[25]A. de Cheveigné and H. Kawahra, "YIN, A fundamental frequency estimator for speech and music," J. Acoust. Soc. Am. **111**, 1917–1930 (2002).
[26]R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, 2nd ed. (Wiley-Interscience, New York, 2000).
[27]H. Stark and J. W. Woods, *Probability and Random Processes with Applications to Signal Processing*, 3rd ed. (Prentice-Hall, Englewood Cliffs, NJ, 2001).
[28]R. A. Fisher, "The statistical utilization of multiple measurements," Annals of Eugenics **8**, 376–386 (1938).
[29]A. Papoulis, *Probability, Random Variables and Stochastic Processes*, 4th ed., (McGraw-Hill, New York, 1991).
[30]M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables* (Dover, New York, 1972).
[31]L. J. Bain and M. Engelhardt, *Introduction to Probability and Mathematical Statistics*, 2nd ed. (Duxbury, Pacific Grove, CA, 2000).

Intelligibility and listener preference of telephone speech in the presence of babble noise

Joseph L. Hall[a] and James L. Flanagan
Avaya Labs Research, 233 Mt. Airy Road, Basking Ridge, New Jersey 07920

(Received 7 May 2009; revised 8 October 2009; accepted 9 October 2009)

Two procedures for improving the intelligibility of wideband telephone speech in the presence of competing babble noise are evaluated. One procedure is differentiation, the other consists of equalizing the speech spectrum by applying the inverse of the average spectrum of formant amplitudes for adult male speakers ("formant equalization"). Speech processed by these two methods was evaluated both for intelligibility and for listener preference. Both methods produced a clear increase in intelligibility compared to unprocessed wideband telephone speech. Formant equalization was found to be preferred over differentiation, more so at low signal-to-noise ratios than at high ones. © *2010 Acoustical Society of America.* [DOI: 10.1121/1.3263603]

PACS number(s): 43.60.Dh, 43.66.Dc, 43.66.Lj, 43.71.Gv [WMC] Pages: 280–285

I. INTRODUCTION

A typical environment in teleconferencing has multiple listeners attending to speakerphone output. Often, side conversations develop, resulting in masking of the desired signal by "babble" (multiple simultaneous speech sources) noise. This report aims to evaluate two simple techniques for improving the intelligibility and quality of speech received via telephone in the presence of babble noise generated at the receiving end. In two separate experiments, we measured intelligibility of and listener preference for unprocessed speech, for differentiated speech, and for speech equalized by the inverse of the average formant amplitude spectrum.

Although this study is concerned with speech presented over speakerphones, both speech and noise were presented diotically through headphones. This procedure provides greater control over signal levels and spectra than does signal presentation through loudspeakers. It follows that these results are equally applicable to single and to multiple loudspeaker situations.

The bandwidth of speech over the switched telephone network historically has been limited to approximately 300 Hz–3.3 kHz (TIA/EIA-810-A, 2000). This bandwidth represented a compromise between the conflicting criteria of intelligibility, speech naturalness, and cost. With the introduction of Voice-over-Internet-Protocol service (VoIP), the limitations of the switched network no longer apply, and *wideband* telephony provides a frequency range of approximately 100 Hz–7 kHz, as shown in the send and receive templates for wideband telephone in Fig. 1 (TIA/EIA-920, 2002).

The frequency response for conventional wideband telephony is essentially flat, as shown in Fig. 1. This frequency response may be desirable for some applications, but for speakerphone use by a conference group in a noisy room it is *not* the most intelligible. It is well established (French and Steinberg, 1947; Miller and Nicely, 1955) that while much of the energy of speech is in low-frequency vowels, much of the intelligibility is in high-frequency consonants. Methods of improving speech intelligibility by enhancing high-frequency content have been investigated in the past (e.g., Thomas and Ohley, 1972; Griffiths, 1968). More recently, there have been studies of more elaborate methods for improving speech intelligibility using sophisticated speech processing (e.g., Yoo *et al.*, 2007; Kim *et al.*, 2009).

We describe in this report the effect of two spectral modifications on the perception of speech in the presence of babble noise: (1) differentiation and (2) formant equalization, corresponding to the inverse of formant amplitudes measured for adult male talkers (Peterson and Barney, 1952). Our rationale for choosing these two modifications for study is as follows: Differentiation is perhaps the simplest way of boosting high frequencies. It can be realized in a sampled-data system with a first-order backward difference. Formant equalization is slightly more complex, but it is attractive because it is more specifically tailored toward speech than is differentiation (Cassel, 1964; Flanagan *et al.*, 1964). The effect of these spectral modifications on speech quality as well as on speech intelligibility is investigated.

These spectral modifications were compared to unmodified speech in two separate experiments. Experiment I used the diagnostic rhyme testing (DRT) procedure described in Sec. 8.4 of ANSI S3.2 (ANSI S3.2, 1999; Voiers, 1971) to measure the intelligibility of isolated single words. Experiment II used the comparison category rating method described in Annex B of ITU-T Recommendation P.800 (ITU-T Recommendation P.800, 1996) to measure listener preference for the various spectral modifications.

II. PROCEDURES

A. Signal processing

All signal processing was done and all signals were presented at a sampling rate of 22.05 kHz. Figure 1 shows the spectral modification introduced by the two procedures used in this study, differentiation and formant equalization. Differentiation (solid line) was realized as a first-order backward

[a]Author to whom correspondence should be addressed. Electronic mail: jlhall@research.avayalabs.com

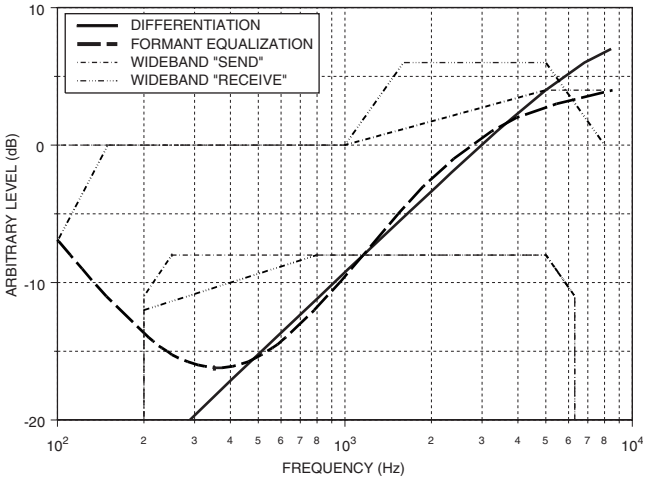


FIG. 1. Spectral modification introduced by differentiation (solid line) and by formant equalization (dashed line). z-transforms used for differentiation and for formant equalization are given in the Appendix. The figure also shows the handset "send" (dash-dot line) and "receive" (dash-triple dot line) frequency response masks for wideband telephone (TIA/EIA-920, 2002).

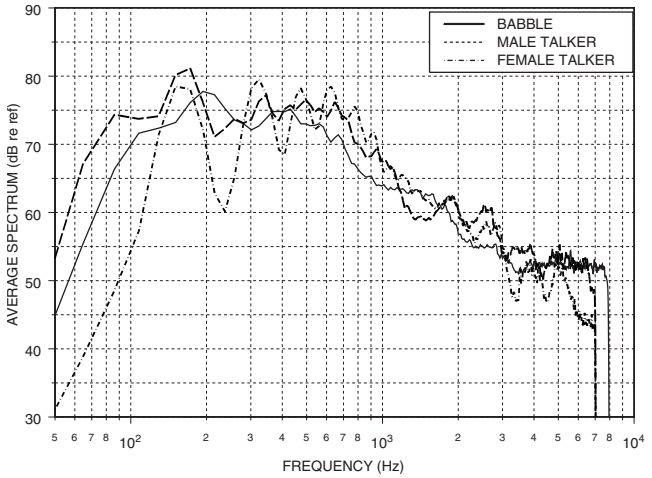


FIG. 2. Average spectrum of speech babble used as masker (solid line). The babble shows a high-frequency noise floor about 25 dB below the peak at 2 kHz. Average spectra for unprocessed speech from the two talkers used in experiment 1 (dashed lines). Note that the spectrum for babble, representing masking generated locally, is band limited to 8 kHz, while spectra for the two talkers, representing speech received over a wideband telephone system, are band limited to 7 kHz.

difference, $y_k = x_k - 0.995x_{k-1}$. This processing produces a 6 dB/octave boost over the broadband frequency range of interest, decreasing slightly as one-half the Nyquist frequency of 11.025 kHz is approached. Formant equalization (dashed line) approximates the inverse of average formant amplitudes as measured for adult male talkers (Peterson and Barney, 1952), as described in Sec. I. The procedure is described in Flanagan *et al.* (1964). Specific z-transforms used for differentiation and for formant equalization are given in the Appendix. The vertical displacement of these two curves in Fig. 1 is arbitrary: All speech signals were set to the desired levels by applying ITU-T Recommendation P.56 (ITU-T Recommendation P.56, 1993) after the spectral modification.

B. Factors common to experiment I and experiment II

Listeners sat inside a sound-treated listening booth (WhisperRoom Inc. MDL6060E 60×60 in. enhanced isolation room). Speech samples were presented diotically (binaural, same signal to left and right ears) through Sennheiser HD250 linear II headphones. The experiments were controlled by a Tucker-Davis Technologies System II workstation with a high-quality 16-bit digital/analog (D/A) converter (Tucker-Davis DA3–4 four-channel 16-bit instrumentation D/A).

The babble used as a masker in both experiments was made up of 72 10-s samples of speech, approximately evenly divided between 20 male and 20 female talkers. The speech samples were from a corpus recorded with a high-quality microphone in an anechoic room at 48 kHz sampling rate and subsequently stored at 16 kHz (Schroeter, AT&T Bell Laboratories internal memo, 1992). We first upsampled the speech samples from 16 to 22.05 kHz and then scaled levels of the samples to simulate uniformly spaced talkers in a region around the listener (16 samples at 0 dB re ref., 24 samples at −3.5 dB re ref., and 32 samples at −6 dB re ref.). The speech samples were windowed and offset to obtain a 10-s signal that could be looped endlessly with no transients.

Figure 2 shows the average spectrum of the babble masker. This spectrum was never modified. Four test subjects took part in this study: two young men and two older men, all of whom produced normal, age-corrected Békésy audiograms, administered by the same laboratory system used for the experiments. Statistical analysis and plotting were done with program S-Plus for Windows 6.0.

III. EXPERIMENT I: INTELLIGIBILITY

A. Procedures

As described in Sec. I, experiment I used a DRT procedure (ANSI S3.2, 1999) to measure the intelligibility of isolated single words. The test materials were subsets from a government-sponsored corpus produced by the Arcon Corporation, which included utterances for DRTs under several environmental conditions (Quatieri *et al.*, 2006; Dusan *et al.*, 2007). DRT materials consist of pairs of words differing only in the initial consonant, such as "bond" and "pond." One word of a pair was presented over headphones in each trial, then after a 500-ms interval the two candidate words were presented through a visual display and the subject was required to indicate which of the two words had been presented. The next trial began 500 ms after the subject responded.

The test materials, initially at a sampling rate of 48 kHz, were downsampled to 22.05 kHz, bandpass filtered from 90 Hz to 7 kHz in order to remove unwanted low-frequency transients, processed using the two methods described above, and finally set to the desired listening level through application of ITU-T Recommendation P.56 (ITU-T Recommendation P.56, 1993).

We used utterances from two talkers, one male and one female. Figure 2 shows average spectra for unprocessed speech from these two talkers. For processed speech, these spectra were modified, as shown by the solid and dashed curves in Fig. 1. The test corpus contained 232 utterances

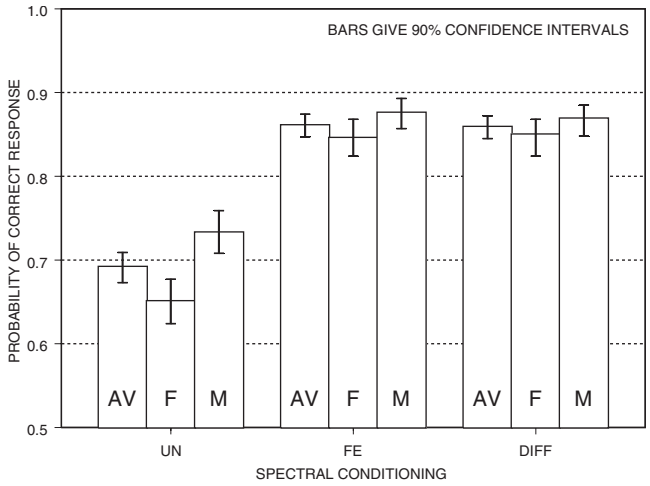


FIG. 3. Intelligibility (probability of a correct response) versus spectral conditioning for the average of the two talkers (AV), for the female talker (F), and for the male talker (M). UN: unprocessed. FE: formant equalization. DIFF: differentiation. Results averaged over four experimental subjects. Vertical bars give 90% confidence intervals. Babble masker at 70 dB SPL and speech at −6 dB SNR.

from 116 word pairs. Each of the 464 (2×232) utterances was presented either unprocessed, differentiated, or formant equalized, resulting in a total of 1392 (3×464) judgments. Each of the four test subjects made all 1392 judgments, administered in eight sessions of 174 stimuli per session spread out over several days.

The interfering babble was delivered continuously at a 70 dB sound pressure level (SPL) (dB re 0.0002 $\mathrm{dyn/cm^2}$), representative of high conference room background, and the test words were delivered at a signal-to-noise ratio (SNR) of −6 dB, a value that we had determined in preliminary experiments to provide the most sensitive discrimination conditions. The subject held a QSI Corporation QTERM-II data-entry terminal that provided both a small text window for visual display of the candidate words and pushbuttons for the subject's response.

B. Results

Results averaged over the four experimental subjects are shown in Fig. 3. This figure shows probability of a correct response (that is, correct selection of the target word, or word intelligibility) for unprocessed (UN), formant-equalized (FE), and differentiated (DIFF) speech. Results are shown averaged over the two talkers (AV), for the female talker (F), and for the male talker (M). Vertical bars give 90% confidence intervals.

C. Discussion of results

Several relationships are suggested in Fig. 3. First, the intelligibility for unprocessed speech is lower than that for speech processed by either formant equalization or differentiation. Average intelligibility for unprocessed speech is approximately 70%, but this number increases to approximately 85% for either processing method. Second, unprocessed speech is significantly more intelligible for the male talker than for the female talker. For both formant equalization and differentiation, processed speech is slightly more intelligible for the male talker than for the female talker, but the difference is not statistically significant.

With one male and one female talker in our stimulus set, we are unable to tell whether or not this difference is attributable to gender of the talker. The advantage of the male talker might be attributed to the fact that the male vocal tract is physically longer than the female and the male vocal-cord pitch is usually about an octave lower, providing more resolution of the spectral envelope. Another consideration is that the formant equalization we used is derived from Peterson and Barney (1952) data for male talkers. It could be expected that this might make our formant equalization more effective in increasing intelligibility for the male than for the female talker. In any event, our results show only a very small difference.

IV. EXPERIMENT II: LISTENER PREFERENCE

A. Procedures

As described in Sec. I, experiment II used the comparison category rating method described in Annex B of ITU-T Recommendation P.800 (1996) to measure listener preference for versions of test sentences processed in different ways. Two versions of a pair of test sentences were presented in a given trial. The subject was instructed to listen carefully to the two versions and then to rate the quality of the second pair of sentences compared to the quality of the first pair of sentences on the following seven-point adjectival scale: "first pair much better," "first pair better," "first pair slightly better," "about the same," "second pair slightly better," "second pair better," and "second pair much better." In analyzing the experimental results, the seven-point adjectival scale was mapped to a seven-point numerical scale going from −3 (first pair much better) through 0 (about the same) to +3 (second pair much better). The sign of the result was adjusted appropriately to account for the order of presentation: the two versions being compared were presented in both possible orders, in separate trials. The resulting score is known as the comparison mean opinion score (CMOS).

Written instructions read by the listeners prior to the experiment included the paragraph "What you will hear is a pair of sentences, and then after a short interval another pair of sentences. The two pairs are identical except that they may have been processed differently. Your task is to decide which pair sounds better and by how much. You should base your judgment both on how pleasant-sounding the speech is and on intelligibility, how easy the speech is to understand."

Test sentences used in experiment II came from the same speech corpus used for the masking babble. The test sentences were upsampled from a sampling rate of 16 to 22.05 kHz, processed by formant equalization or differentiation, and set to the desired listening level through application of ITU-T Recommendation P.56 (ITU-T Recommendation P.56, 1993).

There were four talkers: high- and low-pitch female and high- and low-pitch male. For each talker, we used two sentences rich in fricatives and two sentences rich in glides and nasals. The pair of sentences in a trial were the two sentences

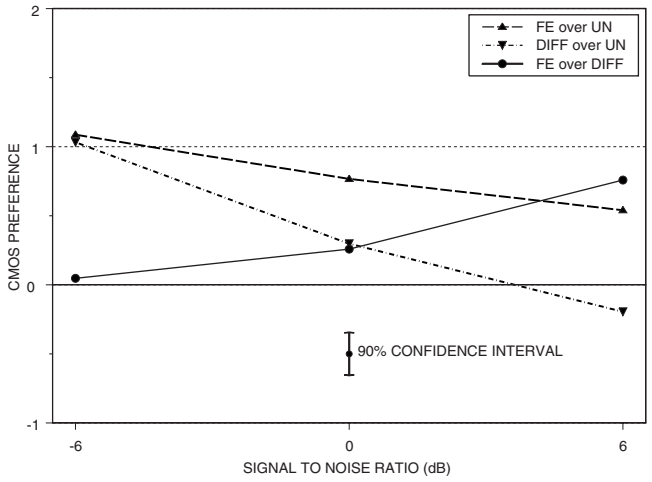


FIG. 4. Listener preference for speech processed by formant equalization over unprocessed speech (dashed line), speech processed by differentiation over unprocessed speech (dash-dot line), and speech processed by formant equalization over speech processed by differentiation (solid line). Babble masker at 70 dB SPL.

rich in fricatives or the two sentences rich in glides and nasals. As in experiment I, the interfering babble was delivered continuously at 70 dB SPL. Comparisons were made at three SNRs, −6, 0, and +6 dB, and there were two replications of the entire stimulus set. Each of the four test subjects made a total of 288 judgments, administered in six sessions of 48 judgments per session spread out over several days.

The subject responded by means of a response box with eight pushbuttons arranged in a row. The leftmost button, colored green, initiated the listening session and also gave the listener the opportunity to repeat a trial, if desired. The remaining seven buttons, colored white, were labeled with the seven response categories. After the listener responded by pushing one of the seven white buttons, the next trial proceeded automatically.

B. Results

The results from experiment II are shown in Figs. 4 and 5. In both of these figures, the three curves show preference for formant equalized over unprocessed ("FE over UN," dashed line), preference for differentiated over unprocessed ("DIFF over UN," dash-dot line), and preference for formant equalized over differentiated ("FE over DIFF," solid line). Figure 4 shows results averaged over all listeners and speech types, and Fig. 5 shows results for speech rich in glides and nasals in the upper panel and speech rich in fricatives in the lower panel.

Figure 4 shows that processed speech, either formant equalized or differentiated, is generally preferred over unprocessed speech. However, the amount of preference depends both on SNR and on type of processing. For speech at a SNR of −6 dB, formant-equalized and differentiated speech are each preferred over unprocessed speech by approximately one CMOS unit. For SNR=0 dB, the preference for formant equalized over unprocessed speech decreases to 0.8 CMOS units and for differentiated over unprocessed to 0.3 CMOS units. For SNR=+6 dB, formant-

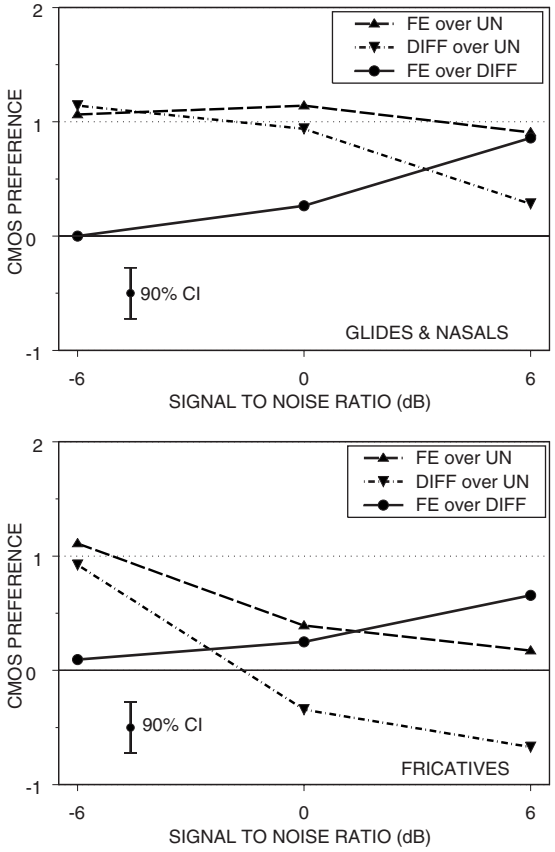


FIG. 5. Similar to Fig. 4 except that results are shown for speech rich in glides and nasals (upper panel) and for speech rich in fricatives (lower panel). Babble masker at 70 dB SPL.

equalized speech is preferred over unprocessed by 0.5 CMOS units, and unprocessed is preferred over differentiated by 0.2 CMOS units.

Consistent with the results described in the above paragraph, formant-equalized speech is generally preferred over differentiated speech, and the amount of this preference increases with increasing speech level. The two are judged "approximately the same" for SNR=−6 dB, but formant equalized is preferred over differentiated by 0.3 CMOS units at SNR=0 dB and by 0.8 CMOS units at SNR=+6 dB.

Figure 5 shows that speech type, rich in glides and nasals versus rich in fricatives, has a substantial effect. While for both speech types preference for formant-equalized over differentiated speech increases with increasing SNR, the preference for processed over unprocessed is greater for glides and nasals than for fricatives. At both SNR=0 dB and SNR=+6 dB, unprocessed is preferred over differentiated for speech rich in fricatives.

C. Discussion of results

A central conclusion to be drawn from these results is that while differentiation and formant equalization are equally effective in increasing speech intelligibility, formant equalization has an advantage over differentiation in terms of speech quality, more so at higher SNRs. This advantage is presumably related to the greater low-frequency (below 400 Hz) content of formant-equalized speech, as shown in Fig. 1.

Our interpretation of these results is that the judgment of speech quality is multidimensional, and the different dimensions are differently influenced by type of processing and SNR. Recall that listeners were instructed to base their judgments "… both on how pleasant-sounding the speech is and on intelligibility, how easy the speech is to understand." Formant equalization and differentiation both boost the higher frequencies, resulting in an unnatural and potentially unpleasant sound. At the lowest SNR, where intelligibility is the lowest, the unnaturalness introduced by processing is overbalanced by the resulting increase in intelligibility. As the SNR, and therefore intelligibility, increases, intelligibility becomes a less dominant factor. The influence of type of speech shown in Fig. 5 is accounted for by the fact that speech rich in fricatives has more high-frequency energy than speech rich in glides and nasals.

V. CONCLUSIONS

These results suggest that if speech intelligibility in the presence of babble interference is taken as the sole criterion, emphasizing the high-frequency end of the spectrum either by differentiation or by formant equalization can produce substantial improvement. This particular observation may be unique for babble masking, such as can occur in teleconferencing. The composite interfering spectrum is speech-like, being heavily weighted in the low frequencies which are effective maskers of higher-frequency components. However, we believe it can be exploited in other domains such as air-to-ground radio, where the interference is broad-spectrum static and continuous noise.

The picture changes when speech quality is considered. Formant equalization is preferred over differentiation, more so at low SNRs than at high ones. Speech quality is a multidimensional percept. Intelligibility is surely a major component, but attributes such as naturalness (as one experiences in face-to-face conversation) are important and contributory. For example, recognition of a familiar voice may be impacted by one-dimensional optimization of intelligibility.

This paper represents a first attempt to quantify interaction among the various perceptual attributes. This subject still constitutes a challenging area of research. Incomplete and preliminary attempts to quantify quality lead us to feel that such judgments are highly subject dependent, with some listeners giving preference to speech naturalness and others showing preference for optimum intelligibility.

ACKNOWLEDGMENTS

Man Mohan Sondhi provided valuable assistance in obtaining formulas for the coefficients of the formant-equalization z-transform shown in the Appendix. We are indebted to the listeners who agreed to take part in these listening experiments.

APPENDIX: z-TRANSFORMS USED FOR DIFFERENTIATION AND FOR FORMANT EQUALIZATION

1. Differentiation (backward difference)

$$y_k = x_k - \alpha \cdot x_{k-1}.$$

This leads to

$$F(z^{-1}) = 1 - \alpha \cdot z^{-1},$$

where z^{-1} is the unit delay operator.

(We used $\alpha = 0.995$ with a sampling frequency of 22.05 kHz.)

2. Formant equalization

From Flanagan *et al.* (1964), we have the Laplace transform of the processing as

$$F(s) = 1 - \frac{\alpha \cdot s}{s^2 + (\sigma_1 + \sigma_2 + \alpha) \cdot s + \sigma_1 \cdot \sigma_2},$$

where s is the complex frequency operator, $\sigma_1 = 2\pi \cdot 408\ \text{s}^{-1}$; $\sigma_2 = 2\pi \cdot 318\ \text{s}^{-1}$; and $\alpha = 2\pi \cdot 1590\ \text{s}^{-1}$.

This leads via the bilinear z-transform $s \rightarrow (2/T) \cdot (1 - z^{-1})/(1 + z^{-1})$ to

$$F(z^{-1}) = 1 + (1 + z^{-1}) \cdot \left[\frac{a_{01}}{b_{11} \cdot z^{-1} + 1} + \frac{a_{02}}{b_{12} \cdot z^{-1} + 1} \right],$$

where

$$a_{01} = \alpha \cdot p/(2/T - s_1), \quad a_{02} = \alpha \cdot q/(2/T - s_2),$$

$$b_{11} = (s_1 + 2/T)/(s_1 - 2/T), \quad b_{12} = (s_2 + 2/T)/(s_2 - 2/T)$$

(T is the sampling interval)

with

$$s_1 = \frac{-(\sigma_1 + \sigma_2 + \alpha) + \sqrt{(\sigma_1 + \sigma_2 + \alpha)^2 - 4 \cdot \sigma_1 \cdot \sigma_2}}{2},$$

$$s_2 = \frac{-(\sigma_1 + \sigma_2 + \alpha) - \sqrt{(\sigma_1 + \sigma_2 + \alpha)^2 - 4 \cdot \sigma_1 \cdot \sigma_2}}{2},$$

and

$$p = s_1/(s_2 - s_1), \quad q = s_2/(s_1 - s_2).$$

ANSI S3.2. (**1999**). "Method for measuring the intelligibility of speech over communication systems," American National Standards Institute, Washington, DC.

Cassel, L. E. (**1964**). "Spectral flattening and its application as a synthesis technique (A)," J. Acoust. Soc. Am. **36**, 1029–1030.

Dusan, S., Flanagan, J. L., Karve, A., and Balaraman, M. (**2007**). "Speech compression by polynomial approximation," IEEE Trans. Audio, Speech, Lang. Process. **15**, 387–395.

Flanagan, J. L., Meinhart, D. I. S., and Cummiskey, P. (**1964**). "Digital equalizer and deequalizer for speech (A)," J. Acoust. Soc. Am. **36**, 1030.

French, N. R., and Steinberg, J. C. (**1947**). "Factors governing the intelligibility of speech sounds," J. Acoust. Soc. Am. **19**, 90–119.

Griffiths, J. D. (**1968**). "Optimum linear filter for speech transmission," J. Acoust. Soc. Am. **43**, 81–86.

ITU-T Recommendation P.56. (**1993**). "Objective measurement of active speech level," International Telecommunication Union, Geneva.

ITU-T Recommendation P.800. (**1996**). "Methods for subjective determina-

tion of transmission quality," International Telecommunication Union, Geneva.
Kim, G., Lu, Y., Hu, Y., and Loizou, P. G. (**2009**). "An algorithm that improves speech intelligibility in noise for normal-hearing listeners," J. Acoust. Soc. Am. **126**, 1486–1494.
Miller, G. A., and Nicely, P. E. (**1955**). "An analysis of perceptual confusions among some English consonants," J. Acoust. Soc. Am. **27**, 338–352.
Peterson, G. E., and Barney, H. L. (**1952**). "Control methods used in a study of the vowels," J. Acoust. Soc. Am. **24**, 175–184.
Quatieri, T., Brady, K., Messing, D., Campbell, J. P., Campbell, W. M., Brandstein, M. S., Tardelli, J. D., and Gatewood, P. D. (**2006**). "Exploring nonacoustic sensors for speech encoding," IEEE Trans. Speech Audio Process. **14**, 533–544.
Thomas, I. B., and Ohley, W. J. (**1972**). "Intelligibility enhancement through spectral weighting," in Proceedings of the 1972 IEEE Conference on Speech Communications and Processing, pp. 360–363.
TIA/EIA-810-A. (**2000**). "Telephone terminal equipment—Transmission requirements for narrowband," U.S. Telecommunications Industry Association, Arlington, VA.
TIA/EIA-920. (**2002**). "Telephone terminal equipment—Transmission requirements for wideband digital wireline telephones," U.S. Telecommunications Industry Association, Arlington, VA.
Voiers, W. D. (**1971**). "Current status of the diagnostic rhyme test (A)," J. Acoust. Soc. Am. **50**, 145.
Yoo, S. D., Boston, J. R., El-Jaroundi, A., and Li, C. (**2007**). "Speech signal modification to increase intelligibility in noisy environments," J. Acoust. Soc. Am. **122**, 1138–1149.

A different approach to use narrowband super-resolution multiple signal classification algorithm on wideband sources

Mohammad Asgari[a)]
Department of Broadcast Engineering, IRIB University, Tehran 1779617165, Iran

Nasim Yahya Soltani
Faculty of Electrical and Computer Engineering, University of Tehran, Tehran 14395-515, Iran

Ali Riahi
Department of Broadcast Engineering, IRIB University, Tehran 1779617165, Iran

(Received 9 January 2009; revised 31 October 2009; accepted 5 November 2009)

There are varieties of wideband direction-of-arrival (DOA) estimation algorithms. Their structure comprises a number of narrowband ones, each performs in one frequency in a given bandwidth, and then different responses should be combined in a proper way to yield true DOAs. Hence, wideband algorithms are always complex and so non-real-time. This paper investigates a method to derive a flat response of narrowband multiple signal classification (MUSIC) [R. O. Schmidt, IEEE Trans. Antennas Propag., **34**, 276–280 (1986)] algorithm in the whole frequencies of given band. Therefore, required conditions of applying narrowband algorithm on wideband impinging signals will be given through a concrete analysis. It could be found out that array sensor locations are able to compensate the frequency variations to reach a flat response of DOAs in a specified wideband frequency. © *2010 Acoustical Society of America.* [DOI: 10.1121/1.3268609]

PACS number(s): 43.60.Fg [EJS] Pages: 286–291

I. INTRODUCTION

During the past few decades, there have been significant researches on source localization algorithms for sensor array signal processing, culminating in the development of subspace-based methods, which asymptotically exhibit infinite resolution, and are, hence, also referred to as "super-resolution" algorithms. An important issue that has been overlooked, however, is the array structure design, which, under practical nonasymptotic conditions, places fundamental limitations on the ultimate performance of the system. The aim of this paper is to investigate and propose a new array design framework tailored to subspace-based direction-finding (DF) algorithms. The conventional approach is to identify the determining factor in their DF performance. For instance, the design of beamformers is based on the structure of the array gain pattern since the width of the mainlobe underlies the resolving power, and the height of the sidelobes represents the power leakage from undesired sources.[1] Huang *et al.*[2] presented an approach to array design for the maximum likelihood estimator (MLE) based on minimizing a measure of the direction-of-arrival (DOA) estimation errors. Using a Taylor series expansion and retaining just the first-order terms, a linear approximation of the MLE, which is valid for large values of SNR [(signal power)/(acoustical noise power)], was obtained, from which an approximate expression for the covariance matrix of the ML estimates was derived. It turned out that the covariance matrix is a function of the true DOAs, the sensor locations, and the noise characteristics. By specifying the number of sensors, the available aperture, the true DOAs, and the noise covariance, it was possible to search for the sensor locations that minimized the determinant of the covariance matrix. It is important to note that all the above array design techniques involve an investigation of the sensor locations that satisfy certain algorithm-dependent criteria.[3] One common drawback, however, is the computational complexity of the search. In this paper, the design of arrays tailored to narrowband subspace-based DF algorithms is addressed, while incoming signals are wideband. We propose an innovative approach to the array design problem, which being not only the first super-resolution array design technique, but also overcomes the computationally intensive search procedure.

On the other hand, wideband signals are exploited for localization for a wide range of applications. A wideband signal is any signal whose energy is distributed over a bandwidth that is large in comparison with the signal's center frequency. For example, ultrawideband (UWB) noise radars use wideband offering low probability of detection (LPD), while achieving good target detection and high resolution.[4] In the acoustic vehicle tracking scenario, the target emits a set of narrowband harmonics.[5] Direct exploitation of an array of raw wideband signals for localization using traditional narrowband techniques is typically problematic. The shortcoming is that the narrowband methods exploit the fact that time delays directly translate to a phase shift in the frequency domain $s(t-t_0) \leftrightarrow S(f)\exp(-j2\pi f t_0)$. It is possible to pass the output of each array element through a narrowband bandpass filter, so it follows $S(f)$ for $f=f_i$. Then, a narrowband DOA

[a)]Author to whom correspondence should be addressed. Electronic mail: m.asgari@iribu.com

 0001-4966/2010/127(1)/286/6/$25.00

estimator is applied. However, this methodology ignores valuable data in other frequency bands that can enhance estimation. Researchers have proposed a number of methods to exploit the richer structure of the array output available with wideband sources.[6] For example, Su and Morf[7] modeled the wideband sources as a rational transfer function driven by white noise, and used modal decomposition to estimate the delays on each mode. Agrawal and Prasad[6] used an array manifold vector for wideband sources instead of the conventional narrowband array manifold, assuming that the power spectral density of the signal is flat.

This paper is organized as follows: In Sec. II, problem formulation and some preliminaries are presented. In Sec. III, we analyze a very famous subspace-based algorithm, multiple signal classification (MUSIC), and discuss the problems of narrowband algorithms in the detection of wideband sources. In Sec. IV, the method of overcoming the problem through array geometry is presented. In Sec. V, the reliability of the method will be tested through practical experiments, and finally, the paper will be concluded in Sec. VI.

II. PROBLEM FORMULATION AND PRELIMINARIES

When the bandwidth of narrowband signals is small in relation to the center frequency f_c, the phase shift remains approximately constant over the bandwidth. The delayed signal in the time domain is

$$S(f)\exp(-j2\pi ft) \approx S(f)\exp(-j2\pi f_c t) \leftrightarrow s(t)\exp(-j2\pi f_c t). \tag{1}$$

The phase shift is now time independent, where the time delay t is a function of the location of the source relative to the array elements. When the signal is mixed with a pure tone at the center frequency (i.e., baseband conversion), the outputs are viewed as approximately constant signals along a linear array due to M far-field sources. In Sec. III, the common model for array outputs will be given in order to analyze the array in both narrowband and wideband cases.

Signals emanating from M remote sources and monitored in the presence of noise at N spatially separated sensors can be mathematically modeled by a weighted sum of steering vectors embedded in noise

$$\underline{x}(t) = A(\theta,\varphi)\underline{s}(t) + \underline{n}(t), \tag{2}$$

where $\underline{x}(t)$, $\underline{s}(t)$, and $\underline{n}(t)$ are received, emitted, and additive white Gaussian (AWG) noise signal vectors, respectively. $A(\theta,\varphi)$ is an $N\times M$ matrix containing array manifold vectors as

$$A(\theta,\varphi) = [\underline{a}_g(\theta_1,\varphi_1),\underline{a}_g(\theta_2,\varphi_2),\ldots,\underline{a}_g(\theta_M,\varphi_M)].$$

Here, $\underline{a}_g(\theta,\varphi)$ is array manifold vector

$$\underline{a}_g(\theta,\varphi) = \underline{g}(\theta,\varphi) \odot \underline{a}(\theta,\varphi), \tag{3}$$

where $\underline{g}(\theta,\varphi)$ is sensor gain vector. $\odot$ is Hadamard or element by element multiplication sign and

$$\underline{a}(\theta,\varphi) = \exp[(-j\pi\underline{d}_x\cos\theta - j\pi\underline{d}_y\sin\theta)\cos\varphi], \tag{4}$$

where $\underline{d}_x$ and $\underline{d}_y$ are the array sensor locations on x and y axes, respectively. Also, θ and φ are DOAs that are interested in estimating. Estimation of DOAs is performed through eigenvalue decomposition of R_{xx} as the covariance matrix of received signals. It can be shown that[8]

$$R_{xx} = A(\theta,\varphi)R_{ss}A^H(\theta,\varphi) + \sigma^2 I, \tag{5}$$

where R_{ss} and σ^2 are covariance matrix of $\underline{s}(t)$ and the noise power, respectively, I is an identity matrix, and H denotes conjugate transpose operator. Through decomposition of R_{xx} into signal and noise subspaces and applying orthogonal property, MUSIC function derives

$$\xi(\theta) = \frac{\underline{a}^H(\theta)E_nE_n^H\underline{a}(\theta)}{\underline{a}^H(\theta)\underline{a}(\theta)} = \mathrm{tr}(P_\theta E_n E_n^H), \tag{6}$$

where E_n represents $M-N\times N$ noise subspace matrix containing noise eigenvectors, tr(.) is trace function of a given matrix, and P_θ is defined as a projection matrix, which shows the orthogonality between array manifold and noise subspace as

$$P_\theta = \underline{a}(\theta)[\underline{a}^H(\theta)\underline{a}(\theta)]^{-1}\underline{a}^H(\theta). \tag{7}$$

Note that elevation would be ignored for simplicity without any lost in generalization.

III. NARROWBAND MUSIC ALGORITHM ANALYSIS AND WIDEBAND SIGNALS

To understand MUSIC algorithm behavior against wideband signals, the basic elements of equation $\xi(\theta)$, which contain frequency, are extracted in this section. As a matter of fact, $\xi(\theta)$ is a function of different parameters while it seems it consists of only one parameter θ. One of these parameters is frequency f. Here, we are trying to extract this parameter through an exact analysis of $\xi(\theta)$, and in Sec. IV, we will deal with the issue of how to make $\xi(\theta)$ frequency independent.

In order to reach our goal, we refer to Eq. (7) and start from

$$\underline{a}^H(\theta)\underline{a(\theta)} = |I_1|^2 + |I_2|^2 + \ldots + |I_N|^2 = Q, \tag{8}$$

where I_i is a function of $\underline{g}(\theta,\varphi)$ and Q is a constant. So P_θ changes to

$$\underline{a}(\theta)[\underline{a}^H(\theta)\underline{a}(\theta)]^{-1}\underline{a}^H(\theta) = Q^{-1}\underline{a}(\theta)\underline{a}^H(\theta).$$

Now $\underline{a}(\theta)\underline{a}^H(\theta)$ is an $N\times N$ matrix with ith and kth elements

$$I_iI_k\exp(j\beta\cos\varphi(Dx_{ik}\cos\theta + Dy_{ik}\sin\theta)), \tag{9}$$

where

$$\begin{aligned} Dx_{ik} &= dx_i - dx_k, \\ Dy_{ik} &= dy_i - dy_k, \end{aligned} \quad 1 \le i,k \le N. \tag{10}$$

Here, $\beta=(2\pi/v)f$ and v are constant phase number and the velocity of wave in environment, respectively.

In this research, the main goal is to use narrowband algorithms directly on wideband signals, without any more complex computations. This is done only through a proper array configuration. So huge computation and analysis must be performed to remove the frequency parameter from the MUSIC function mathematically and make it proper to apply on wideband signals. Since this is not a simple task, we can

TABLE I. Estimated values of azimuth in different frequency scales of an equi-space circular array with quarter wavelength radius ($\lambda_1/4$). $N=24$ and SNR=10 dB.

	$\theta°$				
F	18	22.5	30	45	60
1	17.1887	22.9183	31.5127	45.8366	60.1606
3	114.591	22.9183	171.887	123.185	80.2141
5	77.3493	45.8366	48.7014	143.239	60.1606
7	42.9718	103.1324	0	68.7549	60.1606
10	68.7549	177.6169	157.563	14.3239	65.8901

modify the function to be fixed or constant in a given frequency band. Referring to previous equations, we see that the only parts containing the frequency are exponential functions such as $\exp(j\beta\cos\varphi(Dx_{ik}\cos\theta+Dy_{ik}\sin\theta))$. It can be seen that the argument of the exponential function depends directly on frequency. There are two other parameters except the frequency, DOA or (θ,φ), and sensor location differences or (Dx,Dy) along two x and y axes. Through analysis, we yield the exponential function in Eq. (9) that is the only part of $\xi(\theta)$, which comprises frequency as a variable. Now, exploiting this analysis, to keep $\xi(\theta)$ constant in a wide frequency bandwidth, we have to allocate the places of sensors in a way

$$(\beta Dx_{ik}\cos\theta+\beta Dy_{ik}\sin\theta)\cos\varphi=C_{ik}. \tag{11}$$

For a given DOA, Eq. (11) will be simply shown as

$$\begin{aligned}\beta dx_i&=C_{i1},\\ \beta dy_i&=C_{i2},\end{aligned}\quad \forall i, \tag{12}$$

where C_{ik}, C_{i1}, and C_{i2} are arbitrary constants. Also Eq. (12) is valid for all i's, and all C_i's must be the same values. Here, we introduce F as frequency scale that is $F=f_h/f_l$, where f_h and f_l are the highest and lowest frequencies, respectively, in the given band $[f_l,f_h]$. In wideband frequency systems, F is applied mostly instead of frequency bandwidth $\mathrm{BW}=f_h-f_l$. Usually, the system that works with $F>1$ is called wideband, and for $F\gg1$ or $F\geq10$, it is called frequency independent system (Ref. 9, Chap. 6). In Table I, DOA estimation of an equi-space circular array with quarter wavelength radius $(\lambda_1/4)$, where λ_1 is the wavelength that depends on $F=1$, is shown. First column is for different frequency scales and the first row is for true DOAs. The cells in Table I are filled by the estimated values of DOA through MUSIC algorithm in corresponding frequency scales. The results of narrowband algorithm in different frequency scales are not the same, as we expected. It shows that it is not permitted to apply a narrowband algorithm directly on wideband sources. In Sec. IV, we present another table, which clarifies that, in particular cases, it is allowed to use the narrowband algorithm. As a matter of fact, we will describe how to make a narrowband algorithm frequency independent.

IV. ARRAY GEOMETRY DETERMINATION

So far, we analyzed an array to detect wideband sources through narrowband algorithms. Here, the results of Secs. II and III will be used to design and define an array geometry or array sensor locations.

In Sec. III, we came to an agreement to choose dx_i and dy_i proportional to the inverse of frequency in order to make the exponential function frequency independent. It means that $Dx_i\cos\theta+Dy_i\sin\theta\propto1/f$. Since f belongs to a frequency range, it can get an infinite number of values, and so, the number of positions (d_x,d_y) will be infinite! Therefore, we define specific frequencies f_k, where $k=1,2,\ldots,K$, and try to make $\xi(\theta)$ have the same values in these frequencies. For every f_k, a set of positions could be derived. So, a K set of array geometry can be obtained. However, we are interested in finding just one array acting the same in all the frequencies. The problem can be solved if we consider kth sensor acts in kth frequency f_k and when frequency changes to f_j, the jth sensor plays the kth element role. In this way, every element acts just in one frequency and the positions of sensors are in a periodic state. As a matter of fact, every two frequencies f_n and f_{n+1} are related to each other through

$$\tau=\frac{D_n}{D_{n+1}}=\frac{\lambda_n}{\lambda_{n+1}}=\frac{f_{n+1}}{f_n}, \tag{13}$$

where $D_n=Dx_n\cos\theta+Dy_n\sin\theta$ and λ is wavelength. τ is called similarity factor that shows how much array behaviors in f_n and f_{n+1} are the same. Proper functions are exponential for this reason. The following equation shows a periodic curve, which one sensor in every period is placed:

$$\tau=\exp(2\pi|k_0|), \tag{14}$$

where k_0 is an arbitrary number. The curve shape of above equation is such a spiral geometry. For the convenience, we consider a spiral array sensor with

$$\rho=\rho_0\exp(k_0\vartheta), \tag{15}$$

where ρ_0 is spiral radius and ϑ represents angle.

Considering Eq. (15) in $\xi(\theta)$, it will be periodic

$$\xi(\gamma\upsilon)=\xi(\upsilon), \tag{16}$$

where $\gamma=\tau^k$. If $\beta D_n=\alpha\tau^{-n}$, then for every period with the same performance, only one sensor is taken. This leads different sensors to behave vice-versa substitutionally in different frequencies, and so, Eq. (13) is satisfied. This subject was tested for antenna array gain in Ref. 10.

The crucial subject is the proper value selection of α and τ to make the MUSIC function constant over a broadband frequency. Experiments clear that considering $1<\alpha<2$ and $0<\tau<1$ is completely suitable for wideband applications.

Here we compute the sensor positions of a one-dimensional, 24 element array with $\tau=0.98$ and $\alpha=1.9$ as follows:

$$\begin{aligned}d_x=[&-2.4213,-2.3728,-2.3254,-2.2789,-2.2333,\\ &-2.1886,-2.1449,-2.1020,-2.0599,-2.0187,\\ &-1.9783,-1.9388,1.9388,1.9783,2.0187,\end{aligned}$$

TABLE II. Estimated azimuth for a linear periodic array. The first row and column show true azimuths and frequency scales, respectively. The other values are obtained from simulation of MUSIC algorithm for azimuth direction.

	$\theta°$				
F	18	22.5	30	45	60
1	19.0901	21.9549	30.5493	45.8366	60.1606
3	19.0901	24.8197	30.5493	44.8732	60.1606
5	17.1887	22.9183	28.6479	44.8732	60.1606
7	17.1887	22.9183	30.5493	44.8732	60.1606
10	19.0901	21.9549	30.5493	44.8732	60.1606

2.0599,2.1020,2.1449,2.1886,2.2333,

2.2789,2.3254,2.3728,2.4213].

Table II shows the estimated result of the array in different azimuths for different frequencies. In comparison with Table I, the estimated azimuths are approximately close to the actual ones, while both arrays are linear. Meanwhile, Fig. 1 clears simulation results for true azimuth $\theta=\pi/6$ and SNR=10 dB. As it is obvious the occurred error on detection of narrowband MUSIC algorithm in the given bandwidth is rather negligible. Therefore, the array shape is a crucial and effective parameter to make wideband algorithms to be real-time.

In the last case, we assumed that the array is linear. For planar array, e.g., spiral, presume that λ_h and λ_l are highest and lowest wavelengths, respectively. Then, using Eq. (15), and assuming $\rho_0=\lambda_h$ and $R\propto\lambda_l$, it gives

$$\frac{\rho_i}{\rho_0}=\frac{\lambda_{N-i}}{\lambda_h}=\exp(k_0\vartheta_i), \tag{17}$$

and so

$$\vartheta_i=\frac{1}{k_0}\ln\frac{\lambda_{N-i}}{\lambda_h}. \tag{18}$$

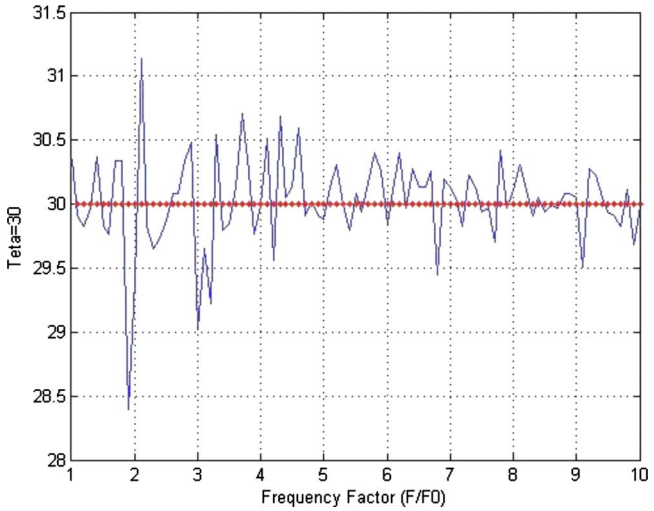


FIG. 1. (Color online) Azimuth detection versus frequency. This array is able to decrease the detected azimuth error in a wideband frequency range. $N=24$, $\tau=0.98$, $\alpha=1.9$, true azimuth $\theta=\pi/6$, and SNR=10 dB.

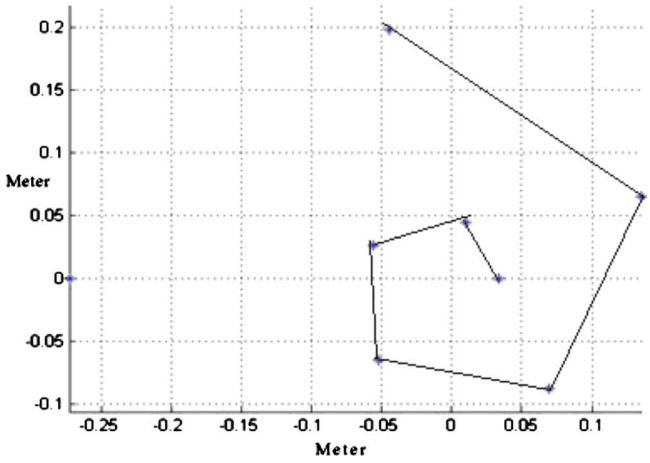


FIG. 2. (Color online) This illustrates sensors locations of a spiral array to perform in a wideband frequency range.

The locations of sensors can be computed through these two equations, when R is considered as the biggest radius of spiral. Figure 2 shows the locations for an array with $[f_l,f_h]=[314.81,2500]$ Hz. Connecting the points in this figure shows a spiral. Applying this array to detect targets with true azimuths $\theta=22.5°$, 30°, 60°, and 90° through MUSIC simulations in Fig. 3.

V. EXPERIMENTAL RESULTS

Here, we set up a measurement system to evaluate our methods in DF. The configuration is made up of an AKG microphone (c-417 model)[11] (Denmark) and a Delta 1010 card,[12] which is connected through an amplifier circuit to linear (or planner) array in x axis with $d_x=[0\ \ 0.092\ \ 0.185\ \ 0.277\ \ 0.370\ \ 0.462\ \ 0.555\ \ 0.647]$ m (Fig. 4). The test is done on three frequencies, i.e., 1837, 3490, and 6337 Hz. Table III shows the result of estimation algorithm in three different frequencies for three given true azimuths. The emitted signal is originated from a loudspeaker fed with a computer. These three frequencies are produced by MATLAB programming. This is clear that for every azimuth angle, different estimations for different frequencies can be obtained. Therefore, the results of this algorithm are not valid except for the designed frequency.

Now, we test this algorithm with above frequencies on a planner array with spiral shape with $d_x=[0.092\ \ 0.065\ \ 0\ \ -0.065\ \ -0.092\ \ -0.065\ \ 0\ \ 0.0654]$ m and $d_y=[0,0.065,0.092,0.065,0,-0.065,-0.092,-0.065]$ m as array sensor positions in x and y axes, respectively. Table IV shows that those results yielded for linear array in the last example. This shows a good compromise between the presented methods in this paper and the actual results. Therefore, in order to exploit such a narrowband algorithm against wideband signals, we can only find some array geometries (sensor positions), whose characteristics such as azimuths or even elevations are not affected with frequency variations. For both tests, it is considered that SNR=10 dB, $L=192\,000$ samples, sampling time=1 s, and $\varphi=0°$.

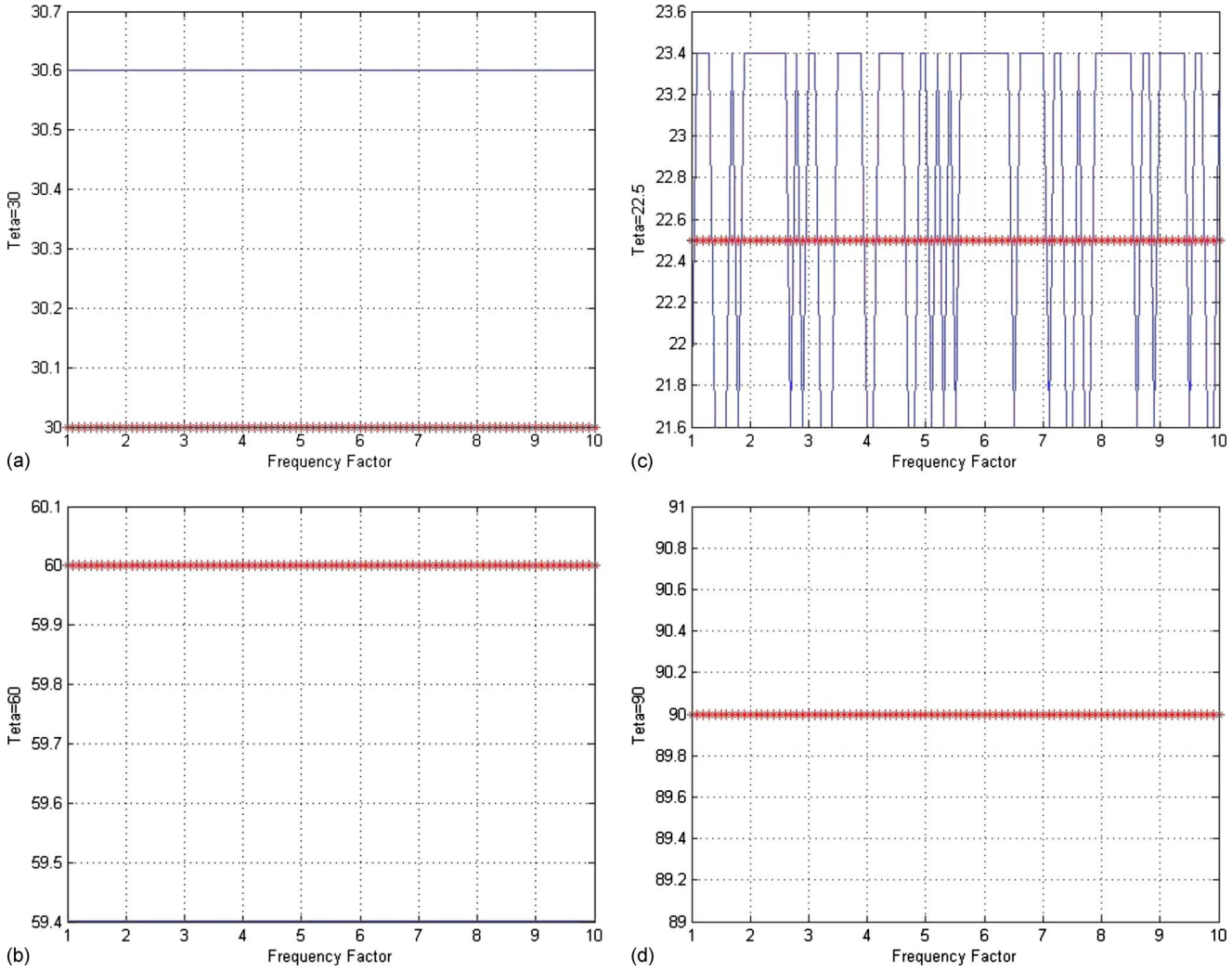


FIG. 3. (Color online) Simulation results of MUSIC algorithm for a spiral array with sensors locations as in Fig. 2. Azimuth versus frequency scale, true azimuth: (a) θ=30°, (b) θ=60°, (c) θ=22.5°, and (d) θ=90°; + denotes true azimuth direction and − denotes estimated azimuth through MUSIC algorithm.

VI. CONCLUSION

In this paper, a new method for using narrowband algorithm against wideband signals was discussed to detect and estimate their parameters such as azimuth. Since the estimated parameter is valid only for one frequency and in other frequencies in the band, wrong responses are derived. This problem is solved through designing the position of array sensors the way that in every frequency, the MUSIC response is the same. So, we analyzed the narrowband algorithm and pulled out the sections of the formula, which were imposed on the frequency change. It was shown that these section behaviors were neutralized through appropriate array configurations. In fact, we substitute the effect of a sensor position in one frequency with the others. This means that

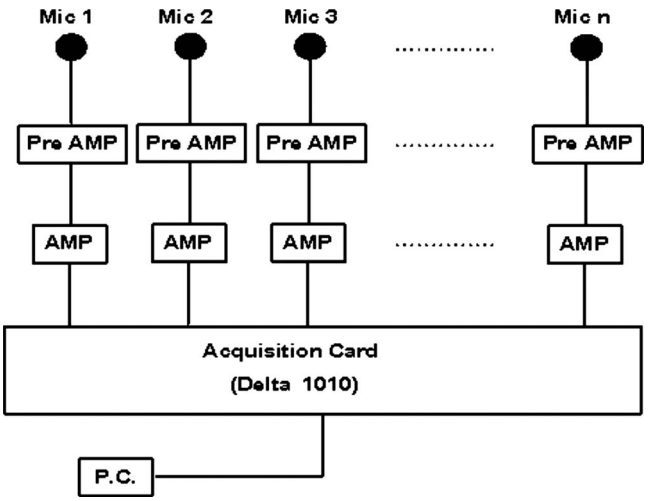


FIG. 4. Block diagram of practical measurement system.

TABLE III. Measurement results of a linear array in different azimuths and frequency scales. The first row and column show frequencies and true azimuths, respectively. The other values are measured from the setup system in Fig. 4 for azimuth direction.

	f (Hz)		
θ°	1837	3491	6337
25.56	0	99.3	150.4
90	90.3	90.2	90
135	134.3	110.89	63.1

TABLE IV. Measurement results of a spiral array in different azimuths and frequencies. First row and column show frequencies and true azimuths, respectively. The other values are measured from the setup system in Fig. 4 for azimuth direction. Good agreement between true and measured values of azimuth direction can be seen.

	f (Hz)		
$\theta°$	1837	3491	6337
25.56	22.6	26.9	28.2
90	90.5	91.3	88.8
135	128.8	132.1	135.2

every sensor acts in one certain frequency and so N sensors act the same in N frequency. Thus, increasing the number of sensors increases the flatness of frequency response of azimuth estimation through MUSIC algorithm. Such an array behavior with the above performance has already been introduced in some lectures to produce wideband beams in different applications. Spiral (array) antenna or nonlinear array antenna holds this property. Applying their formulas in MUSIC equations shows that a periodic geometry of array is able to provide a periodic MUSIC algorithm. In other words, the periodical geometry allows the algorithm to have the same value in every frequency. Therefore, it acts very flat in a wideband system.

Moreover, using a narrowband algorithm against wideband signals decreases the processing time and makes the algorithm a real-time one. This is because it removes numerous narrowband algorithms in the structure of wideband algorithm, as every one needs its own processing time to obtain a true result.

[1]Y. T. Lo, "A mathematical theory of antenna arrays with randomly spaced elements," IEEE Trans. Antennas Propag. **12**, 257–268 (1964).
[2]X. Huang, J. P. Reilly, and M. Wong, "Optimal design of linear array of sensors," in Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (1991), pp. 1405–1408.
[3]R. M. Leahy and B. D. Jeffs, "On the design of maximally sparse beamforming arrays," IEEE Trans. Antennas Propag., **39**, 1178–1187 (1991).
[4]M. I. Skolnik, G. Nemhauser, and J. W. Sherman, "Dynamic programming applied to unequally spaced arrays," IEEE Trans. Antennas Propag. **12**, 35–43 (1964).
[5]D. Lake, "Harmonic phase coupling for battlefield acoustic target identification," Proceedings of the IEEE Conference on Acoustics, Speech, and Signal Processsing (ICASSP'98), Seattle, WA (1998).
[6]M. Agrawal and S. Prasad, "Broadband DOA estimation using spatial-only modeling of array data," IEEE Trans. Signal Process. **48**, 663–670 (2000).
[7]G. Su and M. Morf, "The signal subspace approach for multiple wideband emitter location," IEEE Trans. Acoust., Speech, Signal Process. **31**, 1502–1522 (1983).
[8]A. Swindlehurst and T. Kailath, "A performance analysis of subspace-based methods in presence of modal errors. I. MUSIC algorithm," IEEE Trans. Signal Process. **40**, 1758–1774 (1992).
[9]W. G. Stutzman and G. A. Thiele, *Antenna Theory and Design*, 2nd ed. (Wiley, New York, 1998).
[10]D. H. Werner, P. L. Werner, and A. J. Ferraro, "Frequency independent features of self-similar fractal antennas," Proceedings of the IEEE AP-S International Symposium and URSI Radio Science Meeting, Vol. **3**, pp. 2050–2053.
[11]Professional miniature condenser microphone, retrieved from http://www.akg.com/site/products/powerslave,id,274,pid,274,nodeid (Last viewed April, 2008).
[12]PCI audio interfaces, retrieved from http://www.m-audio.com/products/en_us/Delta1010-main.html (Last viewed April, 2008).

Distribution theory approach to implementing directional acoustic sensors

Dean J. Schmidlin
Department of Electrical and Computer Engineering, University of Massachusetts Dartmouth, 285 Old Westport Road, North Dartmouth, Massachusetts 02747-2300

(Received 15 February 2009; revised 3 November 2009; accepted 5 November 2009)

The objective of directional acoustic sensors is to provide high directivity while occupying a small amount of space. An idealized point sensor achieves this objective from a knowledge of the spatial partial derivatives of acoustic pressure at a point in space. Direct measurement of these derivatives is difficult in practice. Consequently, it is expedient to come up with indirect methods. The use of pressure sensors to construct finite-difference approximations is an example of such a method. This paper utilizes the theory of distributions to derive another indirect method for estimating the various spatial partial derivatives of the pressure. This alternate method is then used to construct a multichannel filter which processes the acoustic pressure by mean of three-dimensional integral transforms throughout a 6ε-length cube centered at the origin. The output of the multichannel filter is a spatially and temporally filtered version of the pressure at the origin. The temporal filter is a lowpass Gaussian filter whose bandwidth is inversely proportional to ε. Finally, the lattice method for numerical multiple integration is utilized to develop a discrete-spatial version of the multichannel filter. © *2010 Acoustical Society of America.* [DOI: 10.1121/1.3268610]

PACS number(s): 43.60.Fg, 43.30.Wi [EJS] Pages: 292–299

I. INTRODUCTION

This paper is concerned with the further development of the theory of directional acoustic sensors. Cray *et al.* (2003) considered a limited combined receiver that measures pressure gradients to any order along the z axis. The normalized noise power (normalized with respect to the noise power out of an omnidirectional sensor) of the directional receiver was represented as a quadratic form whose matrix is a Hankel matrix. The method of Lagrange multipliers was used to maximize the directivity (array gain) against three-dimensional (3D) isotropic noise, and a matrix expression was given for the weights of the directional receiver. Schmidlin (2007a) extended the results to a directional acoustic sensor operating in all three dimensions. The same Hankel matrix was obtained and explicit formulas were derived for the coefficients or weights of the sensor as well as a formula for the maximum directivity index (DI) (array gain). The maximum directivity index achievable by a directional acoustic sensor of order ν is $20\log(1+\nu)$. This formula is the same as the one given by Franklin (1997) for the case of a superdirective line array. The directivities calculated for sensor orders going from 1 to 5 match the directivities determined by Cray (2002).

With regard to the design and implementation of directional acoustic receivers, Cray (2001) clearly demonstrated that a velocity sensor can be more sensitive to subsonic noise sources than an omnidirectional pressure sensor and concluded that reducing a directional sensor's response to subsonic self-noise will remain a challenging task. Hines *et al.* (2003, 2000) and Humphrey *et al.* (2003) analyzed the performance of a fifth order superdirective line array consisting of six pressure sensors separated by a distance of 0.16 m. It was demonstrated that impressive gains can be achieved despite the presence of degrading system noise. The various pressure gradients were estimated by finite-difference approximations.

Schmidlin (2007b) introduced a generalized function approach for estimating spatial partial derivatives of pressure. This approach will be utilized herein to develop a theoretical scheme for implementing directional acoustic sensors of arbitrary order. Section II defines a directional acoustic sensor of order ν by means of a one-dimensional (1D) discriminating polynomial of degree ν and two associated discriminating functions (one a 1D function and the other a 2D function). This definition makes clear the presence of null contours in the $\theta\phi$-plane. In Sec. III, the 2D discriminating function is applied to the problem of discriminating among plane waves. The result is a multichannel spatial filter. A generalized function approach is then presented for the estimation of spatial partial derivatives. The final part of Sec. III uses the distribution theory approach to come up with an approximation to the spatial filter. The implementation of this filter is the concern of Sec. IV.

II. DIRECTIONALITY OF GENERALIZED ACOUSTIC SENSORS

A. Directional acoustic sensors of order ν

Figure 1 illustrates a directional acoustic sensor situated at the origin of a rectangular coordinate system. The objective of the sensor is to discriminate between two wavefronts, one arriving from the "look" direction represented by the unit-vector $\mathbf{u}_L$ and the other arriving from an arbitrary direction represented by the unit-vector $\mathbf{u}$. The directional capability of the acoustic sensor can be described by a one-

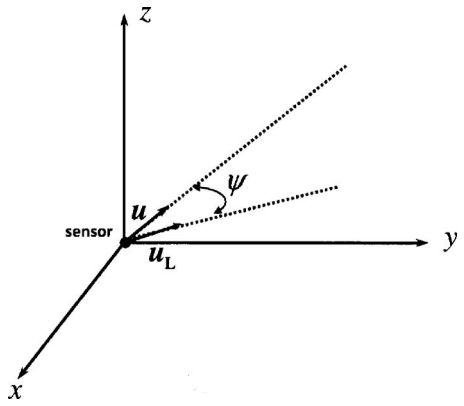


FIG. 1. Directional acoustic sensor at the origin of a rectangular coordinate system.

dimensional discriminating polynomial $d(u)$ and two associated discriminating functions $d_1(\psi)$ and $d_2(\theta,\phi)$, where

$$d(u)=\sum_{n=0}^{\nu} a_n u^n, \tag{1}$$

$$d_1(\psi)=\sum_{n=0}^{\nu} a_n \cos^n \psi, \tag{2}$$

$$d_2(\theta,\phi)=\sum_{n=0}^{\nu} a_n(\mathbf{u}\cdot\mathbf{u}_L)^n, \tag{3}$$

$$\mathbf{u}=\sin\phi\cos\theta\mathbf{i}+\sin\phi\sin\theta\mathbf{j}+\cos\phi\mathbf{k}, \tag{4}$$

$$\mathbf{u}_L=\sin\phi_L\cos\theta_L\mathbf{i}+\sin\phi_L\sin\theta_L\mathbf{j}+\cos\phi_L\mathbf{k}, \tag{5}$$

where θ and θ_L are azimuthal angles and ϕ and ϕ_L are elevation angles. The dot product $\mathbf{u}\cdot\mathbf{u}_L$ has the value

$$\mathbf{u}\cdot\mathbf{u}_L=\sin\phi\sin\phi_L\cos(\theta-\theta_L)+\cos\phi\cos\phi_L=\cos\psi, \tag{6}$$

where the angle ψ is the angle between the two unit-vectors $\mathbf{u}$ and $\mathbf{u}_L$ measured in the plane formed by these two vectors (see Fig. 1).

The degree of the discriminating polynomial is ν, which is defined as the order of the directional acoustic sensor. Silvia (2001) investigated theoretically and experimentally the dyadic sensor, which is a directional acoustic sensor of order 2. The polynomial coefficients a_i of $d(u)$ are determined from some pre-selected criterion. One notable criterion is the maximization of the *directivity index*, as defined by Nuttall and Cray (2001). The DI is the decibel measure of the improvement in the signal-to-noise ratio that a directional sensor provides in a 3D isotropic noise field with a perfectly correlated signal, relative to an omnidirectional sensor in the free field. Explicit formulas for the coefficients a_i that maximize the DI have been derived by Schmidlin (2007a). The expressions for the coefficients depend on whether the order ν of the sensor is even or odd. For an even-ordered sensor, the coefficients are given by

$$a_{2i}=\beta(-1)^i\binom{m}{i}(2i+1)(2i+3)\cdots(2i+2m+1),$$

$$i=0,1,\ldots,m,$$

$$a_{2i+1}=-\frac{2(m-i)}{2i+1}a_{2i},\quad i=0,1,\ldots,m-1, \tag{7}$$

where

$$m=\tfrac{1}{2}\nu,$$

$$\beta=\frac{(-1)^m}{(2m+1)^2 2^m m!}. \tag{8}$$

If the order of the sensor is odd, then

$$a_{2i}=\beta(-1)^i\binom{m}{i}(2i+1)(2i+3)\cdots(2i+2m+1),$$

$$i=0,1,\ldots,m,$$

$$a_{2i+1}=\frac{2i+2m+3}{2i+1}a_{2i},\quad i=0,1,\ldots,m, \tag{9}$$

where

$$m=\tfrac{1}{2}(\nu-1),$$

$$\beta=\frac{(-1)^m}{(m+1)^2 2^{m+2} m!}. \tag{10}$$

The value of the maximum directivity index is $20\log(1+\nu)$. For sensors of orders 1, 2, and 3, the values of the sensor coefficients are given by

$$\boldsymbol{a}=[1/4\quad 3/4], \tag{11}$$

$$\boldsymbol{a}=[-1/6\quad 1/3\quad 5/6], \tag{12}$$

$$\boldsymbol{a}=[-3/32\quad -15/32\quad 15/32\quad 35/32]. \tag{13}$$

The maximum DIs are 6.02, 9.54, and 12.04 dB, respectively. The acoustic sensor of order 1 and of order 2 has been referred to in the acoustics literature as the vector sensor and the dyadic sensor, respectively. Figure 2 are plots of $d_1(\psi)$. For all three cases, the value of $d_1(\psi)$ is 1 for $\psi=0$, which corresponds to the wavefront arriving in the "look" direction. For the situation where the wavefront is arriving in the opposite direction ($\psi=180°$), the function $d_1(\psi)$ are equal to $-1/2$, $+1/3$, and $-1/4$ for $\nu=1,2,3$, respectively. For an arbitrary value of ν, $d_1(180°)=(-1)^\nu/(\nu+1)$. This suggests that one measure of the effectiveness of a directional acoustic sensor is the quantity

$$20\log\left(\left|\frac{d_1(0°)}{d_1(180°)}\right|\right)=20\log(1+\nu), \tag{14}$$

which has the same value as the maximum directivity index. A glance at Fig. 2 reveals that as the order of the directional acoustic sensor increases, all of the peaks and valleys, other than at $\psi=0$, decrease in magnitude.

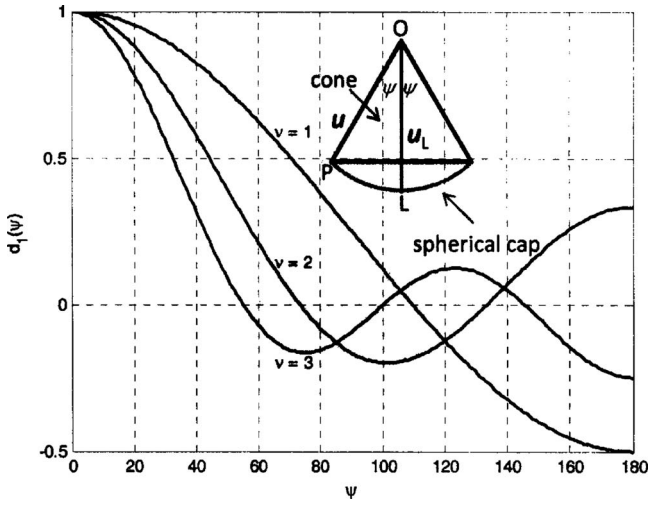


FIG. 2. Discriminating function $d_1(\psi)$ for sensors of orders 1, 2, and 3.

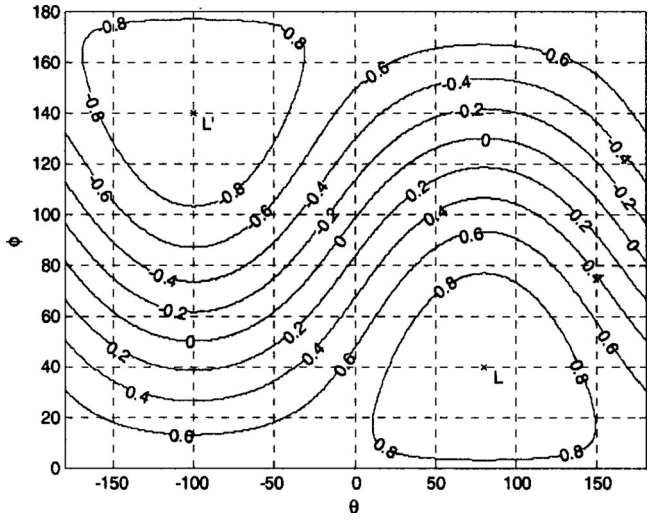


FIG. 3. Null contours as a function of a zero of the discriminating polynomial.

B. Null contours in the $\theta\phi$-plane

Another property that is apparent in Fig. 2 is that the number of zero crossings increases as the order of the sensor increases. The zero crossings occur at $\psi=109.5°$ for $\nu=1$, at 73.2° and 133.6° for $\nu=2$, and at 54.9°, 100.4°, and 145.4° for $\nu=3$. For a sensor of order ν, the discriminating function $d_1(\psi)$ experiences ν zero crossings. A value of ψ at which $d_1(\psi)=0$ is important because then the sensor perfectly discriminates against a wavefront that arrives at an angle of ψ relative to the look direction. Figure 2 also illustrates a conical sector of a unit-sphere centered about the origin of the rectangular coordinate system. This sector is defined by the two unit-vectors $\mathbf{u}$ and $\mathbf{u}_L$. As $\mathbf{u}$ rotates about $\mathbf{u}_L$ at a constant angle ψ, it becomes the generator of a right circular cone. The radius of the base-circle of the cone is $\sin\psi$, and its height is $\cos\psi$. As the unit-vector $\mathbf{u}$ rotates, it determines all of the directions of arrival of wavefronts that have the same angle ψ relative to the look direction. Thus, a null of the function $d_1(\psi)$ corresponds to a contour of nulls of $d_2(\theta,\phi)$ in the $\theta\phi$-plane.

In order to determine the null contours of a directional acoustic sensor, it is first necessary to factor the discriminating polynomial $d(u)$ as follows:

$$d(u) = a_\nu \prod_{i=1}^{\nu} (u - b_i), \tag{15}$$

where $b_i, i=1,\ldots,\nu$ are the roots or zeros of $d(u)$. The discriminating function $d_2(\phi,\theta)$ can be written as an analogous fashion, that is,

$$d_2(\theta,\phi) = a_\nu \prod_{i=1}^{\nu} (\mathbf{u}\cdot\mathbf{u}_L - b_i). \tag{16}$$

Equations (11)–(13) give the optimal values (for maximum DI) of the sensor coefficients for $\nu=1,2,3$. For these cases, the b_i coefficients are $b_1=-1/3$ for $\nu=1$, $b_1=-0.6899$ and $b_2=0.2899$ for $\nu=2$, and $b_1=-0.8228$, $b_2=0.5753$, and $b_3=-0.1811$ for $\nu=3$. For a maximum DI sensor of order ν, if the ν b_i coefficients are arranged in order of decreasing magnitude, i.e., $|b_1|>|b_2|>\cdots>|b_\nu|$, then the value of b_1 will be negative and the ones that follow will alternate in sign. Since all of the b_i are not greater than 1 in magnitude, each of the equations $\mathbf{u}\cdot\mathbf{u}_L=b_i$ defines a null contour in the $\theta\phi$-plane. Figure 3 provides contour plots of $\mathbf{u}\cdot\mathbf{u}_L=b$ for b going from -0.8 to $+0.8$ in steps of 0.2. The point L corresponds to the directions $\theta=\theta_L=80°$ and $\phi=\phi_L=40°$, whereas the point L' corresponds to the directions $\theta=\theta_{L'}=-180+\theta_L=-100°$ and $\phi=\phi_{L'}=180°-\phi_L=140°$. In terms of three-dimensional space, the point L is located on the unit-sphere at the tip of the vector $\boldsymbol{u}_L$. The point L' is on the unit-sphere diametrically opposite to the point L.

With the assistance of Fig. 3, one can observe the behavior of the null contour caused by a root b of the discriminating polynomial. The value of b is indicated along each null contour. Consider the behavior as b moves from 0 to -1. The contour starts off being oscillatory with a minimum and a maximum value of ϕ occurring when $\theta=\theta_{L'}=-100°$ and $\theta=\theta_L=80°$, respectively. When $b=-\cos\phi_L=-0.7660$, the null contour changes from being an open contour to a closed one. As b continues toward -1, the circumference of the contour decreases until the contour becomes a single point at $b=-1$. A similar description of the null contour holds as b moves from 0 to +1.

From Fig. 3 and a knowledge of the zeros of the discriminating polynomial of the acoustic sensor, one can get an idea of the null contours of the sensor. Contour plots of $d_2(\theta,\phi)$ are given in Figs. 4 and 5 for acoustic sensors of orders 2 and 3, respectively. Both sensors are weighted to produce maximum DI. The dyadic sensor represented by Fig. 4 contains two null contours. One null contour is close to the point L' and corresponds to the equation $\mathbf{u}\cdot\mathbf{u}_L=b_1=-0.6899$, whereas the other null contour is closer to the point L and corresponds to the equation $\mathbf{u}\cdot\mathbf{u}_L=b_2=0.2899$. The acoustic sensor of order 3 represented by Fig. 5 contains three null contours. Since a triad is a group of three things, the sensor may be appropriately referred to as a triadic sensor because of its three null contours. The location of these contours tends to significantly reduce the level of $|d_2(\theta,\phi)|$ in the space between the two points L and L'.

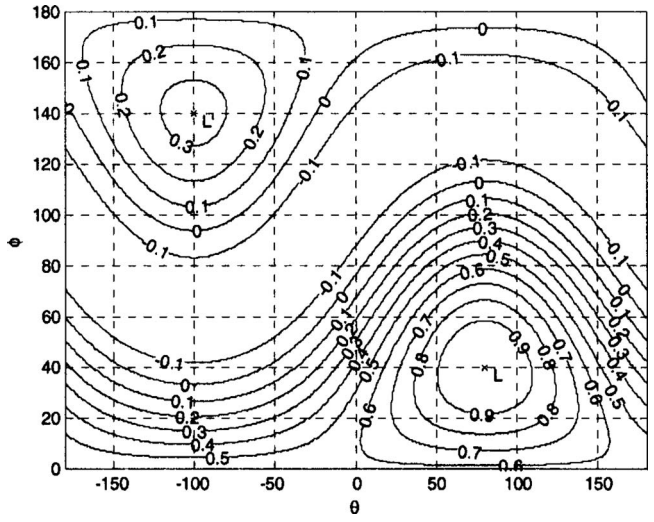


FIG. 4. The discriminating function $d_2(\theta,\phi)$ for a dyadic sensor.

III. DISCRIMINATING AMONG PLANE WAVES

A. Application of the discriminating function

In Fig. 6, a plane wave propagates toward the origin of the rectangular coordinate system. The acoustic pressure at the point (x,y,z) is

$$p(t,\mathbf{r}) = p(t+\mathbf{u}\cdot\mathbf{r}/c) = p\left(t+\frac{u_1x+u_2y+u_3z}{c}\right)$$

$$= p(t,x,y,z). \tag{17}$$

Consider the time function

$$p_{n_1,n_2,n_3}(t) = \left.\frac{\partial^{n_1+n_2+n_3}p(t,x,y,z)}{\partial^{n_1}x\partial^{n_2}y\partial^{n_3}z}\right|_{x=0,y=0,z=0}. \tag{18}$$

Using Eq. (17) to compute this function results in

$$p_{n_1,n_2,n_3}(t) = \frac{u_1^{n_1}u_2^{n_2}u_3^{n_3}}{c^{n_1+n_2+n_3}}\frac{\partial^{n_1+n_2+n_3}p(t)}{\partial t^{n_1+n_2+n_3}}. \tag{19}$$

A second time function of interest is

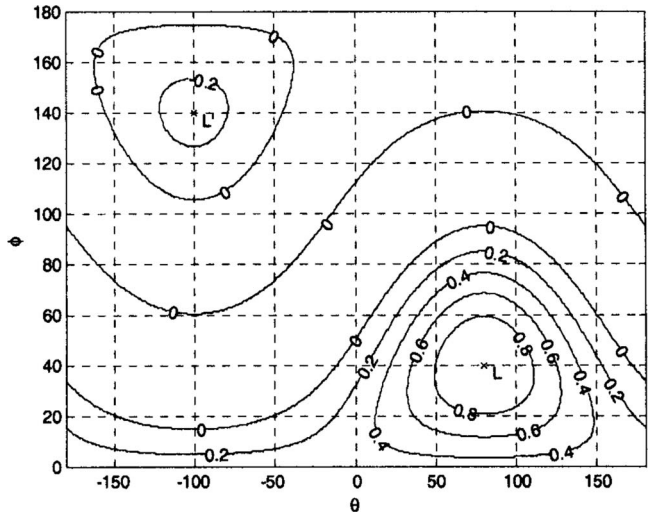


FIG. 5. The discriminating function $d_2(\theta,\phi)$ for a triadic sensor.

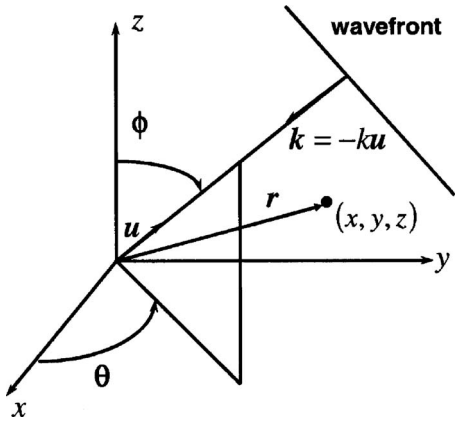


FIG. 6. Plane wave moving toward the origin of a rectangular coordinate system.

$$p_n(t,\mathbf{u}_L) = c^n \sum_{n_1+n_2+n_3=n} \binom{n}{n_1,n_2,n_3} u_{L_1}^{n_1}u_{L_2}^{n_2}u_{L_3}^{n_3}p_{n_1,n_2,n_3}(t), \tag{20}$$

where

$$\binom{n}{n_1,n_2,n_3} = \frac{n!}{n_1!\,n_2!\,n_3!} \tag{21}$$

is the multinomial coefficient. Substituting Eq. (19) into Eq. (20) and applying the multinomial formula lead to

$$p_n(t,\mathbf{u}_L) = (\mathbf{u}\cdot\mathbf{u}_L)^n p^{(n)}(t), \tag{22}$$

where $p^{(n)}(t)$ is the n-th time derivative of the acoustic pressure at the origin. Applying the operator

$$I^n[\ ] = \int\int\cdots\int [\ ]dt_1dt_2\cdots dt_n \tag{23}$$

to both sides of Eq. (22) yields

$$I^n p_n(t,\mathbf{u}_L) = (\mathbf{u}\cdot\mathbf{u}_L)^n p(t). \tag{24}$$

Multiplying both sides of Eq. (24) by a_n and summing from $n=0$ to $n=\nu$ give the desired result

$$p_s(t) = \sum_{n=0}^{\nu} a_n I^n[p_n(t,\mathbf{u}_L)] = d_2(\theta,\phi)p(t). \tag{25}$$

Equations (18), (20), and (25) suggest a processing method for discriminating among plane waves. The output $p_s(t)$ of the processor is equal to the product of the acoustic pressure of the wavefront at the origin and the discriminating function of the sensor. When there are μ wavefronts, then

$$p_s(t) = \sum_{m=1}^{\mu} d_2(\theta_m,\phi_m)p_m(t), \tag{26}$$

where $p_m(t)$ is the pressure at the origin of the m-th wavefront, and the angles θ_m and ϕ_m specify its direction. If ℓ represents the index of a plane wave arriving in the look direction, then ideally, $p_s(t)=p_\ell(t)$.

B. Distribution theory approach to estimating derivatives

The Dirac delta distribution $\delta(x)$ is defined by

$$\int_{-\infty}^{\infty} \delta(x) f(x) dx = f(0). \tag{27}$$

The Dirac delta distribution assigns to each function $f(x)$ its value at $x=0$. The function $\delta(x)$ is not an ordinary function but a generalized function. The integral appearing at the left side of Eq. (27) is not functional but only symbolic. It is not useful in coming up with the value at the right side of Eq. (27). However, one can find an ordinary function $\delta_\varepsilon(x)$ that is a function of a parameter ε and that provides an estimate of the value of $f(0)$, namely,

$$\int_{-\infty}^{\infty} \delta_\varepsilon(x) f(x) dx = \hat{f}(0). \tag{28}$$

As ε approaches zero, the estimate $\hat{f}(0)$ approaches the true value of $f(0)$. That is,

$$\lim_{\varepsilon \to 0} \int_{-\infty}^{\infty} \delta_\varepsilon(x) f(x) dx = f(0). \tag{29}$$

The n-th derivative of the Dirac delta distribution is the distribution defined by

$$\int_{-\infty}^{\infty} \delta^{(n)}(x) f(x) dx = (-1)^n f^{(n)}(0). \tag{30}$$

The distribution $(-1)^n \delta^{(n)}(x)$ assigns to the function $f(x)$ the value of its n-th derivative evaluated at $x=0$. An estimate of this derivative can be obtained by means of the integral

$$\int_{-\infty}^{\infty} (-1)^n \delta_\varepsilon^{(n)}(x) f(x) dx = \hat{f}^{(n)}(0), \tag{31}$$

where the function $\delta_\varepsilon^{(n)}(x)$ is the n-th derivative of the function $\delta_\varepsilon(x)$. The function $\delta_\varepsilon(x)$ is called a nascent delta function. There are many such functions to choose from. One notable example is the function

$$g_\varepsilon(x) = \frac{1}{\varepsilon\sqrt{2\pi}} \exp\left(-\frac{x^2}{2\varepsilon^2}\right), \tag{32}$$

which is a Gaussian function of zero mean and a variance of ε. The first derivative of the Gaussian function is given by

$$g_\varepsilon'(x) = -\frac{1}{\varepsilon^3\sqrt{2\pi}} x \exp\left[-\frac{x^2}{2\varepsilon^2}\right]. \tag{33}$$

All of the derivatives of the Gaussian are smooth functions and can be generated recursively by means of Eqs. (32) and (33), and the relation

$$g_\varepsilon^{(n)}(x) = -\frac{1}{\varepsilon^2}[x g_\varepsilon^{(n-1)}(x) + (n-1) g_\varepsilon^{(n-2)}(x)],$$

$$n = 2, 3, \ldots . \tag{34}$$

The basic estimator that will be used to estimate the n-th derivative of a one-dimensional function at the origin is

$$\hat{f}^{(n)}(0) = \int_{-\infty}^{\infty} (-1)^n g_\varepsilon^{(n)}(x) f(x) dx. \tag{35}$$

For the special case $f(x) = e^{j\kappa x}$, Eq. (35) becomes

$$\hat{f}^{(n)}(0) = (-1)^n \int_{-\infty}^{\infty} g_\varepsilon^{(n)}(x) e^{j\kappa x} dx. \tag{36}$$

The integral on the right side of Eq. (36) represents the Fourier transform of the n-th derivative of the Gaussian function $g_\varepsilon(x)$. The value of this integral is $(-j\kappa)^n$ times the Fourier transform of $g_\varepsilon(x)$, which is $\exp(-\varepsilon^2\kappa^2/2)$. Equation (36) becomes

$$\hat{f}^{(n)}(0) = (j\kappa)^n \exp(-\varepsilon^2\kappa^2/2). \tag{37}$$

The true value of the n-th derivative of $f(x) = e^{j\kappa x}$ evaluated at $x=0$ is

$$f^{(n)}(0) = (j\kappa)^n. \tag{38}$$

As $\varepsilon \to 0$ in Eq. (37), the estimate of the derivative approaches the true value.

Equation (34) provides an iterative means for evaluating the various derivatives of the nascent delta function $g_\varepsilon(x)$. An explicit form can be written for these derivatives by considering the alternative or "probabilist" Hermite polynomials. These polynomials are defined by

$$\mathrm{He}_n(x) = (-1)^n \exp(x^2/2) \frac{d^n}{dx^n} \exp(-x^2/2). \tag{39}$$

Equation (39) together with Eq. (32) leads to

$$g_\varepsilon^{(n)}(x) = (-1)^n \frac{1}{\varepsilon^{n+1}\sqrt{2\pi}} \mathrm{He}_n(x/\varepsilon) \exp(-x^2/2\varepsilon^2). \tag{40}$$

In terms of $g_\varepsilon(x)$, Eq. (40) becomes

$$g_\varepsilon^{(n)}(x) = \frac{(-1)^n}{\varepsilon^n} \mathrm{He}_n(x/\varepsilon) g_\varepsilon(x). \tag{41}$$

The estimator specified by Eq. (35) will prove useful in providing a methodology for computing the function $p_n(t, \mathbf{u}_L)$ defined by Eq. (20).

C. Acoustic signal processing

The objective of this section is to develop the signal processing, which will produce the function $p_s(t)$. The use of generalized functions for evaluating derivatives can be applied to functions of dimension greater than 1. The function specified by Eq. (18) can be written as

$$p_{n_1,n_2,n_3}(t) = (-1)^{n_1+n_2+n_3} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p(t,x,y,z) \times \delta^{(n_1)}(x) \delta^{(n_2)}(y) \delta^{(n_3)}(z) dx dy dz. \tag{42}$$

Utilizing the Gaussian function as the nascent delta function, Eq. (42) can be approximated as

$$\hat{p}_{n_1,n_2,n_3}(t) = (-1)^{n_1+n_2+n_3}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p(t,x,y,z) \times g_\varepsilon^{(n_1)}(x) g_\varepsilon^{(n_2)}(y) g_\varepsilon^{(n_3)}(z)\,dx\,dy\,dz. \quad (43)$$

The utilization of Eq. (41) in Eq. (43) and the result in Eq. (20) lead to

$$\hat{p}_n(t,\mathbf{u}_L) = \frac{c^n}{\varepsilon^n}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p(t,\mathbf{r})\mathrm{He}_n(\mathbf{u}_L\cdot\mathbf{r}/\varepsilon)\hat{g}_\varepsilon(r)\,dx\,dy\,dz, \quad (44)$$

where $\hat{g}_\varepsilon(r)$ is the multidimensional Gaussian function

$$\hat{g}_\varepsilon(r) = \left(\frac{1}{\varepsilon\sqrt{2\pi}}\right)^3 \exp\left(-\frac{r^2}{2\varepsilon^2}\right). \quad (45)$$

In arriving at Eq. (44), use was made of the following relationship given by Eq. (4) on p. 196 on the work of Erdelyi *et al.* (1981):

$$\sum_{n_1+n_2+n_3=n}\binom{n}{n_1,n_2,n_3} u_{L1}^{n_1}u_{L2}^{n_2}u_{L3}^{n_3}\mathrm{He}_{n_1}(x)\mathrm{He}_{n_2}(y)\mathrm{He}_{n_3}(z) = \mathrm{He}_n(\mathbf{u}_L\cdot\mathbf{r}) \quad (46)$$

The acoustic pressure $p(t,x,y,z)$ may be written as

$$p(t,x,y,z) = \frac{1}{2\pi}\int_{-\infty}^{\infty} P(\omega)\exp(j\omega t)\exp(j\omega\mathbf{u}\cdot\mathbf{r}/c)\,d\omega. \quad (47)$$

By substituting Eq. (47) into Eq. (43), it can be shown with the aid of Eqs. (36) and (37) that

$$\hat{p}_{n_1,n_2,n_3}(t) = \frac{u_1^{n_1}u_2^{n_2}u_3^{n_3}}{c^n}p_\varepsilon^{(n)}(t), \quad (48)$$

where

$$p_\varepsilon(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} P(\omega)\exp[-\varepsilon^2\omega^2/2c^2]\exp(j\omega t)\,d\omega. \quad (49)$$

Analogous to the development that led to Eq. (22), one has

$$\hat{p}_n(t,\mathbf{u}_L) = (\mathbf{u}\cdot\mathbf{u}_L)^n p_\varepsilon^{(n)}(t). \quad (50)$$

The output of the processor is similar to that of Eq. (25) and is equal to

$$\hat{p}_s(t) = \sum_{n=0}^{\nu} a_n I^n[\hat{p}_n(t,\mathbf{u}_L)] = d_2(\theta,\phi)p_\varepsilon(t). \quad (51)$$

A comparison of Eqs. (51) and (25) reveals the effect of using the Gaussian function as the nascent delta function. The acoustic pressure $p(t)$ at the origin is replaced by the pressure $p_\varepsilon(t)$, which approaches $p(t)$ as $\varepsilon\to 0$. The output of the processor in the frequency domain is

$$\hat{P}_s(\omega) = \sum_{n=0}^{\nu} a_n(j\omega)^{-n}\hat{P}_n(\omega,\mathbf{u}_L) = d_2(\theta,\phi)P(\omega)\exp[-\varepsilon^2\omega^2/2c^2]. \quad (52)$$

It is clear from Eq. (52) that ε tends to attenuate the processor's output in the frequency domain. According to the rightmost side of Eq. (52), the processing produces a spatially and temporally filtered version of the acoustic pressure at the origin. The temporal filter is a lowpass Gaussian filter. The relationship between ε and the 3dB-down frequency of the filter is

$$\varepsilon = \frac{\sqrt{\ln 2}}{2\pi}\frac{c}{f_{3\mathrm{dB}}} = 0.1325\frac{c}{f_{3\mathrm{dB}}}. \quad (53)$$

An upper bound on ε may be determined by noting the highest frequency of interest, selecting the 3 dB down frequency accordingly, and using Eq. (53) to compute the value of the upper bound.

IV. IMPLEMENTATION OF DIRECTIONAL ACOUSTIC SENSORS

The most important part of the processing for implementing acoustic sensors of arbitrary order is that which generates the time function $\hat{p}_n(t,\mathbf{u}_L)$. It is noted that the limits of the one-dimensional integrals in Eq. (44) go from $-\infty$ to $+\infty$. However, this assignment of limits does not take into consideration the effective duration of the Gaussian function $g_\varepsilon(x)$. Using the "energy concentration" definition of duration discussed by Papoulis (1977), about 99.7% of the energy of a Gaussian function is contained within three standard deviations (3ε) of its mean. It is reasonable then that the infinite limits in the integrals of Eq. (44) be replaced by limits going from -3ε to $+3\varepsilon$. The function of Eq. (44) becomes

$$\hat{p}_n(t,\mathbf{u}_L) = \int_R K_n(\mathbf{u}_L,\mathbf{r})p(t,\mathbf{r})\,d\mathbf{r}, \quad (54)$$

where

$$K_n(\mathbf{u}_L,\mathbf{r}) = \frac{c^n}{\varepsilon^n}\mathrm{He}_n(\mathbf{u}_L\cdot\mathbf{r}/\varepsilon)\hat{g}_\varepsilon(\mathbf{r}\cdot\mathbf{r}). \quad (55)$$

The region R is a 6ε-length cube centered at the origin. For a fixed value of t, Eq. (54) is an integral transform from the function $p(t,\mathbf{r})$ to the function $\hat{p}_n(t,\mathbf{u}_L)$. The function $K_n(\mathbf{u}_L,\mathbf{r})$ is the kernel of the integral transform. There are many methods for evaluating the three-dimensional integral transform defined by Eq. (54). Krommer and Ueberhuber (1998) gave a comprehensive but concise survey of computational integration methods and the fundamental mathematical principles on which they are based. Especially appropriate to the problem at hand are lattice methods because they integrate over the unit hypercube. Lyness (1989) introduced the reader to lattice rules and their generator matrices, and Sloan and Joe (1994) gave a thorough coverage of lattice methods. Once a lattice rule is selected, the estimation of $\hat{p}_n(t,\boldsymbol{u}_L)$ takes on the form

$$\overline{\hat{p}_n}(t,\mathbf{u}_L) = \frac{1}{N}\sum_{i=0}^{N-1} K_n(\mathbf{u}_L,\mathbf{r}_i)p(t,\mathbf{r}_i), \quad (56)$$

where each of the $\mathbf{r}_i$ lies within the region R. The values of N and $\mathbf{r}_i$ depend on the choice of lattice rule. From Eq. (51) and (52), the output of the sensor's multichannel filter is

$$\hat{p}_s(t) = \sum_{n=0}^{\nu} a_n I^n[\hat{p}_n(t,\mathbf{u}_L)] \tag{57}$$

in the time domain and

$$\hat{P}_s(\omega) = \sum_{n=0}^{\nu} a_n (j\omega)^{-n} \hat{P}_n(\omega,\mathbf{u}_L) \tag{58}$$

in the frequency domain. Utilizing the lattice rule to implement the multichannel filter, the filter's output becomes

$$\overline{\hat{p}_s}(t) = \sum_{n=0}^{\nu} a_n I^n[\overline{\hat{p}_n}(t,\mathbf{u}_L)] \tag{59}$$

in the time domain and

$$\overline{\hat{P}_s}(\omega) = \sum_{n=0}^{\nu} a_n (j\omega)^{-n} \overline{\hat{P}_n}(\omega,\mathbf{u}_L) \tag{60}$$

in the frequency domain. The substitution of Eq. (56) into Eq. (59) yields

$$\overline{\hat{p}_s}(t) = \frac{1}{N}\sum_{i=0}^{N-1}\left(\sum_{n=0}^{\nu} a_n K_n(\mathbf{u}_L,\mathbf{r}_i) I^n[p(t,\mathbf{r}_i)]\right). \tag{61}$$

Taking the Fourier transform of $\overline{\hat{p}_s}(t)$ results in

$$\overline{\hat{P}_s}(\omega) = \frac{1}{N}\sum_{i=0}^{N-1}\left(\sum_{n=0}^{\nu} a_n K_n(\mathbf{u}_L,\mathbf{r}_i)(j\omega)^{-n}\right) P(\omega,\mathbf{r}_i). \tag{62}$$

Equations (61) and (62) describe the processing performed by the multichannel filter when implemented by means of a lattice rule. There are N pressure sensors. For each sensor, Eq. (61) indicates that the processing for the i-th sensor consists of a cascade connection of ν integrators, the output of the n-th integrator being multiplied by $a_n K_n(\mathbf{u}_L,\mathbf{r}_i)$, and all of the scaled integrator outputs summed. This processing is done for each sensor, and the results averaged to obtain the output of the multichannel filter. Equation (62) shows what is happening in the frequency domain. The Fourier transform of the pressure at each sensor is multiplied by a polynomial function of ω^{-1} of degree ν and the results averaged together.

The Hermite polynomials $\mathrm{He}_n(u)$ are orthogonal and satisfy the recurrence relation,

$$\mathrm{He}_n(u) = u\mathrm{He}_{n-1}(u) - (n-1)\mathrm{He}_{n-2}(u). \tag{63}$$

Because of the relationship indicated by Eq. (63), the kernel function $K_n(\mathbf{u}_L,\mathbf{r})$ satisfies a similar relation, namely,

$$K_n(\mathbf{u}_L,\mathbf{r}) = \frac{c}{\varepsilon}(\mathbf{u}_L\cdot\mathbf{r}/\varepsilon)K_{n-1}(\mathbf{u}_L,\mathbf{r}) - \left(\frac{c}{\varepsilon}\right)^2 (n-1)K_{n-2}(\mathbf{u}_L,\mathbf{r}). \tag{64}$$

Equation (64) permits the recursive computation of the kernel function for different values of n. The value of c/ε can be chosen with the aid of Eq. (53). The primary characteristic of the kernel function $K_n(\mathbf{u}_L,\mathbf{r})$ can be brought out by making use of Eqs. (41), (46), and (55). The result is

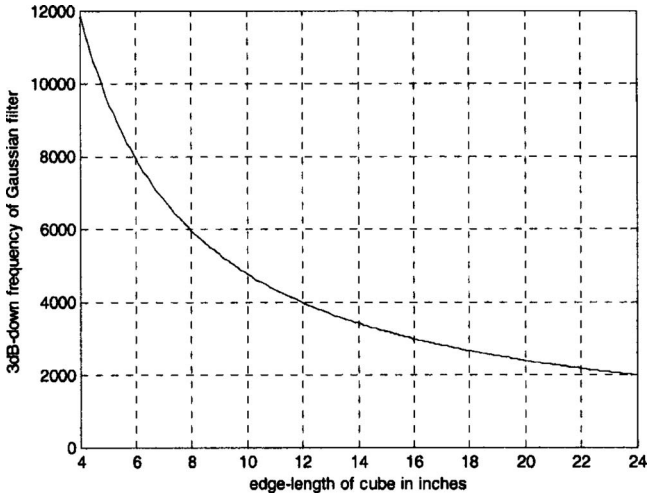


FIG. 7. Variation in the 3 dB down frequency of Gaussian filter with the edge-length of the cube centered at the origin.

$$K_n(\mathbf{u}_L,\mathbf{r}) = c^n \sum_{n_1+n_2+n_3=n} (-1)^n \binom{n}{n_1,n_2,n_3} u_{L1}^{n_1} u_{L2}^{n_2} u_{L3}^{n_3} \times g_\varepsilon^{(n_1)}(x) g_\varepsilon^{(n_2)}(y) g_\varepsilon^{(n_3)}(z). \tag{65}$$

As the value of ε approaches 0, the kernel function $K_n(\mathbf{u}_L,\mathbf{r})$ approaches

$$c^n \sum_{n_1+n_2+n_3=n} (-1)^n \binom{n}{n_1,n_2,n_3} u_{L1}^{n_1} u_{L2}^{n_2} u_{L3}^{n_3} \delta^{(n_1)}(x) \times \delta^{(n_2)}(y)\delta^{(n_3)}(z) \tag{66}$$

Then kernel given by Eq. (66) is a three-dimensional distribution and when operating on the acoustic pressure produces a linear combination of all of the spatial partial derivatives of order n of the pressure evaluated at the origin. The number of these derivatives is $(n+2)(n+1)/2$ and they are all succinctly incorporated into a single kernel function.

The case $\varepsilon\to 0$ corresponds to the idealized directional acoustic sensor, which is a point sensor located at the origin. This idealized sensor is the sensor illustrated in Fig. 1. For a nonzero ε, the directional acoustic sensor is no longer a point sensor but is one consisting of n pressure sensors that lie within a 6ε-length cube centered at the origin. These pressure sensors are not dedicated to estimating the spatial partial derivatives of the acoustic pressure at the origin. Their purpose is to estimate the set of integral transforms specified by Eq. (54) for $n=0,1,\ldots,\nu$. The pressure values from these sensors are used to carry out the estimates by means of Eq. (56). In practice, the value of ε must be large enough to accommodate the n pressure sensors, but not too large so as to filter out desirable frequency components of the acoustic pressure, as described by Eq. (52). Figure 7 shows the variation in the 3 dB down frequency of the Gaussian lowpass filter with the edge-length of the cube. The speed of sound is assumed to be 5000 ft/s. It is important to ascertain the most efficient lattice rules for estimating the integral given by Eq. (54). The goal is to achieve an acceptable accuracy with a small number of pressure sensors.

Cray, B. A. (**2001**). "Directional acoustic receivers: Signal and noise characteristics," in Proceedings of the Workshop on Directional Acoustic Sensors, Newport, RI.
Cray, B. A. (**2002**). "Directional point receivers: The sound and the theory," J. Geophys. Res., [Oceans] **2**, 1903–1905.
Cray, B. A., Evora, V. M., and Nuttall, A. H. (**2003**). "Highly directional acoustic receivers," J. Acoust. Soc. Am. **113**, 1526–1532.
Erdelyi, A. (**1981**). *Higher Transcendental Functions* (Krieger, Malabar, FL), Vol. **II**.
Franklin, J. B. (**1997**). "Superdirective receiving arrays for underwater acoustic application," DREA-CR-97-444, Contractor Report, Defense Research Establishment Atlantic, Dartmouth, Nova Scotia, Canada.
Hines, P. C., Humphrey, V. F., and Young, V. (**2003**). "Performance of a superdirective line array in nonideal environments," in Proceedings of the 146th Meeting of the Acoustical Society of America, Austin, TX.
Hines, P. C., Rosenfeld, A. L., Maranda, B. H., and Hutt, D. L. (**2000**). "Evaluation of the endfire response of a superdirective line array in simulated ambient noise environments," in Proceedings of the Oceans 2000, pp. 1489–1494.
Humphrey, V. F., Hines, P. C., and Young, V. (**2003**). "Experimental performance analysis of a superdirective line array," in Proceedings of the 146th Meeting of the Acoustical Society of America, Austin, TX.
Krommer, A. R., and Ueberhuber, C. W. (**1998**). *Computational Integration* (SIAM, Philadelphia).
Lyness, J. N. (**1989**). "An introduction to lattice rules and their generator matrices," IMA J. Numer. Anal. **9**, 405–419.
Nuttall, A. H., and Cray, B. A. (**2001**). "Approximations to directivity for linear, planar, and volumetric apertures and arrays," IEEE J. Ocean. Eng. **26**, 383–398.
Papoulis, A. (**1977**). *Signal Analysis* (McGraw-Hill, New York).
Schmidlin, D. J. (**2007a**). "Directionality of generalized acoustic sensors of arbitrary order," J. Acoust. Soc. Am. **121**, 3569–3578.
Schmidlin, D. J. (**2007b**). "A generalized function approach to implementing directional acoustic sensors," in Proceedings of the IEEE Underwater Acoustics Signal Processing Workshop, Alton Jones Campus, University of Rhode Island.
Silvia, M. T. (**2001**). "A theoretical and experimental investigation of acoustic dyadic sensors," SITTEL Technical Report No. TP-4, SITTEL Corporation, Ojai, CA.
Sloan, I. H., and Joe, S. (**1994**). *Lattice Methods for Multiple Integration* (Clarendon, Oxford).

Reconstruction of radiating sound fields using minimum energy method

Rolf Bader[a)]
Institute of Musicology, University of Hamburg, Neue Rabenstrasse 13, 20354 Hamburg, Germany

(Received 18 October 2008; revised 10 November 2009; accepted 13 November 2009)

A method for reconstructing a pressure field at the surface of a radiating body or source is presented using recording data of a microphone array. The radiation is assumed to consist of many spherical radiators, as microphone positions are present in the array. These monopoles are weighted using a parameter α, which broadens or narrows the overall radiation directivity as an effective and highly intuitive parameter of the radiation characteristics. A radiation matrix is built out of these weighted monopole radiators, and for different assumed values of α, a linear equation solver reconstructs the pressure field at the body's surface. It appears that from these many arbitrary reconstructions, the correct one minimizes the reconstruction energy. The method is tested, localizing the radiation points of a Balinese *suling* flute, reconstructing complex radiation from a *duff* frame drum, and determining the radiation directivity for the first seven modes of an Usbek tambourine. Stability in terms of measurement noise is demonstrated for the plain method, and additional highly effective algorithm is added for a noise level up to 0 dB. The stability of α in terms of minimal reconstruction energy is shown over the whole range of possible values for α. Additionally, the treatment of unwanted room reflections is discussed, still leading to satisfactory results in many cases. © *2010 Acoustical Society of America.* [DOI: 10.1121/1.3271416]

PACS number(s): 43.60.Fg, 43.60.Gk, 43.75.Kk [AH] Pages: 300–308

I. INTRODUCTION

For reconstructing the sound fields from radiation bodies using microphone arrays, different methods have been proposed. The near-field-acoustic holography[1–3] (NAH) and its equivalent, the statistical optimized nearfield-acoustic-holography[4] (SONAH), decompose the spatial signal for single frequencies at the microphone array into an angular spectrum. This is propagated using the phase information and the wave vector. The reconstruction of the pressure field at any point opposite to the plane array is performed by an inverse Fourier transform (IFT) of the propagated angular spectrum. This approach is widely used. Nevertheless the treatment of evanescent waves restricts this method with some applications. So, when dealing with fine structures of radiations, it is quite appropriate to cover the geometry with many microphones being close to each other. But then the wave vector components along the array plane are large. So the domain of evanescent waves begins at very low frequencies and the reconstruction is more or less only exponentially enlarging and not propagating the angular spectrum amplitudes. When dealing with musical instruments made of wood or bronze, nearly all waves radiated are evanescent waves, as the speed of these waves on the geometry is much faster than the speed of sound in air.

Other approaches try to rebuild the radiation by superposition of spherical harmonics as solutions of the spherical Helmholtz equation. The Helmholtz-least-squares (HELS) method for static[5–7] and transient sounds[8] uses an orthonormalization of these functions, according to the used geometry and microphone array placement. It then calculates the amplitudes of these functions from the recorded sound data and reconstructs the pressure field by inserting new spatial coordinates into the orthonormalized radiation functions. As all possible radiation patterns can, in theory, be represented by an infinite amount of spherical harmonics, this method is straightforward. Still in some cases, the arbitrary choice of a virtual radiation point behind the radiating surface can change the reconstruction. Ochmann,[9,10] therefore, suggested a multipole method, where several radiators are superposed and the optimal locations of these sources are investigated. Ochmann used the Galerkin method to estimate the amplitudes of the poles and showed that the methods using multipole radiators differ mainly by the weighting matrix used in the construction of the linear equation system. All these and related methods such as the full-field equations[10] or the equivalent-sphere method[11] assume virtual radiators within the structure to avoid the singularity at zero radiation distance in the solution of the Helmholtz equation with spherical coordinates (for a review, see Refs. 12).

The boundary-element-method (BEM) for source reconstructions has also been proposed.[14,15] Here, the boundary mesh of the geometry surface determines the pressure field in the geometry and therefore determines the radiation. Still, a complete model of the geometry needs to be built and meshed, and the solution of the pressure field is very time consuming.

Another problem with sound source reconstruction is the treatment of measurement noise and finite reconstruction accuracy, which can easily lead to divergence of any solver. The NAH and SONAH methods use an angular spectrum and, therefore, are able to reduce the problem of noise by cutting high frequency components with a trade-off of accuracy, as also explicitly proposed by the NAH method[16] using a low-pass filter for the angular spectrum. HELS and

[a)]Author to whom correspondence should be addressed. Electronic mail: r_bader@t-online.de

Ochmann methods use the Galerkin method when solving the linear equation system built of the radiation functions. The idea of Galerkin method was to assume orthonormality of the error vectors and so find an error minimum of the solution. So when setting the derivatives of the dependent variables to zero, the minimum error is found. Although the mathematical formulation of the idea of Galerkin method cannot be applied to the minimum energy method, the basic idea is the same as discussed below.

Like Ochmann, the method proposed here also uses many sources to model the radiation of arbitrary geometries. But here, different from the previous methods, as many radiators as microphone recording positions are used. Each radiator is a monopole with a radiation characteristic parameter α shaping directivity. α needs to be chosen in a way to minimize the reconstruction energy. This solution then needs to be the correct one (see below). Different from the HELS and from Ochmann's methods, the radiation points are assumed to be right on the geometry and not slightly behind it to avoid the local infinity, as here, the reconstruction algorithm does not make use of the radiation center point, so one could simply look at this method as a superposition of radiating points similar to Huygens waves leaving the geometry. Indeed, if the radiations of these points show up to be perfectly spherical, Huygens waves hold completely in this method. Nevertheless this is not normally the case and a directivity parameter needs to be estimated. So the minimum energy method does not only reconstruct the sound pressure field at the geometry, but also assumes an overall radiation directivity of that geometry.

As the method only uses monopoles, all dipole or higher order pole radiations are covered by the density of points on the geometry. If two neighboring points show a phase difference of π, a dipole is present. So for a frequency investigated, the density of points needs to cover each wavelength at least by two reconstruction points. As most sources of steel or wood have quite high wave speeds, which additionally grow with frequency because of dispersiveness of bending waves, achieving this density is easy. All methods using a finite number of microphones treat this issue in different ways. Here, it is simply solved by taking the wave number on the geometry into account and placing the reconstruction points accordingly.

As the measurements are also nearfield to cover all evanescent waves, it will appear that the reconstruction pressure at the geometry is then a "sharper" version of the recorded sound field in front of the geometry. This is a fundamental aspect of all reconstruction methods. Still here, by linearly changing a directivity parameter, this "sharpness" can be tuned by hand easily, as discussed in Sec. A1.

Microphone arrays have been used in musical acoustics already[17,16] and we will test the algorithm with a tambourine after we have used artificial sounds to test the reliability of this reconstruction method.

II. METHOD

The radiation from the geometry is assumed to be a superposition of monopoles with a directivity involved. Each point $\mathbf{x}_g^i$ right above the geometry or source has the pressure p_g^i for a single frequency f. Taking the phase changes in the radiating waves with wave vector k in air from the radiation points $\mathbf{x}_g^i$ on the geometry to microphone points $\mathbf{x}$ into account, the pressure at the microphones are

$$p_m^j = \sum_{i=1}^{N} p_g^i R_0^{ij} \tag{1}$$

with radiation matrix

$$R_0^{ij} = \frac{1}{\Gamma_0^{ij}} e^{\imath k r^{ij}}. \tag{2}$$

Here, $r^{ij} = |\mathbf{r}^{ij}|$ (its length) and Γ_0 is the intensity drop-off depending on the radiation angle discussed below.

The task is to calculate the pressure on the geometry p_g^i. Here, two reconstruction stages are performed. First, the recorded time series of the microphones are filtered for a single frequency f to obtain the complex pressure amplitudes p_m^j of this frequency at the single N microphone position vectors $\mathbf{x}_m^j$ with $1 \le i,j \le N$. Second, the minimum energy method is applied, taking the microphone positions and the used points of radiation on the geometry with position vectors $\mathbf{x}_g^i$ into account to finally get the spatial radiation pressures p_g^i for the used frequency.

As arbitrary shapes of geometries can be investigated, the formulation uses the normal vectors $\mathbf{n}^i$ at $\mathbf{x}_g^i$. Only the radiation at the front half plane of each radiation point is considered. So also radiations in all directions around the geometry can be investigated at once if only the different half planes of the radiation points cover a whole sphere. From the normal vector and the distance between the radiation points and the microphones, a matrix for the local directivity is calculated as

$$\beta^{ij} = \left| \frac{\mathbf{r}^{ij}}{\|\mathbf{r}^{ij}\|} \cdot \mathbf{n}^i \right|. \tag{3}$$

It is the absolute value of the inner product between the normal vector $\mathbf{n}^i$ with the normalized distance vector matrix $\mathbf{r}^{ij}$ between microphone j and radiating point i

$$\mathbf{r}^{ij} = \mathbf{x}_g^i - \mathbf{x}_m^j, \quad i,j = 1,2,3,\ldots,N. \tag{4}$$

So in case the microphone position $\mathbf{x}_m^j$ is right in front of the radiation point, then $\beta = 1$. In case it is in an angle of $\pm 90°$ to the radiation point, then $\beta = 0$.

We now define

$$\Gamma_0^{ij}(\alpha) = r^{ij}(1 + \alpha(1 - \beta^{ij})) \tag{5}$$

with $\alpha \ge 0$.

Γ_0 has the following properties.

- If $\alpha = 0$, then the radiation is that of a *perfect sphere* with an amplitude decay of $1/r$ independent of the radiation angle.
- If $\alpha > 0$ and $\beta^{ij} = 1$, so the radiation in the normal direction is investigated, then the *maximum radiation* for a fixed α and r^{ij} is achieved.
- If $\alpha > 0$ and $\beta^{ij} = 0$, so radiation in a direction orthonormal

to the normal direction is investigated, then the *minimum radiation* for a fixed α and r^{ij} is achieved.

Our task is to find an optimum value for α to minimize reconstruction energy so that α is fixed for all radiation positions. Only β changes with radiation angle from the radiation positions to the microphones where all of them have the same phase but different radiation strength depending on this angle. The maximum radiation is always in the normal direction; the minimum radiation is always in the directions orthonormal to the normal direction.

So the radiation matrix is

$$R_0^{ij}(\alpha) = e^{\imath kr}\frac{1}{r^{ij}(1+\alpha(1-\beta^{ij}))}. \tag{6}$$

When writing the pressures as vectors as $\mathbf{p}_g$ and $\mathbf{p}_m$ and the radiation matrix as $\mathbf{R}_0$, we can write Eq. (6) as

$$\mathbf{R}_0(\alpha)\mathbf{p}_g = \mathbf{p}_m. \tag{7}$$

This matrix is not composed out of vectors, which are orthonormal to one another. So we can get into trouble if the matrix is ill-conditioned and the condition number (see Ref. 18)

$$C = \frac{\lambda_{\max}}{\lambda_{\min}} \tag{8}$$

becomes very high. Mathematically, this happens if the vectors become more and more similar. Technically, this can be the case if we zoom into a structure very closely, as then the radiating spheres are coming closer together and so do look more and more the same. We will show in an example below with artificial time series how much we can zoom into a structure and resolve it with the example array also used for real measurements and how this is related to the condition number.

We can only solve this linear equation system if α is known, as elsewhere, we have an unclosed system with N equations and $N+1$ unknown dependent variables. Fortunately, we know that for the correct choice of α, the reconstruction energy we define as

$$E \sim \sum_i |\mathbf{p}_g^i|^2 = \min, \tag{9}$$

where $|\mathbf{p}_g^i|$ is the vector of the absolute amplitudes of the reconstruction pressure without phase information.

In the cases where α is over- or underestimated, the reconstructed pressure energy must be higher than for the correct solution. The reason for E, being a minimum at the correct reconstruction pressure field, follows from the reconstruction process. This can be clearly understood when we take a closer look at the following two cases.

- If α is higher than the correct value, the energy recorded by the microphones is assumed to come from a much narrower region. So this region has to supply all the energies recorded by the microphones. In the more widespread case, more radiation points contribute to the energy measured by the microphones. Therefore, each radiation point does not need to be so strong to represent the measured amplitudes alone. So the reconstruction energy must go down with smaller values of α.
- If α is smaller than the correct value, the influence of radiation points farther away from a microphone has more impact than physically true. In this case, the linear equation solver tries to balance this overestimation by an underestimation of nearly a radiation point. This continues for all points and so the solution diverges into unreasonable high values of the radiation amplitudes quite fast, and therefore, the radiation energy.

When reconstructing the pressure field in free space from the calculated $\mathbf{p}_g$, we now can also define *gamma* of Eq. (2) as

$$\Gamma^i(\alpha) = |\mathbf{x}_g^i - \mathbf{x}|(1+\alpha(1-|\|\mathbf{x}_g^i - \mathbf{x}\| \cdot \mathbf{n}^i|)). \tag{10}$$

As a linear equation system is solved, the amount of solution points $\mathbf{x}_g^i$ must be the same as the points of measured radiation. This is the amount of microphones when using single recordings. With stable radiations and changing microphone positions, an unlimited amount of reconstruction positions can be used.

To apply the method proposed here, the following constraints must be met.

- All radiating surfaces must be taken into consideration.
- The microphones should be placed near the geometry in a distance up to about 10 cm or even nearer. The closer the microphones are to the geometry, the better the condition of the radiation matrix R, as the relative lengths of neighboring radiation points for each microphone become larger.
- The density of reconstruction points must cover two points per wavelength of the waves on the geometry as other wide aliasing will occur.
- All reconstruction is performed for only the half-sphere in the normal direction of each radiation point separately.
- If very complicated geometries are investigated (i.e., violins from both sides) than only values of $\mathbf{R}_0$, for which an influence exit can be used. This does not influence the overall structure of the matrix or the solution.
- Complicated three-dimensional (3D) geometries might scatter waves around it, and therefore, some radiation points near the geometry in the air should be placed.

A. Noise

As discussed above for low noise level, the reconstruction error is small. The noise level may rise because of additional reflections of the room, where the measurements are performed, and so additional signals are present in the recording, which do not originate from the source investigated. The signal to noise ratio may also increase when dealing with a low signal level, where additional amplification noise of the recording system is present or when additional noisy sources are around. For the first case, the method can still treat such signals approximately. For the second case, an additional highly effective noise filter can be used.

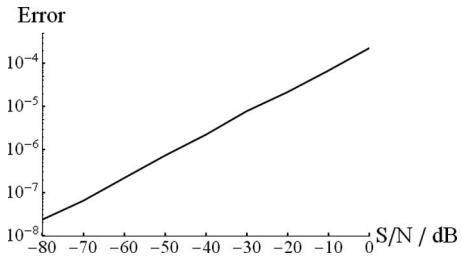


FIG. 1. The error for all noise levels as the mean difference between the original and the noise reconstructions.

1. Relaxation of source dependencies in case of presence of additional arbitrary sources

Additional reflections of walls or other rigid bodies around the source do disturb the recording. As discussed above, the method assumes the radiating body investigated to be the only source around. If such additive sources are part of the recording, the algorithm will try to explain them as part of the radiation pattern of the body under investigation. In some cases, this may lead to unphysical results of the reconstruction because the linear equation solver tries to balance these reflections between all radiation points and may therefore be forced to result in senseless reconstructions.

In such cases, the physical nature of the alpha parameter helps to still lead to reasonable reconstructions. As alpha determines the directivity of each radiation point, narrowing this directivity will ease such cases as additional sources disturbing the recording. This is simply so, because a larger value of alpha will narrow the spatial range of influences of neighboring radiation points to each microphone. So the reconstruction at each point is less dependent on other points around. This relaxation of dependencies always continues when increasing alpha from its minimum value. Indeed, when alpha is very high, the reconstruction will always be the recording itself. So when increasing alpha from its minimum, we are always sure to make the reconstruction more stable, and therefore, reducing the influence of other sources. Still this increase in the alpha value will also blur the reconstruction more and more and so we are no longer sure about the quantitative nature of our results. But as we have no information about additional sources of the room, part of the recording is arbitrary anyway and this information is not retrievable by any algorithm. Still in practice, this relaxation of neighboring point dependencies by increasing alpha has served very well in such cases, where additional reflections are not unreasonably strong and the nature of the radiation pattern can still clearly be seen with low additional source noise around.

The behavior of the method in such a case was tested with additional noise added to the signal. The recording of the duff frame drum was used, as discussed in Sec. III (see also Fig. 5). To the real recording, artificial noise was added in 10 dB steps from −80 to 0 dB with respect to the signal level. For each noise case, the reconstruction was performed, the reconstruction error compared to the unmodified measurement was calculated, and the value of α was determined. Figure 1 shows the error for all noise levels as the mean difference between the original and the noise reconstructions.

TABLE I. Value of α for reconstruction of the duff reconstruction as shown in Fig. 5 for different levels of artificially added noise.

Noise level (dB)	Alpha
No noise added	2.52
−80	2.52
−70	2.52
−60	2.52
−50	2.52
−40	2.52
−30	2.52
−20	2.53
−10	2.58
0	2.64

The error is very low and linearly increases up to the order of a power of –4 for the case of 0 dB noise level, where the noise level is as strong as the signal level. Visually, a difference between the reconstructions of the original recording and the noise cases can only be detected for the −20, −10, and 0 dB cases. Still, the difference is very low as expected from the error plot. The value of α changes for the noise cases, as shown in Table I. But here again, this change is only in the order of the second digit behind the comma, a precision which is not normally used with this method because the reconstruction does not change within this precision range in terms of a visual change in the reconstruction plot.

Still, if the value of α needs to be known with higher precision in Sec. II A 2, an additional method is presented to keep α constant over noise levels.

2. Sound segmentation method in case of additional recording noise

When additional noise is around, still the reconstruction can be obtained with high precision nearly independent of the noise level using a mean Fast-Fourier-Transform (FFT) algorithm, based on phase-locked sound segmentation. The signal $s^i(t)$ of pressure data recorded at any channel i is split into N segments

$$s_n^i(t) = s^i(t_n) \tag{11}$$

with

$$t_0 + \frac{nz}{f} \leq t_n < t_0 + \frac{(n+1)z}{f}, \tag{12}$$

$$n = 1,2,3,\ldots,N, \tag{13}$$

starting from an arbitrary time point t_0, which may be the beginning of a sound and taking the time of one period of the frequency f into consideration. The channel index i will be omitted now. The value of z needs to be an integer, determining the amount of periods of the frequency f in each segment to phase-lock the beginning of each segment to the phase of f at the beginning of the first segment. When performing FFTs over each segment, we obtain N complex pressure amplitudes $\bar{p}_n$, each containing the signal p_n, as well as the noise $\tilde{p}_n$ as follows:

TABLE II. Noise level, reconstruction α, and reconstruction error using the sound segmentation method for noise reduction.

Noise level (dB)	Alpha	Error (referenced to pure data)
−90	0.71	0.004 948
−72	0.71	0.004 845
−60	0.71	0.004 754
−48	0.71	0.004 718
−36	0.71	0.004 704
−24	0.71	0.004 697
−12	0.71	0.004 696
−6	0.71	0.004 697
0	0.71	0.004 691

$$\bar{p}_n = p_n + \tilde{p}_n. \tag{14}$$

As the noise is random, the $\tilde{p}_n$ have random phase relations, while the phases of the p_n signals are phase-locked by definition. So we can retrieve the original complex amplitude by averaging the amplitudes of all segments (for each channel) as

$$p = \frac{1}{N}\sum_{i=1}^{N} \bar{p}_n = \frac{1}{N}\sum_{i=1}^{N} p_n + \tilde{p}_n, \tag{15}$$

$$= \frac{1}{N}\sum_{i=1}^{N} p_n + \sum_{i=1}^{N} \tilde{p}_n = \frac{1}{N}\sum_{i=1}^{N} p_n + \varepsilon, \tag{16}$$

with

$$\varepsilon \to 0 \quad \text{for} \quad N \to \infty. \tag{17}$$

Table II shows results of this method using the fundamental mode of the tambourine discussed in Sec. II B. The original signal shows a low S/N ratio. To simulate more noisy cases to each recorded sound, a different random noise signal was added with increasing level resulting in nine cases with S/N ratio from −90 up to 0 dB. The sound segmentation was performed using $N=100$ and $z=1$, taking 100 segments with only one period of the fundamental frequency of $f=197$ Hz in each segment, starting from t_0, where the instrument was struck. For each noise level, the complex amplitude p was calculated using the method above and the reconstruction was performed, as shown in Sec. II A 1, determining the value of alpha for the minimum reconstruction energy. The reconstruction error was computed by comparing the results of the reconstruction using the original signals with each of the reconstructions using the signal with additional noise as

$$\text{err} = \frac{1}{121}\sum_{i=1}^{121} |p^i_{\text{org}} - p^i_{\text{noise}}|. \tag{18}$$

Table II shows the alpha values and the errors for noisy sounds of S/N ratios ranging from −90 to 0 dB. The error is within the range of 10^{-3}, even for the signal of 0 dB S/N, where the noise is as strong as the signal itself. The averaging over the different segments leads to a stabilization of the signal with respect to its complex amplitude. When inspecting the resulting reconstructions visually for the original and all noise signals, no differences could be seen comparing all reconstructions to one another. The value of alpha is stable throughout, too. Still, it differs from the value of the original sound by the small amount of 0.09. This may be caused by the different FFT integration length, where, for the original sound, a FFT length of 1 s was used. The length of the signal with the segregation method depends on the amount of periods within each segment z, the overall number of segments N, and the frequency f used. In the case shown in Table II, the integration length was $t_{\text{int}}=0.508$ s. Note that the signal needs to be stable in frequency over this integration length, as otherwise, the phase-locking of the segments will no longer hold. Still, the locking-phase itself is arbitrary and so we do not need to care about the time point t_0 where we start the analysis of the sound.

B. Hardware

The hardware used is a microphone array consisting of 128 electret MCE-4000 omnidirectional microphones with a frequency range 20 Hz–20 kHz and a S/N ratio of <58 dB. They are driven by a 9 V external battery voltage supply. The array is flat and equally spaced 11×11 plus an additional row of seven microphones. All microphones are spaced 3.9 cm apart in x- and y-direction. The amplification and digitalization equipment is a RME system of 128 microphone preamplifiers synchronized by a word clock and streamed by a Multichannel Audio Digital Interface (MADI) protocol. Two 64 channel MADI signals are streamed into a Personal Computer (PC) using two Peripheral Component Interconnect (PCI) cards, respectively, and again synchronized by a world clock. The signal is recorded using a Steinberg ASIO driver code written for this purpose. The used sample frequency is 48 kHz with 24 bits/sample for each channel. As the MADI protocol uses 32 bits/sample, a data flow of $128\times48\,000\times32$ bits/s is processed. These data are analyzed and statically and dynamically visualized by a WINDOWS C# code, including a Graphical User Interface (GUI). The recordings were done in the unechoic chamber at the Institute of Musicology in Hamburg.

III. RESULTS

To validate the method itself, six investigations show different aspects of the method. First, the spacing of microphones is discussed in terms of stability of the radiation matrix using an artificial signal. Then, the localization of radiation points is shown by measuring a tone partial of a Balinese suling flute. Although this flute is low in volume, the noise floor is audible and the solution is stable. Still, to discuss noise further, a second artificial signal is reconstructed with and without noise floor to show the error. Then, a complex radiation is shown of the octopole mode of a duff frame drum, comparing the measurement to a finite-element solution using the eigenvalue equation for the membrane. As drums are expected to have a more narrow radiation directivity with higher partials, an Usbek tambourine is shown to

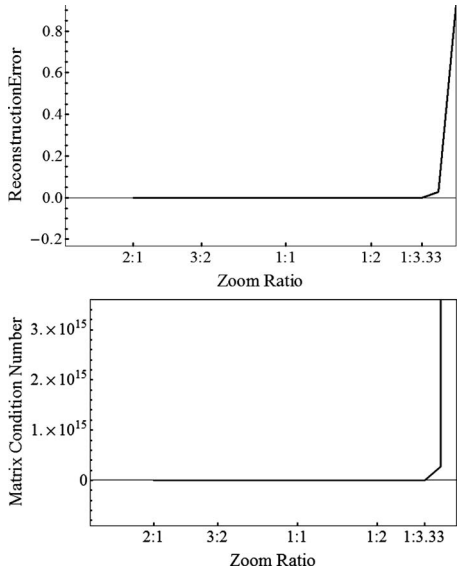


FIG. 2. Reconstruction error and condition number of radiation matrix $\mathbf{R}_0$ vs zoom factor of microphone array.

have increasing α values with higher mode shapes. Finally, the accuracy of the reconstruction value of α is shown again with artificial signals by comparing the input α to the α for minimal reconstruction energy.

A. Microphone spacing limits

As the radiation matrix does not consist of perfectly orthonormal vectors, the reconstruction precision is validated by varying the microphone spacing.

A radiating sphere of 2000 Hz is assumed in front of the microphone at position (6,6) in front of the array. The virtual array is spaced equivalent to the real array (see Sec. II B). To make the visualizations perfectly symmetric, only 121 points (11×11) were used and not all 128 points were possible with the hardware. The radiation at the microphone points is theoretically calculated, assuming a monopole radiation. The radiating strength was assumed to be 1 Pa, which is perfectly reconstructed as the correct minimum energy value for alpha was calculated by the algorithm.

To test different microphone spacings, we zoom in and out of the structure in a way that the equally spaced grid is preserved, but the grid constant is changed. A zoom of 1:1 results in the resolution of the radiating grid equal to the resolution of the array grid at 3.9 cm. If the zoom becomes larger, we look at larger radiating areas where the resolution also becomes larger, so, a zoom of 2:1 results in a grid constant of 2×3.9 cm^2 at the radiator. Here, the matrix is stable up to large zoom factors, as the condition number of the matrix is still small. This is because the vectors of the radiation matrix $\mathbf{R}_0$ differ from one another, still enough for the matrix to be well-conditioned.

If the zoom becomes smaller, we zoom into the radiating area and so a zoom of 1:2 results in a resolution of 3.9 cm/2 of the radiator. When zooming in, the condition number can reach high values and so a threshold of reconstruction precision is reached. Figure 2 shows the matrix condition number (Ref. 13) and the reconstruction error defined as $\text{err} = 1/\mathbf{N}\Sigma_n \mathbf{p_g} - \mathbf{p_m}$ for different zoom values. When zooming out, err is zero and the condition number is low. If we zoom into the radiation area up to a zoom of 1:3.33, we still have perfect reconstruction. Beyond this point, errors occur, which perfectly corresponds to the condition number reaching high values. So with this array, we are able to zoom in to a precision of about 1 cm and we are able to represent an area of about 10×10 cm^2 by 121 points. This is normally more than enough precision to cover the vibrations of usual plates and structures in the normal hearing range.

B. Localization of radiating blowing and tone holes of a Balinese suling flute

The Balinese suling flute is a bamboo end-blown flute, where the labium is at the top of the flute with an opposite radiation direction compared to the tone holes. It has a soft and gentle sound and exists in a small version of 25.5 cm in length (used here) and as a larger instrument played in *gamelan* suling orchestras. This smaller instrument is normally played solo or with the *gamelan gong* or *gamelan kebyar* using circular breathing and highly sophisticated tonal ornamentations. The instrument was collected during a field work of the author in Bali in 1999, where the building process was documented. This small suling is high pitched, and therefore, suitable for investigating higher frequency reconstructions.

As the suling radiates rather simply with a two-point radiation at the blowing hole and at the first open tone hole, it is used as an example for reconstructing point sources. The instrument was played 10 cm in front of the array, where it was held down, so the labium is at the very top of the reconstruction field and the first open radiating tone hole is in the middle of the array. As the labium is rectangle with 5 mm

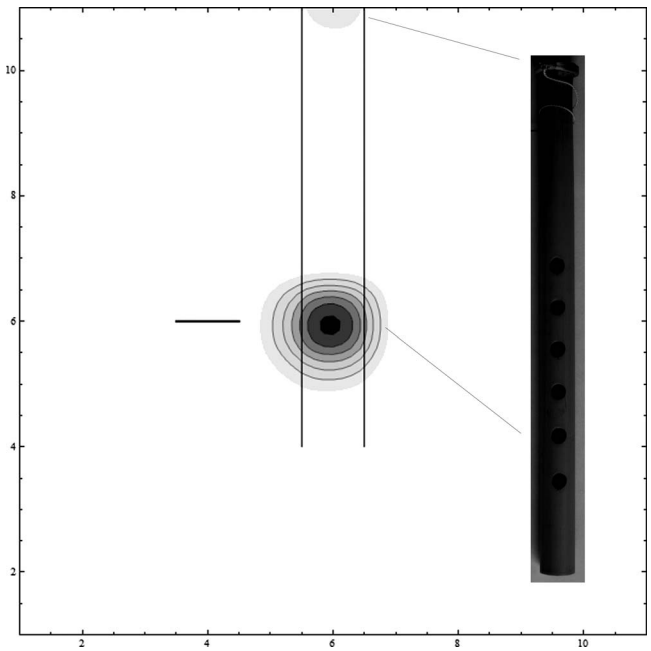


FIG. 3. Reconstruction of the radiation of a Balinese suling flute. The two vertical lines indicate the flute, and the vertical line shows the position of the first open tone hole. The blowing hole is at the top of the plot. A radiation directivity value of α=0.0 measured here indicates monopole radiation.

TABLE III. Three complex modes of the Usbek tambourine reconstructed pressure field on the tambourine surface, from top to bottom: 158, 252, and 874 Hz. From left to right: real part, imaginary part, absolute part and phase plot (normalized to maximum absolute amplitude each).

Position	(6,6)	(2,2)	(10,10)	(3,8)	(7,1)
Input amplitude	1	0.5	0.7	0.1	0.2
Reconstruction amplitude without noise	1	0.5	0.7	0.1	0.2
Reconstruction amplitude with noise (−58 dB)	1.007 18	0.502 29	0.693 74	0.104 44	0.188 13

side length and the tone hole is circular with a diameter of only 7 mm, we expect a nearly monopole radiation from both points.

Figure 3 shows the reconstructed radiation of the fourth partial with a frequency 2370 Hz of a played note with a fundamental of 590 Hz. The flute is graphically indicated by the two vertical black lines. The position of the first open sound hole is shown by the horizontal line in the middle of the plot. As expected, the radiation takes place not only at the tone hole but also at the blowing hole at the very top of the flute where the labium is placed. This radiation is much less here, as the labium is at the opposite side of the flute and only a small part of the radiation is present at the side of the tone hole. The presence of radiation from the blowing hole to the opposite direction of its opening may be caused either by the turbulent sound production process at the labium or by scattering and refraction of the labium radiation around the bamboo tube. The reconstruction value of alpha for minimum reconstruction energy is $\alpha=0.0$, which is a perfect monopole. The value for alpha only holds for the plane in the normal direction and is expected here because both holes are very small in diameter, and therefore, can freely radiate.

C. Five spherical sources with and without noise

The sound of the suling flute above has quite low loudness and so the measurement had a noise floor. This noise did not disturb the reconstruction considerably. Still, the algorithm should be tested for signals with and without a noise floor using an artificial signal. Five sources with different amplitudes are placed at the positions shown in Table III. The signals are calculated theoretically using a computer algorithm and the reconstructions were only done using the algorithm without noise correction. A signal to noise ratio of −58 dB was used. To test high frequencies, all sources radiate with 20 kHz. With recorded signals of 20 kHz, reconstruction would depend crucially on the appearance of any systematic error such as small deviations of the microphone spacing from the values used in the algorithm or small distance variations between the sources and the microphones.

In both cases, the reconstruction is perfect in terms of the amplitudes, as can be seen in Fig. 4. In the no-noise case, the reproduction is with machine size precision. In the noisy case, the error is very small and can be neglected. The noise case also shows additional low noisy radiation from all over the field. This is not consistent with the input, as no noise sources were located at any other point in the field. Still, as the noise is stochastic, there is no direct connection between the noise and the sources. In no way such a noise background of sources can be redirected to the sources alone. Still, this is a minor problem, as the purpose here was to correctly reconstruct the amplitudes above the noise floor, which was done successfully.

Also note that the machine size precision reconstruction of a computationally constructed signal is only of importance, as the reconstruction also depends on the choice of the minimum reconstruction energy. Here, the algorithm was indeed able to find the correct input value of alpha by determining the minimum energy point.

D. Octopole mode of duff frame drum

To investigate a complex radiation pattern, the duff, a frame drum from Pakistan, was investigated, taking the octopole radiation mode into consideration. The duff has a diameter of 45 cm and was investigated using a microphone spacing of 4.75 cm. It was struck once at its rim to get the typical sound of the instrument, which is nearly metallic, consisting of strong high partials. The listener is often able to distinguish them as single frequencies in the sound, making the pitch of the instrument arbitrary.

Figure 5 compares the reconstruction of the octopole mode of the duff of 513 Hz to a finite-element solution of the eigenvalue equation of a simple membrane

$$\frac{T}{\sigma}\left(\frac{\partial^2 u}{\partial x^2}+\frac{\partial^2 u}{\partial y^2}\right)=-\omega^2 u. \tag{19}$$

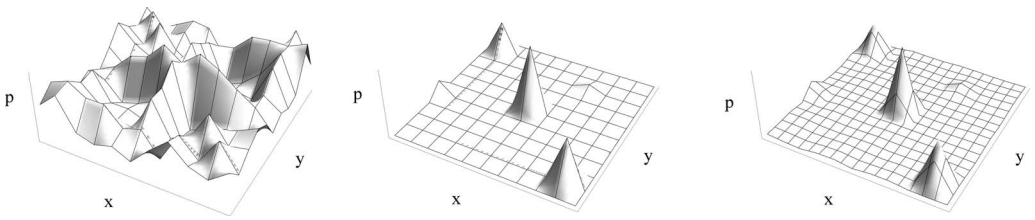


FIG. 4. Artificial measured pressure field at the virtual microphone array (left), reconstruction without noise (middle), and reconstruction with noise (−58 dB) (left) for five spherical point sources of a frequency of 20 kHz.

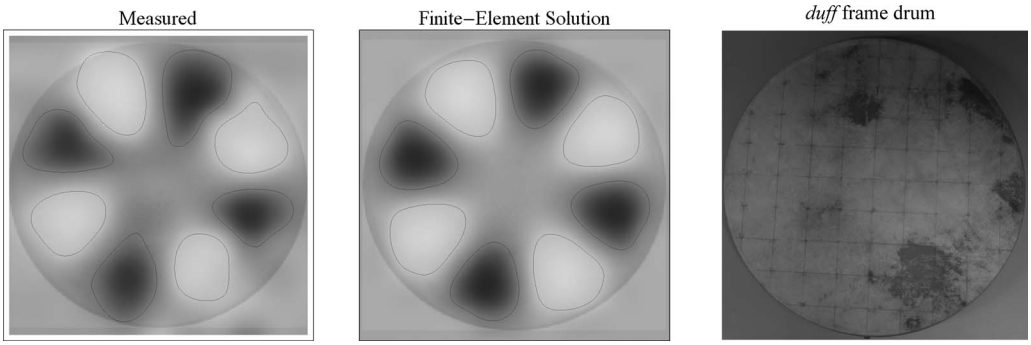


FIG. 5. Comparison between a reconstruction of the octopole mode of a duff frame drum with a finite-element solution using the eigenvalue equation for the membrane.

The displacement u for the octopole mode is plotted in Fig. 5 using homogeneous values for the tension T and the area density σ all over the membrane. The skin of the duff is not perfectly homogeneous and we cannot also expect the tension of the membrane to be stable all over. Here, both effects are neglected, still keeping in mind that they affect the solution to some extent.

E. Directivity values for tambourine mode radiations

Finally, the directivity value of alpha is tested with respect to realistically representing the overall width or narrowness of a radiation. Therefore, a tambourine used in the music of Usbekistan is used. It is round with a membrane diameter of 21 cm, a rim thickness of about 4 mm, and a depth of about 4 cm. This example is used to investigate the well studied geometry of a round membrane. Nevertheless, to make the investigation more interesting, this tambourine was also taken as it has been played for many years in an ensemble located in Samarkant. The membrane is made out of goat skin, while the rim is wooden. The tambourine clearly appears to be used and the part of the membrane, which was beaten normally with the hand or a stick, is clearly marked and presumably a bit thinner. So the membrane is not at all homogeneous as before with the duff drum, which makes the sound of the tambourine very beautiful, as much beating can be heard and the higher partials are enhanced during the sound continuing, depending on the place and kind of striking. Because of the inhomogeneity of the membrane, we would expect degenerated modes to appear.

The tambourine was placed 3 cm before the microphone array. It is held in parallel to the array plane in a way that the middle of the membrane was right before the middle microphone of the array, which is (6,6). It was stuck by a wooden stick to enhance the overtone structure even more. As with the duff, all analysis data presented here were taken from this one strike. There is no averaging in the data over several strikes. This is important for musical instrument investigations, as in this field, it is interesting to use methods, which are able, not only to distinguish one instrument from another, but also to make assumptions about the range and possibilities of single instruments.

Table IV shows the values of the radiation characteristic α for the first seven modes of the tambourine. The fundamental mode is a monopole. The following dipole, quadruple, and hexapole modes are all degenerated into two modes. As expected, the values of α increase with more complex radiation patterns, monopole, dipole, quadruple, and hexapole radiations. This was expected, as the more complex the radiation patterns, the more directional their radiation.

F. Accuracy of directivity value alpha

To also test the accuracy of the directivity value of alpha for the minimum energy reconstruction, a virtual plate as wide as the microphone array fixed at all sides and radiating with its eigenmode of two node lines with a frequency of 200 Hz was constructed again using a computer algorithm. Here, the radiation characteristic was changed by changing the value of α. The value of $\alpha=0$, which is a spherical radiation used in the previous cases, is no longer realistic for a plate showing a more narrow radiation characteristic. To test the validity of the minimum energy reconstruction method, values for α from $\alpha=0$ to $\alpha=1000$ were used. The plate radiation was altered according to the known value of α. Then the reconstruction was performed for different values of the reconstructing α, looking for the value of α with the least reconstruction energy. For a successful test, the reconstruction α must match the α that the plate radiation was constructed with.

Figure 6 shows the construction vs the reconstruction value of α in the range of $0 \leq \alpha \leq 1000$. As can be seen the α used for a radiation was perfectly found by the algorithm. For values around $\alpha \sim 100$, the precision is in the order of 10^{-8}, and for $\alpha \sim 1000$, the precision is still in the order of 10^{-8}. So indeed the minimum of the reconstruction energy is located right at the correct radiation value of α.

TABLE IV. Radiation characteristic values α for the modes of the Usbek tambourine. The more complex the mode shapes, the higher the value of α, which means the sharper the radiation directivity.

Frequency (Hz)	194	316	326	436	444	540	566
n-pole	Monopole	Dipole	Dipole	Quadruple	Quadruple	Hexapole	Hexapole
Alpha (radiation characteristic)	0.8	2.2	1.8	4.8	4.8	10.0	9.0

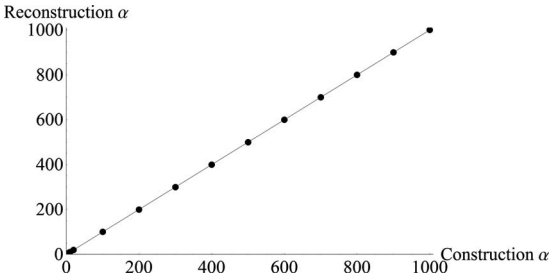


FIG. 6. Radiation characteristic value α vs reconstruction value of α with minimum energy for a radiating plate with an eigenmode shape of two node lines.

Also the phase differences at the plate are reconstructed according to their radiation. This is also true for the radiating sphere and the five-point case.

IV. CONCLUSION

The minimum energy method for reconstructing pressure fields recorded by a microphone array using the radiation characteristic value α is suitable for examining radiation patterns in detail. The reconstruction method works for single source localization, for multiple source localization, and for complex radiation patterns. It is suitable in the frequency range 200 Hz–20 kHz at least and shows much stability in the presence of measurement noise. With a given microphone array grid zooming into and out of a plane is possible and so the reconstruction surface can be investigated according to its geometry. Here also curved structures can be taken into consideration where the normal vectors at the reconstruction points are known. Still there is a threshold for zooming, in which is around 1:3.33 to 1:4 for the regular array used in this investigation, but this seems to be more than enough for reconstructing complex radiations within the frequency range of hearing. The radiation characteristic needs to be estimated for reconstruction, which follows the fact that this value is correct when the reconstruction energy is at a minimum. With artificial sources the reconstruction of α is within a precision of at least 10^{-7} all over the range, from a perfect spherical radiator with $\alpha=0$ to a pure directional radiation of $\alpha\sim1000$. When applying the radiation characteristic method to a tambourine, it appears that α increases with the complexity of the mode shapes as expected. More complex radiation patterns show a more narrow radiation directivity, which is reflected with this value.

So the minimum energy method with nearfield recordings can be used when complex geometry patterns need to be investigated. As the nearfield recording already shows much of the pattern to be reconstructed, this makes measurements very sure and reconstruction very precise. Further investigations need to consider time-dependent phenomena and try to reconstruct the time-dependent radiation pressure at the radiating surface.

[1]E. G. Williams, J. D. Maynard, and E. Skurdzyk, "Sound source reconstructions using a microphone array," J. Acoust. Soc. Am. **6**, 341–344 (1980).
[2]J. D. Maynard, E. G. Williams, and Y. Lee, "Nearfield acoustic holography: I. Theory of generalized holography and the development of NAH," J. Acoust. Soc. Am. **78**, 1395–1413 (1985).
[3]E. G. Williams, B. H. Houston, and P. C. Herdic, "Fast Fourier transform and singular value decomposition formulations for patch nearfield acoustical holography," J. Acoust. Soc. Am. **114**, 1322–1333 (2003).
[4]J. Prager, "Approximate reconstruction of sound fields close to the source surface using spherical nearfield acoustical holography," J. Acoust. Soc. Am. **122**, 2067–2073 (2007).
[5]Z. Wang and S. F. Wu, "Helmholtz equation-least-squares method for reconstructing the acoustic pressure field," J. Acoust. Soc. Am. **102**, 2020–2032 (1997).
[6]N. Rayess and S. F. Wu, "Experimental validations of the HELS method for reconstructing acoustic radiation from a complex vibrating structure," J. Acoust. Soc. Am. **107**, 2955–2964 (2000).
[7]S. F. Wu, "On reconstruction of acoustic pressure fields using the Helmholtz equation least squares method," J. Acoust. Soc. Am. **107**, 2511–2522 (2000).
[8]S. F. Wu, H. Lu, and M. S. Bajwa, "Reconstruction of transient acoustic radiation from a sphere," J. Acoust. Soc. Am. **117**, 2065–2077 (2005).
[9]M. Ochmann, "The source simulation technique for acoustic radiation problems," Acustica **81**, 512–527 (1995).
[10]M. Ochmann, "The full-field equations for acoustic radiation and scattering," J. Acoust. Soc. Am. **105**, 2574–2584 (1999).
[11]L. Bouchet and Th. Loyau, "Calculation of acoustic radiation using equivalent-sphere methods," J. Acoust. Soc. Am. **107**, 2387–2397 (2000).
[12]M. B. S. Magalhães and R. A. Tenenbaum, "Sound source reconstruction techniques: A review of their evolution and new trends," Acta. Acust. Acust. **90**, 199–220 (2004).
[13]G. B. Arfken and H. J. Weber, *Mathematical Methods for Physicists* (Elsevier, Amsterdam, 2005).
[14]W. A. Veronesi and J. D. Maynard, "Digital holographic reconstruction of sources with arbitrarily shaped surfaces," J. Acoust. Soc. Am. **85**, 588–598 (1989).
[15]M. R. Bai, "Application of BEM (boundary element method)-based acoustic holography to radiation analysis of sound sources with arbitrarily shaped geometries," J. Acoust. Soc. Am. **92**, 533–549 (1992).
[16]R. Scholte, N. B. Roozen, and I. Lopez Arteaga, "Regularization in PNAH by means of L-curve," Proceedings of the Forum Acusticum Budapest (2005), pp. 2579–2583.
[17]L. M. Wang and C. B. Burroughs, "Acoustic radiation from bowed violins," J. Acoust. Soc. Am. **110**, 543–555 (2001).
[18]G. Bissinger, E. G. Williams, and N. Valdivia, "Violin f-hole contribution to far-field radiation via patch near-field acoustical holography," J. Acoust. Soc. Am. **121**, 3899–3906 (2007).

Short pulse multi-frequency phase-based time delay estimation

Said Assous,[a] Clare Hopper, and Mike Lovell
Department of Geology, Ultrasound Research Laboratory, University of Leicester, Leicester LE1 7RH, United Kingdom

Dave Gunn, Peter Jackson, and John Rees
Ultrasound Research Laboratory British Geological Survey, Keyworth NG12 5GG, United Kingdom

(Received 6 January 2009; revised 9 October 2009; accepted 9 October 2009)

An approach for time delay estimation, based on phase difference detection, is presented. A multiple-frequency short continuous wave pulse is used to solve the well-known phase ambiguity problem when the maximum distance exceeds a full wavelength. Within an unambiguous range defined with the lowest frequency difference between components, the corresponding phase difference is unique and any distance within this range can be determined. Phase differences between higher frequency components are used to achieve a finer resolution. The concept will be presented and the effectiveness of the approach will be investigated through theoretical and practical examples. The method will be validated using underwater acoustic measurements, simulating noisy environments, demonstrating resolutions better than a 50th of a wavelength, even in the presence of high levels (−5 dB) of additive Gaussian noise. Furthermore, the algorithm is simple to use and can be easily implemented, being based on phase detection using the discrete Fourier transform.
 [DOI: 10.1121/1.3263602]

PACS number(s): 43.60.Jn, 43.60.Ac, 43.60.Bf, 43.60.Qv [EJS] Pages: 309–315

I. INTRODUCTION

Time delay estimation is a fundamental step in source localization or beamforming applications and has attracted considerable research attention over the past few decades in many fields including radar, sonar, seismology, geophysics, ultrasonics, communication, and medical imaging. Various techniques are reported in the literature[1–5] and a state of the art review can be found in Ref. 6 which concerns critical techniques, limitations, and recent advances that have significantly improved the performance of time delay estimation in adverse environments. These techniques can be classified into two broad categories: correlator-based approaches and system-identification-based techniques. Both categories can be implemented using two or more sensors; in general, more sensors lead to increase robustness due to greater redundancy. When the time delay is not an integral multiple of the sampling rate, however, it is necessary to either increase the sampling rate or use interpolation.[6] In this paper, we present a new approach for time delay estimation, based on the received signal phase information. This avoids barriers encountered by alternative approaches based on phase information, which are limited by the need to use a reference signal, usually provided by a coherent local oscillator.[7] Ambiguities in such phase measurement, caused by the inability to count integer number of cycles (wavelengths), are resolved using the Chinese remainder theorem (CRT) taken from number theory, where wavelength selection is based on pair-wise relatively prime wavelengths.[7,8] However, the CRT is not entirely robust, in the sense that small errors in its remainders may induce a large error in the determined integer.[9,10] Another phase-based measurement approach, adopted to ensure accurate positioning of commercial robots, uses two or more frequencies in a decade scale in a transmitted signal. In this, the phase shift of the received signal with respect to the transmitted signal is exploited for ranging.[11,12] However, this approach is valid only when the maximum pathlength/displacement is less than one wavelength, otherwise a phase ambiguity will appear.

The time delay estimation approach proposed here is based on the use of local phase differences between specific frequency components of a short continuous wave (cw) received signal pulse. Within an unambiguous range defined by the lowest frequency difference between components, the corresponding phase difference is unique and any distance within this range can be determined. Phase differences between higher frequency components are used to achieve a finer resolution. Thus our approach overcomes the need to cross-correlate the received signal with either a reference signal or the transmitted signal. Unlike conventional methods, this approach is not limited by phase ambiguity; therefore most practical situations, where the range to be determined is beyond one wavelength, can be accommodated.

II. METHODOLOGY

A. Concept

The inspiration for development of the technique comes from the observation that bats have been shown to have exceptional resolution with regard to target detection when searching during flight.[13–15] Au and Simmons,[15] somewhat controversially, concluded that bats using pulses with a center frequency of about 80 kHz (40 kHz bandwidth) can achieve a distance resolution in air approaching 20 μm. At

[a]Author to whom correspondence should be addressed. Electronic mail: sa251@le.ac.uk.

this frequency, and using $\lambda/2$ as the guide for resolution, we see that the resolution reported by Au and Simmons is about 200 times better than that predicted conventionally. We sought to develop a methodology which would allow us to approach a resolution comparable to that achieved by such creatures which is still better than any man-made system.

Consider an acoustic pulse containing a single frequency component f_1 with an initial zero phase offset. This pulse is emitted through the medium, impinges on a target, is reflected and returns. The signal is captured and its phase measured relative to the transmitted pulse. Given this situation, we cannot estimate the distance to and from an object greater than one wavelength away (hence, usually, we would estimate the time of arrival of the pulse and assume a value for the velocity of sound in the medium to estimate the distance to the target).

For simplicity, assume the pulse contains a single cycle of frequency f_1 of wavelength λ_1. The distance D to the target can be expressed as

$$D = n_1\lambda_1 + r_1, \tag{1}$$

where $\lambda_1 = v/f_1$, n_1 is an integer, r_1 is a fraction of the wavelength λ_1, and v is the speed of sound in the medium.

r_1 can be expressed as follows:

$$r_1 = \lambda_1 \times \frac{\phi_1}{360}, \tag{2}$$

where ϕ_1 is the residual phase angle in degrees. Combining Eqs. (1) and (2) and rearranging

$$D = n_1\lambda_1 + \lambda_1\frac{\phi_1}{360},$$

$$= n_1\frac{v}{f_1} + \frac{\phi_1}{360}\frac{v}{f_1},$$

$$D = \frac{v}{f_1}\left(n_1 + \frac{\phi_1}{360}\right). \tag{3}$$

If we transmit a second frequency component f_2 within the same pulse, then it will also have associated with it a wavelength λ_2 and a residual phase ϕ_2; similarly,

$$D = \frac{v}{f_2}\left(n_2 + \frac{\phi_2}{360}\right). \tag{4}$$

Equations (3) and (4) can be solved by finding Eq. (4) − Eq. (3) $\times(\lambda_2/\lambda_1)$ and rearranged to give

$$D = \left(\frac{\lambda_1\lambda_2}{\lambda_1 - \lambda_2}\right)\left((n_2 - n_1) + \frac{(\phi_2 - \phi_1)}{360}\right). \tag{5}$$

Using $v = f \times \lambda$, we obtain

$$D = \frac{v}{\Delta f}\left(\Delta n + \frac{\Delta\phi}{360}\right), \tag{6}$$

TABLE I. Parameters used in example 1.

R (mm)	v (mm/μs)	f_1 (kHz)	λ_2 (mm)	n_1	f_2 (kHz)	λ_2 (mm)	n_2
1500	1.50	200	7.50	200	201	7.462	201

where $\Delta f = f_2 - f_1$, $\Delta n = n_2 - n_1$, and $\Delta\phi = \phi_2 - \phi_1$.

Knowing $D = v \times t$, we deduce the time delay t as

$$t = \frac{1}{\Delta f}\left(\Delta n + \frac{\Delta\phi}{360}\right). \tag{7}$$

If we impose the condition that $\Delta n \leqslant 1$, then Eq. (7) can be solved. This restriction on Δn is imposed as follows.

- A distance D is selected within which we require an unambiguous range measurement.
- Select a frequency f_1 within the bandwidth of the system, and its corresponding wavelength λ_1 [from Eq. (1)].
- Similarly, using Eq. (1), select frequency f_2 with its corresponding wavelength λ_2 such that the number of cycles is $n_2 = n_1 + 1$.

Considering Eq. (6), the maximum range is achieved by this approach when $\Delta n = 1$;

$$R = \frac{v}{\Delta f}. \tag{8}$$

Therefore, R is the maximum unambiguous range that can be achieved using the two frequencies f_1 and f_2 as described above. As the phase differences $\Delta\phi$ will be unique within this range R, any distance within this range can be determined unambiguously.

B. Example

We demonstrate in this example how we could measure a range by using two frequencies and their related phase differences.

Consider an unambiguous range R; two frequencies f_1 and f_2 comprising integer number of cycles n_1, n_2 of wavelength λ_1, λ_2, respectively, as shown in Table I. Within this range R, consider a distance to target $d = 1000.1234$ mm we wish to estimate. Assume that a short cw pulse comprising these two frequencies f_1 and f_2 is sent toward the target at the distance d.

Using f_1, f_2 and Eqs. (1) and (6) for this distance d gives an integer number $n_1 = 133$ and a residual fraction of cycle of $r_1 = 0.349\,786$ corresponding to a residual phase of $\phi_1 = 125.923°$. Similarly for the frequency f_2, we find the residual phase $\phi_2 = 5.953°$ corresponding to $n_2 = 134$ and $r_2 = 0.016\,535\,6$. Thus, $\Delta\phi = \phi_2 - \phi_1 = -119.970°$.

We use this value in the formula given in Eq. (6), since $\Delta\phi$ is negative; this means $\Delta n = 1$. Using v from Table I, $\Delta f = 1$ kHz, and substituting into Eq. (7) give a first estimate of the range $\hat{d}_{f_1 f_2} = 1000.1233$ mm (caret is used here to mean an estimate). The unambiguous range R [Eq. (8)] is independent of the frequencies used depending only on the difference in frequency Δf.

TABLE II. Parameters used in example 2.

R (mm)	v (mm/μs)	f_3 (kHz)	λ_3 (mm)	n_3	f_4 (kHz)	λ_4 (mm)	n_4
1500	1.5	210	7.142	210	300	5.0	300

Note that in practice such resolution may not be achievable and limitations must be considered. For example, if the accuracy of estimating the phase is within $\pm 0.5°$, then the phases in the example above become $\phi_1 = 126.0$ and $\phi_2 = 6.0$, giving $d = 1000.0$ mm implying an error of 0.1234 mm. To get a finer resolution, we have to add other frequency components.

C. Using multiple frequencies through a "Vernier approach"

We demonstrate in this example how we could obtain a finer resolution in a range by using more than two frequencies and their related phase differences.

In Eq. (6), we imposed the condition that $\Delta n \leqslant 1$. The values of frequencies f_1 and f_2 were chosen to ensure this condition and to obtain a first estimate of the distance $\hat{d}_{f_1 f_2}$, and an estimate of the time delay $\hat{t}_{f_1 f_2}$ [Eq. (7)].

Introducing a third frequency f_3 as shown in Table II, such that $f_3 - f_1 = 10 \times (f_2 - f_1)$; f_2 differs from f_1 by 1 kHz and f_3 differs from f_1 by 10 kHz.

Again from Eq. (1), for f_3 and $d = 1000.1234$ mm, $\phi_3 = 6.219°$, which we would measure as 6.5°. Thus, $\Delta\phi_{13} = \phi_3 - \phi_1 = -119.5°$. We add 360° to give 240.5°. However, Δn_{13} between frequencies f_1 and f_3 is now 7 (in fact, 6 since we have already added in 360° to make the phase difference positive).

Using Eq. (6) with $\Delta\phi_{13}$ and different values of Δn_{13} (0–6) to get different distance estimate $\hat{d}^k_{f_1 f_3}$, where $k = 0, \ldots, 6$, in this case.

Applying Eq. (6) recursively for $\Delta n_{13} = 0, \ldots, 6$ to calculate $\hat{d}^k_{f_1 f_3}|_{k=0,6}$ selecting $\hat{d}^k_{f_1 f_3}$ closest in value to $\hat{d}_{f_1 f_2}$ as the optimum value $\hat{d}_{f_1 f_3} = 1000.2083$ mm$|_{k=6}$. Hence, a new best time delay estimate $\hat{t}_{f_1 f_3}|_{k=6}$.

Note that the new best estimate distance is with an error of 0.0849 mm.

If a fourth frequency f_4 is introduced, as shown in Table II, such that the $\Delta f_{14} = 100.0$ kHz, using Eq. (1) again gives $\phi_4 = 8.8848°$ which we measure as 9°. Thus, $\Delta\phi_{14} = \phi_4 - \phi_1 = -117°$ which gives $\Delta\phi_{14} = 249.5$ after adding 360°. Note that $\Delta n_{14} = 66$ in this case.

Similarly, select the estimate $\hat{d}_{f_1 f_4}$ $(\Delta n = 0, 1, 2, \ldots, 66)$ closest in value to $\hat{d}_{f_1 f_3}$. This occurs at $\Delta n_{14} = 66$ giving $\hat{d}_{f_1 f_4} = 1000.125$ mm. Taking this as the best estimate, the final error is 0.0016 mm = 1.6 μm.

Note that the frequencies need not be in decadal scale variation; however, they should be selected within the bandwidth of the system. All the selected frequencies should be a multiple of the lowest difference ensuring an integer number of cycles; in this way, we ensure that the phase calculated by the discrete Fourier transform (DFT) will be correct (i.e., corresponding exactly to the frequency bins). Thus, this example is reminiscent of the operation of a Vernier gauge as follows:

- $\Delta\phi_{12}$, related to the frequencies f_1 and f_2, gives the first estimate of the distance $\hat{d}_{f_1 f_2}$, hence $\hat{t}_{f_1 f_2}$.
- A higher frequency f_3 is then used (decade difference) to measure the same range but with a finer resolution. So a more accurate approximation to the measured range is obtained $\hat{d}_{f_1 f_3}$.
- Similarly, the measured range $\hat{d}_{f_1 f_4}$ corresponding to $\Delta\phi_{14}$ within f_1 and f_4, will give the ultimate estimate of the measured range d.
- Consequently, the maximum distance and the minimum resolution achieved are determined by the choice of the frequencies f_1, f_2, f_3, and f_4.

An example algorithm calculating the phase at each frequency and outlining the iterative phase difference measurements is given in the Appendix.

D. Phase offset measurement calibration

In the numerical example above, it is assumed that all phases are accurately transmitted and received, with no phase error on transmission or reception, and that all frequencies have zero phase offset with respect to each other. In practice this is almost certainly not the case and phase offsets between frequencies should be accounted for as discussed below.

Considering two frequencies $f_1 = 200.0$ kHz and $f_2 = 201$ kHz, and assuming the speed of sound in water ($v = 1.5$ mm/μs), from Eq. (8), the unambiguous range $R = 1500$ mm and Δn is 0 or 1. Considering the above

$$t = \frac{D}{v} = \frac{n + \phi/360}{f}. \tag{9}$$

Consider two distances d_1, d_2 corresponding to two "times" t_1 and t_2 such that the number of cycles n is the same for both frequencies over these distances, and assuming the phase measured includes a phase offset for that frequency. As an example, suppose the unknown phase offset for f_1 is 10°, for f_2 is 30°, and assume $d_1 = 100$ mm.

From Eq. (9), the term $(n_1 + \phi_1/360)$ would be calculated as 13.33 cycles, where $\phi_1 = 120°$. The "measured" $\phi_1 = 120 + 10 = 130°$ ($\phi_{1\text{measured}} = \phi_{1\text{distance}} + \phi_{1\text{offset}}$).

Similarly, for f_2 we obtain 13.40 cycles, where $\phi_2 = 144°$. The measured $\phi_2 = 144 + 30 = 174°$; from Eq. (7), $t_1 = 12.22$ μs. The actual time should be 6.66 μs.

Assume a second distance $d_2 = 200$ mm. Using Eq. (9), for f_1 we obtain 26.66 cycles, which gives $\phi_1 = 240°$. The measured $\phi_1 = 240 + 10 = 250°$. Similarly, for f_2 we obtain 26.80 cycles, which gives $\phi_2 = 288°$. The measured $\phi_2 = 288 + 30 = 318°$. Thus, using Eq. (7), $t_2 = 18.88$ μs. The actual time should be 13.33 μs.

A linear relationship could be deduced between t and d as follows:

$$t = m \times d + c, \tag{10}$$

where the slope $m=1/1.5=0.66$ is the speed of sound measured as 1 mm per 0.66 μs or $1/0.66=1.5$ mm/μs. The intercept $c=5.55$ μs is a measure of the relative phase between f_1 and f_2. Since $\Delta f_{12}=10$ kHz, 1 cycle is 100 μs long; consequently, the offset of 5.55 μs $\equiv 360\times(5.55/100)=20°$, which is equal to the relative phase (30–10) between the two frequencies. If we had known the phase offset between the two frequencies (20°), then in the calculation of times we would have obtained for t_1 a new phase difference of $(174-130-20)=24°$ giving a time for $t_1=6.66$ μs.

Similarly, for t_2 we obtain a new phase difference of $(318-250-20)=48°$ giving a time for $t_2=13.33$ μs. Both t_1 and t_2 are now correct. Note the following.

- If we assumed d_1 was 100 mm but it was actually, say, 120 mm and that d_2 was 200 mm but it was actually 220 mm, then we obtain the phase offset as 15.2°. The slope of Eq. (10) above, however, is unaffected. For example, such uncertainty may arise if the distance traveled by the wave within the transducers is not taken into consideration.
- If the temperature changes, so v changes; this changes the slope of Eq. (10) but not the time intercept or the phase offset. For example, if $v=1.6$ mm/μs, then equation Eq. (10) becomes $t=(1/1.6)\times d+5.55=0.625\times d+5.55$.

III. APPLICATION

A. Experiment

To demonstrate this approach, a series of measurements was performed in a water tank measuring $1530\times1380\times1000$ mm^3. Two broadband ultrasonic transducers (Alba Ltd., Glasgow, UK), having a wide bandwidth with a center frequency between 100 and 130 kHz, operate as both transmitters (Tx) and receivers (Rx) of ultrasound with a beam width of around 10° at the center frequency, where −3 dB bandwidth is 99 kHz (72–171 kHz). These were mounted on a trolley, movable in the X-Y directions. Linear encoders (Newall Ltd., Leicester, UK) were used to measure displacement of the rails in the x-direction and a software provided readouts of transducer positions. The temperature in the tank was measured by thermocouples calibrated using a quartz thermometers which are traceable to national standard. They were positioned on the four sides panels of the tank and recorded 19.80 ± 0.05 °C during the experiment. The experiment setup is shown in Fig. 1. The transmitter was driven by a 20 V peak-to-peak waveform consisting of short cw pulse comprising four frequencies (70, 71, 80, and 170 kHz). A modular system comprising a 16-bit arbitrary waveform generator (Ztec ZT530PXI) and a 16-bit digital storage oscilloscope (Ztec ZT410PXI) was used for transmit/receive process. Software was used to control signal transmission and acquisition. Distances between Tx and Rx of 305.800, 345.778, 481.128, 535.131, 535.252, 624.515, 759.182, and 862.887 mm were selected to be within an unambiguous range of $R\approx1500$ mm [Eq. (8)], as set by the linear encoders. A set of ten signals, noise free signal and signals with different signal to Gaussian noise ratios (SNR=−20,−10, −5,−2,0,2,5,10,20 dB), was transmitted at each distance

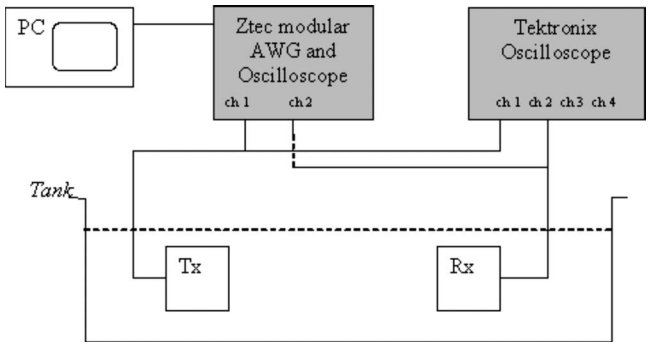


FIG. 1. Schematic diagram of the experimental setup.

described above. Note that the added Gaussian noise was generated by software and added to the original signal in each case. Before transmitting, each signal was multiplied by a Tukey window (cosine-tapered window with a 0.2 taper ratio) to reduce the "turn on" and "turn off" transients of the transducers. At each distance, three repeat pulses were transmitted and received for each SNR and each distance. Furthermore, 60 repeat pulses were transmitted and received while keeping the distance constant at 862.887 mm to assess the repeatability of the system (see Fig. 3). The sampling frequency F_s was set to 10 MHz, giving a number of samples $N=20000$ and a 2 ms pulse length. A DFT was then applied to the received pulses to obtain the magnitude and phase information for each of the four frequency components, using a window of $[N/2+1:N]$ for each received signal. This gave a resolution $F_s/N/2=1$ kHz which was consistent with the smallest step between the four frequencies comprised in the pulse.

Figure 2(a) shows the noise free transmitted (top) and received (bottom) signals when Tx and Rx were 305.800 mm apart, while Fig. 2(b) shows transmitted (top) and received (bottom) signal in the greatest noise case SNR (−20 dB). Note the DFT reports phase with respect to cosine, whereas sine waves were used in this experiment. Sine waves are returned with a phase of −90° relative to cosine waves by the DFT. This was not an issue, since relative phase differences were used. Using the phase for each distance obtained by the DFT, the phase-based time delay algorithm given in the Appendix was applied to obtain the corresponding estimated times for each phase difference $\Delta\phi_{12}$, $\Delta\phi_{13}$, $\Delta\phi_{14}$, $\Delta\phi_{23}$, $\Delta\phi_{24}$, and $\Delta\phi_{34}$, for the pairs f_1f_2, f_1f_3, f_1f_4, f_2f_3, f_2f_4, and f_3f_4, respectively.

Using a simple calculation of the first estimate by $t_{12}=\Delta\phi_{12}/(f_2-f_1)$ as a first estimate using Eq. (7) gave corresponding estimated times $\hat{t}_{12}$, $\hat{t}_{13}$, $\hat{t}_{14}$, $\hat{t}_{23}$, $\hat{t}_{24}$, and $\hat{t}_{34}$, respectively. For each distance, $\hat{t}_{14}$ should be the best estimate (i.e., the greatest Δf). The sensitivity of the algorithm to noise was reduced by the following.

- Limitation of the noise beyond the sensitivity of the transducer. The transducer act with its bandwidth as a bandpass filter.
- Signals were averaged 32 times during acquisition. When acquiring the signal with averaging, this reduces the noise level.
- The DFT operates on sine waves preferentially, even in the

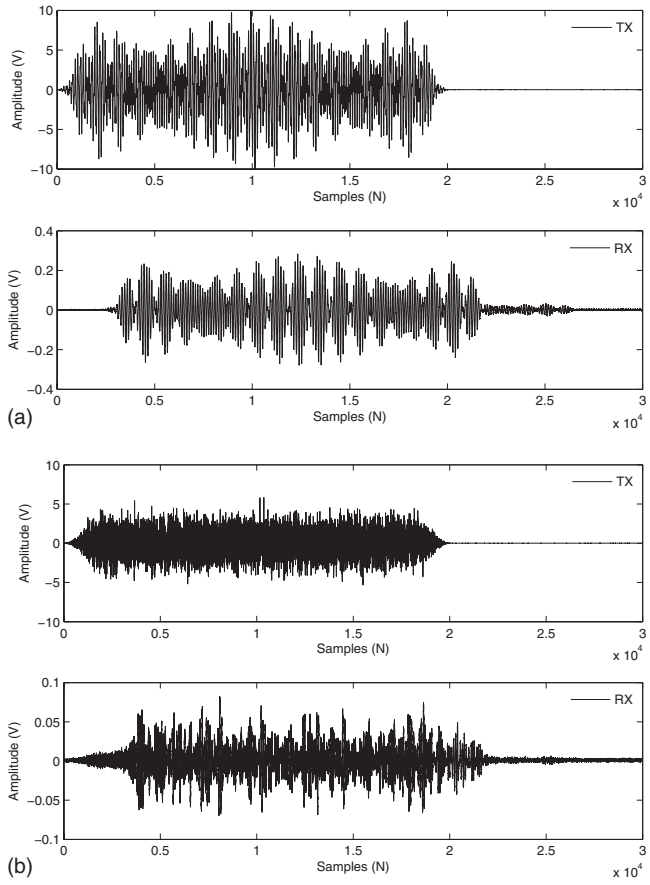


FIG. 2. Examples of transmitted and received short four components cw signals. Without (upper) and with (lower) added noise (−20 dB).

presence of noise. As the transmitted frequencies are known, when applying the DFT for the received signal, we will look just for those frequencies; hence any other frequencies are seen as noise and will not affect the frequencies of interest.

- Gradient estimation using a least squares fitting is noise tolerant. When doing the least squares estimation, any outlier which is basically noise will be removed.

B. Results and discussion

The best estimated time delays $\hat{t}_{14}$ for the eight different distances and using different signal-to-noise ratios are shown in Table III. For each distance, ten estimations were given corresponding to different signal-to-noise ratios. The column 2 corresponding to $\hat{t}_{14,Org}$, represents the estimated times when no noise was added to the signal. Using $\hat{t}_{14,Org}$ as best estimate time reference related to noise free signal, we can see that the technique became more robust to noise when the signal-to-noise ratio become greater than 2 dB. The variation is within 50, 30, 20, and 10 μs when the SNRs are 2, 5, 10, and 20 dB, respectively.

For the two rows in Table III corresponding to the distances 535.131 and 535.252 mm, where the displacement is about 120 μm, the estimated time for this displacement is about 120 ± 20 ns when the SNR is greater than −5 dB. As the sampling frequency used in this experiment was 10 MHz, using the conventional method (e.g., cross-correlation) to estimate the time delay, we would not get a best time resolution of less than $1/10^7 = 100$ ns, which corresponds to a displacement of 150 μm, when assuming a speed of sound of 1500 m/s. Hence, such displacement could not be measured without interpolation or oversampling. Moreover, if the cross-correlation approach was used to estimate the time delay in this experiment, all the measurements in Table III would be an integer of the sampling rate; in this case, 100 ns, as the technique is sampling-frequency dependent.

Figure 3(a) shows the time-distance plots, with the least squares fitting (linear regression), corresponding to the most noisy case (SNR=−20 dB). The error plot (bottom) shows the residuals with a norm of about 170 μs. Figure 3(b) shows the case of the best SNR (20 dB) and the corresponding residual error norm equal to 0.268 μs. The estimated time shown in y-axis corresponds to the best estimate $\hat{t}_{14}$ which is based on the phase difference between the two components $f_1 = 70$ kHz and $f_4 = 170$ kHz. The time delay estimation values with and without SNR ratios can be seen to be almost identical for each distance except for the two lowest SNR −20 and −10 dB. The data plotted in Fig. 3(b) (top) corresponds to a linear equations of the form $t = md + c$, where c is a measure of the phase offset for the two frequencies equal to $c \times (f_4 - f_1) \times 360$ and $1/m = v$ (mm/μs) is the estimated sound velocity in water, where m is the slope. Solving $t = md + c$ by least squares fitting gave the best estimate m, where $(1/m)$ is the speed of sound v in water.

Table IV represents the results after a least squares fitting estimation, showing the estimated sound velocity and the related offset for each situation (original and −20–20 dB

TABLE III. The estimated time delays for the eight distances with different SNR in Rx.

Distance (mm)	Estimated time (μs)									
d_{meas}	$\hat{t}_{14,Org}$	$\hat{t}_{14,-20\ dB}$	$\hat{t}_{14,-10\ dB}$	$\hat{t}_{14,-5\ dB}$	$\hat{t}_{14,-2\ dB}$	$\hat{t}_{14,0\ dB}$	$\hat{t}_{14,2\ dB}$	$\hat{t}_{14,5\ dB}$	$\hat{t}_{14,10\ dB}$	$\hat{t}_{14,20\ dB}$
305.800	242.05714	328.14350	241.40850	241.86129	241.95502	241.99356	242.01367	242.03214	242.04162	242.04872
345.778	269.18036	366.64520	268.79819	269.06062	269.12812	269.14930	269.16954	269.17011	269.18868	269.18072
481.128	360.43477	347.04536	360.06666	360.31378	360.37864	360.39371	360.41009	360.40717	360.42855	360.43632
535.131	396.63307	492.90101	396.20888	396.50459	396.56616	396.59312	396.60089	396.62446	396.63076	396.62781
535.252	396.76493	493.05600	396.31506	396.62782	396.69481	396.73015	396.72628	396.73571	396.75997	396.75431
624.515	456.86671	442.81417	456.07791	456.61695	456.74634	456.78030	456.81829	456.82773	456.84870	456.85793
759.182	547.80206	522.71162	537.17804	547.60860	547.70913	547.73865	547.75749	547.78524	547.79431	547.80029
862.887	617.79228	391.21736	607.22052	617.63769	617.70199	617.72574	617.75122	617.78719	617.78503	617.80452

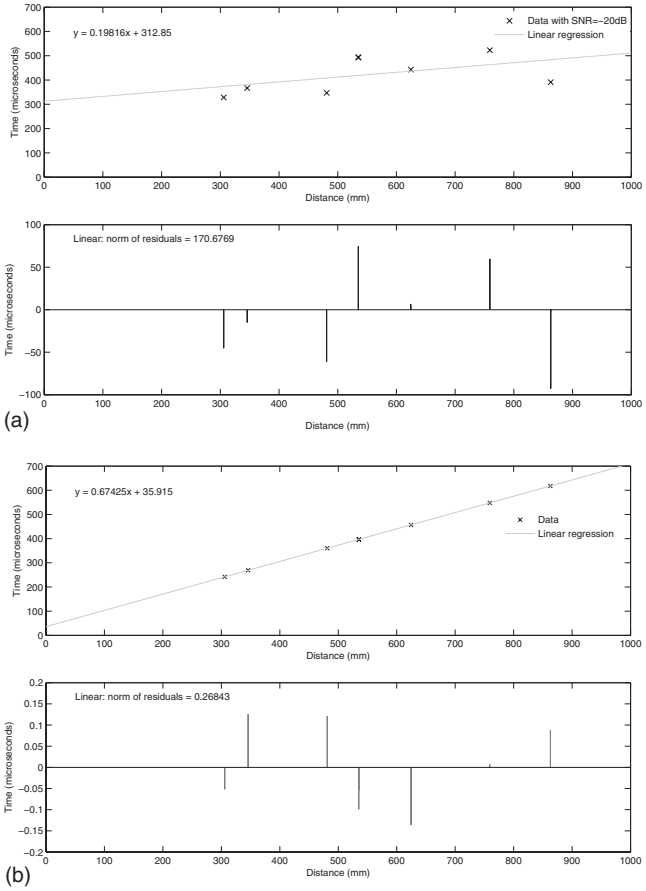


FIG. 3. Estimated times versus distances, −20 dB SNR case (top), noise free case (bottom), and their corresponding least squares fitting showing the residual errors.

SNR). The speed of sound is estimated to be 1483 m/s ±0.2 m/s when the SNR is greater or equal to −5 dB and an average of temperature in the tank of 19.87 °C.

The speed of sound in pure water with similar temperature was reported in Ref. 16 to be 1481.727 m/s and in Ref. 17 to be 1482.36 m/s. Figure 4 shows the speed of sound variation versus different SNR ratios in a logarithmic scale; the approach can be seen to be less sensitive to noise when the SNR is greater than −10 dB.

Table V shows the measured distances and the corresponding estimated distances using the equations of Table IV. In this case, the estimated distance is given by $\hat{d}=(t-c)\times v$, where t is the estimated time corresponding to the best estimate $\hat{t_{14}}$, c is the time offset due to the time delay through the transducers, and v is the estimated sound velocity in water. The differences between the measured and estimated distances lie within the range 5–315 μm (see Table V) when the SNR is greater than −10 dB. The maximum difference represents 1/50 of the wavelength at the central frequency of the transducer (100 kHz).

TABLE IV. Least squares fitting equations for different SNR in Rx and the corresponding estimated speed of sound from the slope.

Curve (dB)	$t=md+C$	$\hat{v}$ (mm/μs)	Phase offset
Noise free	$t=0.674\,182\times d+35.901\,766$	1.483 279	212.463 589
−20	$t=0.198\,160\times d+312.848\,259$	5.046 422	102.537 332
−10	$t=0.653\,855\times d+44.229\,354$	1.529 392	152.256 749
−5	$t=0.674\,171\times d+35.798\,682$	1.483 303	208.752 539
−2	$t=0.674\,186\times d+35.871\,635$	1.483 270	211.378 847
0	$t=0.674\,182\times d+35.901\,766$	1.483 279	212.463 589
2	$t=0.674\,197\times d+35.911\,676$	1.483 247	212.820 339
5	$t=0.674\,244\times d+35.900\,430$	1.483 142	212.415 469
10	$t=0.674\,221\times d+35.926\,900$	1.483 193	213.368 412
20	$t=0.674\,250\times d+35.914\,968$	1.483 130	212.938 860

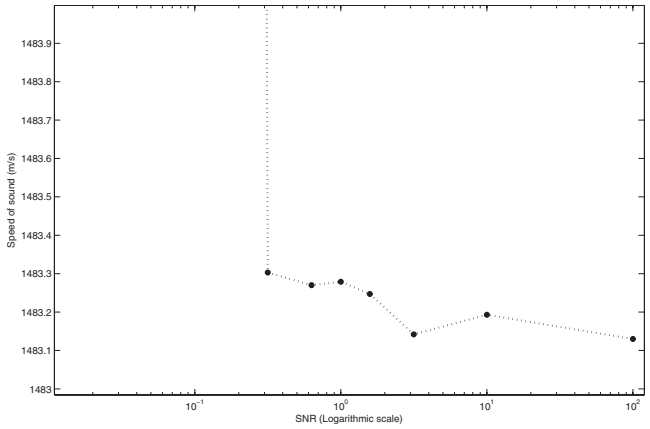


FIG. 4. Estimated speed of sound versus different SNRs (SNR= −5 to 20 dB).

IV. CONCLUSION

In this paper, a time delay estimation approach based on phase differences between components of the received short cw pulse signal is demonstrated. A validation underwater experiment showed that a resolution greater than 1/50 of the wavelength was achievable. Using local phase difference information, no ambiguity in phase measurement arises; hence there is no need to use the Chinese remainder theorem or a coherent local oscillator to overcome the wellknown phase ambiguity problem. As phase information is usually regarded to be useless in most correlator-based techniques, the approach developed in this paper is distinctive. Compared to correlator-based approaches, which have an estimation accuracy on the order of the used bandwidth, the suggested approach is not limited by the bandwidth but only by the ability to measure phase differences accurately. This leads to much improved performance once the SNR is sufficiently high. This approach is tolerant to additive Gaussian noise when the SNR is acceptable. Consequently, the technique offers the potential to outperform animals in subwavelength measurements. Although a bat can achieve a resolution of 20 μm in air,[15] potentially we would get a resolution of 4 mm/360 =11 μm. As a new approach, we expect that the algorithm will be improved to the point where one degree of phase can be resolved, suggesting that performance similar to the bat will be achieved by such technology in the near future.

ACKNOWLEDGMENTS

This work was undertaken as part of the Biologically Inspired Acoustic Systems (BIAS) project, funded by the RCUK Basic Technology Programme Grant Number EP/

TABLE V. The measured (using the encoders) and the estimated distances with different SNR in Rx.

Measured (mm) d_{meas}	Estimated distances (mm) $\hat{d}_{org}$	$\hat{d}_{-20\ \text{dB}}$	$\hat{d}_{-10\ \text{dB}}$	$\hat{d}_{-5\ \text{dB}}$	$\hat{d}_{-2\ \text{dB}}$	$\hat{d}_{0\ \text{dB}}$	$\hat{d}_{2\ \text{dB}}$	$\hat{d}_{5\ \text{dB}}$	$\hat{d}_{10\ \text{dB}}$	$\hat{d}_{20\ \text{dB}}$
305.800	305.785	77.186	301.564	305.653	305.677	305.691	305.700	305.722	305.707	305.723
345.778	346.017	271.482	343.453	345.998	345.982	345.971	345.979	345.972	345.972	345.963
481.128	481.372	172.573	483.039	481.354	481.331	481.312	481.311	481.289	481.298	481.307
535.131	535.065	908.622	538.314	535.036	535.007	535.005	534.991	535.004	534.993	534.984
535.252	535.260	909.404	538.477	535.218	535.198	535.209	535.177	535.169	535.185	535.171
624.515	624.408	655.862	629.877	624.200	624.270	624.280	624.308	624.294	624.308	624.313
759.182	759.290	1059.059	753.911	759.169	759.193	759.196	759.193	759.197	759.198	759.192
862.887	863.105	395.483	861.034	863.043	863.011	863.007	863.011	863.020	863.007	863.017

C523776/1. The BIAS project involves collaboration between the British Geological Survey, Leicester University, Southampton University, Leeds University, Edinburgh University, and Strathclyde University. This paper is published with the permission of the Executive Director of the British Geological Survey (NERC).

APPENDIX: TIME DELAY ESTIMATION ALGORITHM USING PHASE DIFFERENCES

Input: Set of phases $(\phi_1,\phi_2,\phi_3,\phi_4)$ and frequencies (f_1,f_2,f_3,f_4)
Output: Set of estimated time delay $(\hat{t}_{12},\hat{t}_{13},\hat{t}_{14},\hat{t}_{23},\hat{t}_{24},\hat{t}_{34})$
foreach *Received signal, k* **do**
  Calculate DFT (f_1,f_2,f_3,f_4,k);
  Get $\phi_1,\phi_2,\phi_3,\phi_4$
end
foreach *(i,j)=(1,3);(1,4);(2,3);(2,4);(3,4)* **do**
  $\Delta\phi_{ij}=\phi_j-\phi_i$;
  if $((\Delta\phi_{ij})<0)$ **then**
    $\Delta\phi_{ij}=\Delta\phi_{ij}+360.0$;
  end
end
$\Delta\phi_{12}=\frac{\Delta\phi_{12}}{360.0}$, $\Delta f_{12}=f_2-f_1$, $\hat{t}_{12}=\frac{\Delta\phi_{12}}{\Delta f_{12}}$, $\hat{t}_{old}=\hat{t}_{12}$
foreach *(i,j)=(1,3);(1,4);(2,3);(2,4);(3,4)* **do**
  if $((\hat{t}_{old}\times\Delta f_{ij})<1.0)$ **then**
    $\hat{t}_{new}=\frac{\Delta\phi_{ij}}{\Delta f_{ij}}$;
    nmin$=-1$;
  end
  *min*1.0*e*+06;
  for each $n=0$ *to* 1000 **do**
    Calculate $\hat{t}_{ij}=(n+\Delta\phi_{ij})/\Delta f_{ij}$;
    $\Delta t=|\hat{t}_{ij}-\hat{t}_{old}|$;
    if $(\Delta t<min)$ **then**
      $min=\Delta t$;
      $\hat{t}_{new}=\hat{t}_{ij}$;
      nmin$=n$;
    end
  end
end

[1]G. H. Knapp and G. C. Carter, "The generalised correlation method for estimation of time delay," IEEE Trans. Acoust., Speech, Signal Process. **24**, 320–327 (1976).
[2]G. C. Carter, "Time delay estimation for passive sonar signal processing," IEEE Trans. Acoust., Speech, Signal Process. **29**, 463–470 (1981).
[3]G. C. Carter, "Coherence and time delay estimation," in *Signal Processing Handbook*, edited by C. H. Chen (Dekker, New York, 1998), pp. 443–482.
[4]R. E. Boucher and J. C. Hassab, "Analysis of discrete implementation of generalised cross-correlator," IEEE Trans. Acoust., Speech, Signal Process. **ASSP-29**, 609–611 (1981).
[5]J. Chen, J. Benesty, and Y. A. Huang, "Time delay estimation via linear interpolation and cross correlation," IEEE Trans. Speech Audio Process. **12**, 509–519 (2004).
[6]J. Chen, J. Benesty, and Y. A. Huang, "Time delay estimation in room acoustic environments: An overview," EURASIP J. Appl. Signal Process. **2006**, 1–19.
[7]L. Belostotski, T. L. Landecker, and D. Routledge, "Distance measurement with phase stable CW radio link using the Chinese remainder theorem," Electron. Lett. **37**, 521–522 (2001).
[8]C. E. Towers, P. D. Towers, and D. C. Jones Julian, "The efficient Chinese remainder theorem algorithm for full-field fringe phase analysis in multi-wavelength interferometry," Opt. Express **12**, 1136–1143 (2004).
[9]X. Xiang-Gen and L. Kejing, "A generalized Chinese remainder theorem for residue sets with errors and its application in frequency determination from multiple sensors with low sampling rates," IEEE Signal Process. Lett. **12**, 768–771 (2005).
[10]O. Goldreich, D. Ron, and M. Sudan, "Chinese remaindering with errors," IEEE Trans. Inf. Theory **93**, 1330–1338 (2000).
[11]A. K. T. Lee, J. Lucas, and L. E. Virr, "Microcomputer-controlled acoustic rangefinding technique," J. Phys. E: J. Sci. Instrum. **22**, 52–58 (1989).
[12]M. Yang, S. L. Hill, B. Bury, and J. O. Gray, "A multifrequency AM-based ultrasonics system for accuracy distance measurement," IEEE Trans. Instrum. Meas. **43**, 861–866 (1994).
[13]H. Hartridge, "Acoustical control in the flight of bats," Nature (London) **156**, 490–494 (1945); **156**, 692–693 (1945).
[14]J. A. Simmons, "The resolution of target range by echolocating bats," J. Acoust. Soc. Am. **54**, 157–173 (1973).
[15]W. L. Whitlow and J. A. Simmons, "Echolocation in dolphins and bats," Phys. Today **60**(9), 40–45 (2007).
[16]V. A. Del Grosso and C. W. Mader, "Speed of sound in pure water," J. Acoust. Soc. Am. **52**, 1442–1446 (1972).
[17]N. Bilaniuk and G. S. K. Wong, "Speed of sound in pure water as a function of temperature," J. Acoust. Soc. Am. **93**, 1609–1612 (1993).

Distortion product otoacoustic emission phase and component analysis in human newborns

Carolina Abdala[a)]
Division of Communication and Auditory Neuroscience, House Ear Institute, 2100 West Third Street, Los Angeles, California 90057

Sumitrajit Dhar
Roxelyn and Richard Pepper Department of Communication Sciences and Disorders, Hugh Knowles Center, Northwestern University, Evanston, Illinois 60208

(Received 2 September 2009; revised 2 November 2009; accepted 5 November 2009)

Apical distortion product otoacoustic emissions (DPOAEs) are comprised of at least two components, as evidenced by the interference pattern of alternating maxima and minima known as fine structure. DPOAE fine structure is produced by the shifting phase relationship in the ear canal, between the generator and characteristic frequency (CF) component of the response. Each component arises from a different cochlear region and, according to theory, reflects a distinct generation mechanism. The analysis of DPOAE components and phase in newborns may provide a window into targeted aspects of cochlear physiology during development. $2f_1-f_2$ DPOAE fine structure was recorded from 15 adults and 14 newborns using a swept-tone technique. DPOAE group delay, as well as magnitude and phase of each component, was compared between age groups. Results show narrower fine structure spacing, a longer group delay (steeper phase gradient) in low frequencies, and a stronger relative contribution from the CF component in newborns. The prolonged group delay for low-frequency DPOAEs could indicate immature basilar membrane motion in the apex of the cochlea and warrants further investigation. The enhanced contribution from the CF component may have implications for clinical practice as well as for theories of cochlear maturation. © *2010 Acoustical Society of America.* [DOI: 10.1121/1.3268611]

PACS number(s): 43.64.Bt, 43.64.Jb, 43.64.Kc, 43.64.Ha [BLM] Pages: 316–325

I. INTRODUCTION

It is now consensually accepted that ear canal recordings of apical ($fdp<f_1<f_2$) distortion product otoacoustic emissions (DPOAEs) include at least two components: one from the generator region, where traveling waves evoked by f_1 and f_2 overlap maximally (nearer f_2), and one from the characteristic frequency (CF) region of the DPOAE. These two components not only arise from different sites on the cochlea but, according to recent theory (Shera and Guinan, 1999), also reflect two distinct generation mechanisms. The generator site near f_2 produces inter-modulation distortion from nonlinearities in outer hair cell (OHC) function. The CF component is thought to be produced predominantly by irregularities along the basilar membrane, which scatter input energy via linear, coherent reflections. These two components will mainly be referred to throughout this report as the generator (i.e., distortion, overlap, and nonlinear sources) and the CF component (i.e., reflection and linear sources).

The existence of two DPOAE components from disparate cochlear regions (Heitmann *et al.*, 1998; Talmadge *et al.*, 1999; Mauermann *et al.*, 1999a,1999b) and distinct generation processes (Shera and Guinan, 1999) suggests that the manner in which DPOAEs are interpreted should be reexamined and possibly refined. Each DPOAE component may be uniquely sensitive to cochlear pathologies of differing etiology. As such, the more we learn about these sources, their normal characteristics in the adult cochlea, and their maturational time course, the more likely their eventual application to effectively detect auditory pathology.

Interaction between the two DPOAE components, each with distinct phase behavior as a function of frequency, produces a pattern of alternating maxima and minima known as DPOAE fine structure. To our knowledge, only one other publication (from our joint laboratories) has defined fine structure in newborns. Dhar and Abdala (2007) described newborn fine structure and found it to be more prevalent, deeper, and with slightly broader frequency spacing. They did not compare DPOAE phase between the two age groups or examine individual DPOAE components. Although it was a preliminary study using a less than optimal recording protocol (i.e., discrete frequency versus swept tone), the age differences reported, nevertheless, suggest a peripheral auditory immaturity in newborns.

DPOAE age effects can only be fully interpreted after considering the impact of immature outer and middle ear functions. Past work has used the DPOAE input/output function to model the effects the immature middle ear has on forward transmission (of stimulus) and reverse transmission (of DPOAE) (Abdala and Keefe, 2006; Keefe and Abdala, 2007). For high frequencies (4000–6000 Hz), these models have shown that a significant portion of the immaturity in peripheral auditory function can be accounted for by an im-

a)Author to whom correspondence should be addressed. Electronic mail: cabdala@hei.org

mature conductive system that attenuates sound energy as it passes through the middle ear cavity toward the cochlea. The model also suggests that forward transmission through the middle ear should be relatively adult-like by six months, yet at this age, DPOAE ipsilateral suppression is still not adult-like (Abdala *et al.*, 2007). This inconsistency suggests residual DPOAE immaturities not easily explained by outer and middle ear factors.

Given the early morphological maturity of OHCs (Pujol *et al.*, 1998) and relatively normal OAEs in newborns (once the middle ear is considered), it is not likely that this peripheral immaturity involves cochlear micromechanics. However, passive motion of the basilar membrane determined by its physical properties may be later developing and could account for residual immaturities, as noted in laboratory animals (Mills and Rubel, 1996; Overstreet *et al.*, 2002). Additionally, modulation of the cochlea by the medial efferent system may show postnatal immaturity, as OHC innervation by medial olivocochlear fibers does not commence until sometime around the third trimester and may not be refined until later (Lavigne-Rebillard and Pujol, 1988). The measurement of DPOAE phase and the examination of individual DPOAE components in newborns provide a window into cochlear function during development.

Three research areas were addressed in this study: (1) Earlier findings were replicated, evaluating basic fine structure features in newborns with an innovative swept-tone methodology. (2) Phase characteristics of the DPOAE were compared in newborns and adults. (3) Generator and CF components of the DPOAE were evaluated individually in both age groups to assess whether source contribution to the composite DPOAE is mature at birth.

II. METHODS

A. Subjects

15 normal-hearing adults and 14 newborn infants participated in this study. The 15 normal-hearing adults had a mean age of 25.4 years (range=18–32 years): 6 were males and 9 were females. One ear was tested per subject including eight right and seven left ears. Adults had normal audiograms [<15 dB hearing level (HL) between 500 and 8000 Hz] and type A tympanograms with static compliance between 0.4 and 1.5 cm^3 and peak pressure between −50 and 150 daPa. The neonates were term born (delivered between 38 and 41 weeks of gestation) and were tested on average within 43.2 h of birth (range=17–72 h). Nine were females and five were males. Mean birthweight was 3132 g (range =2520–4425 g). All newborn subjects included in this study passed the click-evoked auditory brainstem response hearing screen at 30 dB HL and had no high-risk factors for hearing loss.

B. Signal analysis and instrumentation

Signal generation and recording were controlled using custom software developed by Dr. Carrick Talmadge and run on an Apple Macintosh G4 computer via a MOTU 828 Mk II input/output device (24 bits/44 100 Hz). Stimulus tones were presented to the subjects' ear canal via ER2 insert phones and DPOAEs were recorded with the ER10B+ microphone. The output of the microphone was preamplified and then passed through an analog high-pass filter with 300 Hz cutoff frequency before being digitized by the MOTU and stored on disk. DPOAE recordings were made over $2f_1-f_2$ frequencies spanning three-octaves, between 500 (f_2=782 Hz) and 4000 Hz (f_2=6256 Hz), using fixed stimulus levels of 65 (L1) and 55 (L2) dB sound pressure level (SPL) and a constant stimulus frequency ratio of $f_2/f_1=1.22$. The stimulus tones were swept at a rate of 8 s/octave for optimal definition of DPOAE fine structure (Long *et al.*, 2008). For newborns, eight such sweeps were averaged in each condition, whereas six sweeps were averaged in adults.

DPOAE level and phase estimates were obtained using a least-squares-fit (LSF) algorithm as described by Long *et al.* (2008) and yielded estimates at every 2 Hz around 500–1000 Hz and every 6 Hz around 4000 Hz. The noise floor was estimated similarly except that every other temporal window was inverted and the pair summed to cancel the signal. Stimulus levels were calibrated in a Zwislocki coupler. The two transducers in the ER10B+ were individually equalized to produce flat constant drive voltage frequency responses of up to 7000 Hz. System distortion was below 30 dB SPL for the stimulus levels used in these recordings. The DPOAE phase estimate was corrected for variations in primary tone phase by subtracting $2\phi_1-\phi_2$, where ϕ_1 and ϕ_2 are phases of the lower- and higher-frequency stimulus tones, respectively.

C. Protocol

All adult subjects were tested at the House Ear Institute (Los Angeles, CA). They received an audiogram and standard tympanometry as well as a brief questionnaire to screen for study inclusion. Following this, they were seated for DPOAE testing in a cushioned arm chair within a sound-attenuated booth. Six three-octave primary tone sweeps (each lasting 24 s) were recorded for off-line analysis. Duration of the entire protocol, including preliminary screening, was approximately 60 min.

Newborn subjects were tested at University of Southern California + Los Angeles County Medical Center. They were swaddled and fed, if necessary, prior to testing, and then positioned on their side or back in an Eckels (ABC) sound-attenuated isolette (www.ekel.ca). A soft rubber-tip probe was fit into the meatus of the ear canal and taped to the pinna securely with surgical tape. The test did not commence until the infant was sleeping soundly. Eight three-octave primary tone sweeps (each lasting 24 s) were recorded for off-line analysis. Sweeps were manually paused if the infant moved significantly or if there was any atypical increase in ambient room noise, and then repeated once the noise returned to normal baseline. While one experimenter monitored infant status visually, another monitored the peak activity recorded by the microphone on the computer screen. Either experimenter could initiate the software pause feature as needed. These necessary delays and sweep repetitions prolonged test sessions significantly, which led to a relatively low 30%–40% success rate with newborn infants. The entire

infant protocol including preparation and calming of the infant averaged approximately 90 min but never exceeded 2 h.

D. Data analysis

1. Fine structure classification

Fine structure features were extracted with an automatic algorithm implemented previously by Dhar and Abdala (2007) and detailed below.

(a) *DPOAE level.* The median values for DPOAE level and the noise floor were computed over every three successive data points. Data points where the signal-to-noise ratio (SNR) between level and noise floor medians was less than 6 dB were eliminated.
(b) *Maxima.* Fine structure maxima were identified based on the first and second derivatives of the DPOAE level function and the relationship between them. Data points where the first derivative was equal to zero were identified as extrema, and then further classified as a maximum or minimum based on the second derivative being negative (maxima) or positive (minima). This process was checked by one observer familiar with the morphology of DPOAE fine structure. The visual check was implemented to eliminate peaks associated with noise that were erroneously identified by the program as maxima. These errors typically included maxima that were artificially tall and narrow and mimicked the noise floor configuration in the low-frequency range. In newborns, 78% of the maxima eliminated during visual analysis were in the low-frequency region below 1000 Hz, and in adults, 73% were <1000 Hz. Prevalence of DPOAE fine structure was quantified by counting the number of maxima in each 1/3 octave interval.
(c) *Depth and spacing.* The depth of each identified fine structure period was computed as $20\log_{10}(P_{\max}/P_{\text{av\_min}})$, where $P_{\max}$ is the DPOAE level at a maximum and $P_{\text{av\_min}}$ is the average DPOAE level of the preceding and following minima. Frequency spacing of fine structure was computed both in hertz and as $f/\Delta f$, where f is the geometric mean between two adjacent minima and Δf is the frequency separation between them (Shera, 2003). Only fine structure periods with depth ≥ 2.5 dB and spacing ratio ≤ 25 were accepted. Additionally, values that were more than 2 standard deviations from the mean were eliminated.

2. DPOAE phase

DPOAE phase as a function of frequency was characterized as group delay (i.e., phase gradient delay), which is the negative of the slope of phase. Starting with the initial phase value at the lowest frequency, a linear regression function was fitted to five consecutive phase values with a two-point overlap. The slope of the line was calculated to yield the group delay in milliseconds.

3. Component separation

DPOAE level was separated into its two constituent components using an inverse fast Fourier transform (IFFT) algorithm. A MATLAB-based analysis software (NIPR) developed by Talmadge *et al.* (1999) uses a variant of the IFFT algorithms described by Dhar *et al.*, 2002 and Withnell *et al.*, 2003 to separate the DPOAE generator and CF components based on their respective group delays. During IFFT, DPOAE complex amplitude measured in the frequency domain is multiplied by a moving Welch window (400 Hz window width and 50 Hz steps). The IFFT converts each windowed data set into the time domain, after which a fixed time window filter is applied to extract the desired delay component in the time domain. These filtered windows of data are then transformed back to the frequency domain by FFT and the complex amplitudes of the generator (around f_2) and CF component (at $2f_1-f_2$) are reconstructed.

4. Statistical analysis

Although the swept-tone paradigm provided DPOAE estimates every 2–6 Hz, resulting in 400–500 values of level and phase across the three-octave range, all variables were averaged into 500-Hz-wide frequency bins for statistical analysis and display.

(a) *Composite DPOAE.* DPOAE level, phase (group delay), fine structure prevalence, spacing, and depth were computed for each individual subject and compared across frequency and age using Analyses of Variance (ANOVAs) with repeated measures on frequency.
(b) *DPOAE components.* Level was compared across component, age, and frequency using a three-way ANOVA with repeated measures on component and frequency. Phase of each DPOAE component as a function of frequency was calculated with a linear regression function. For phase measurements, only values from 1000 to 4000 Hz were included because data <1000 were variably present in each subject. The resulting slope of phase measurements were compared across age and component.

III. RESULTS

A. DPOAE fine structure

Newborns showed more fine structure periods per given frequency interval than adults ($f=6.76; p=0.01$). There was an interaction between frequency and age ($f=3.27; p=0.007$) indicating that the age difference was greatest in the higher-frequency range above 2000 Hz. Overall, newborns had 2.0 fine structure periods per 1/3 octave while adults had 1.44.

DPOAE level did not show an age effect. However, an interaction between frequency and age was present ($f=3.76; p=0.003$). As evident from the data presented in Fig. 1, infants showed a trend toward higher levels than adults below 3000 Hz. Figures 2(a)–2(d) show an example of DPOAE fine structure from two randomly selected newborn and adult subjects.

For both adults and infants, around 1500 Hz, the frequency width of fine structure periods was about 100 Hz; at 3000 Hz, spacing ranged from 175 to 225 Hz. Frequency spacing values were converted to spacing ratio values ($f/\Delta f$)

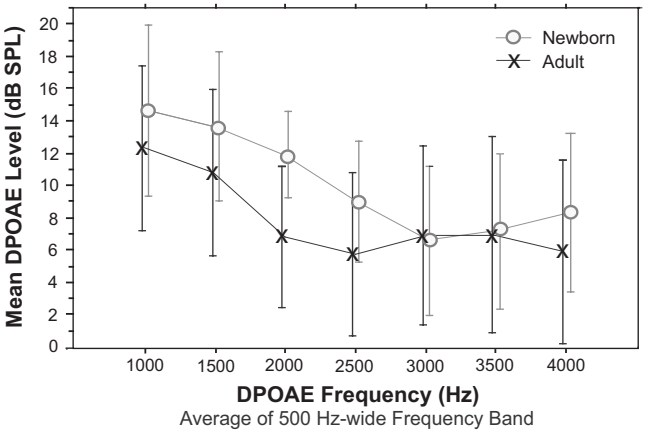


FIG. 1. DPOAE level as a function of DPOAE frequency for 15 adults and 14 newborns. Level was averaged into 500-Hz-wide frequency bands. The frequency displayed represents the upper limit of this band. Error bars = ± 1 SD.

to accommodate the logarithmic frequency scale of the cochlea. Results of a two-way ANOVA (frequency × age) on spacing ratio showed an effect of both age and frequency on DPOAE fine structure spacing [Fig. 3(a)]. Infants had narrower fine structure spacing than adults ($f=13.07; p=0.0005$) and the spacing ratio increased as frequency increased ($f=5.06; p=0.001$).

The infant fine structure data were strongly biased toward high-frequency values relative to adult data because fine structure periods were more prevalent from 2500 to 4000 Hz in newborns. In this high-frequency range, infants had 29 values compared to the 15 data points available from the adults. A secondary ANOVA was also conducted, examining age effects on spacing for the frequency range through 2500 Hz only (similar to the frequency range tested in a previous publication from our joint laboratories, Dhar and Abdala, 2007). This analysis showed no effect of age on the spacing between DPOAE fine structure periods. Although there was a general trend for deeper fine structure in newborns ($p=0.08$) especially in the high frequencies as evident in Fig. 3(b), neither age nor frequency produced significant effects on depth [Fig. 3(b)].

B. DPOAE phase

The slope of DPOAE phase was quantified by calculating a measure of group delay or phase gradient. This index represents the rate or slope of phase as it changes across frequency. Group delay was near 0 ms for both age groups in the mid- to high frequencies with a typical break from this near-zero constant at frequencies below 2000 Hz (Fig. 4). There was a significant difference in group delay between newborns and adults ($f=9.77; p=0.004$). This difference was present in low frequencies only, as seen in Fig. 4. The mean delay for newborns for the lowest-frequency band was 2.35 ms, while for adults, it was 1.74 ms. In the frequency interval between 1000 and 1500 Hz, the newborn group delay was

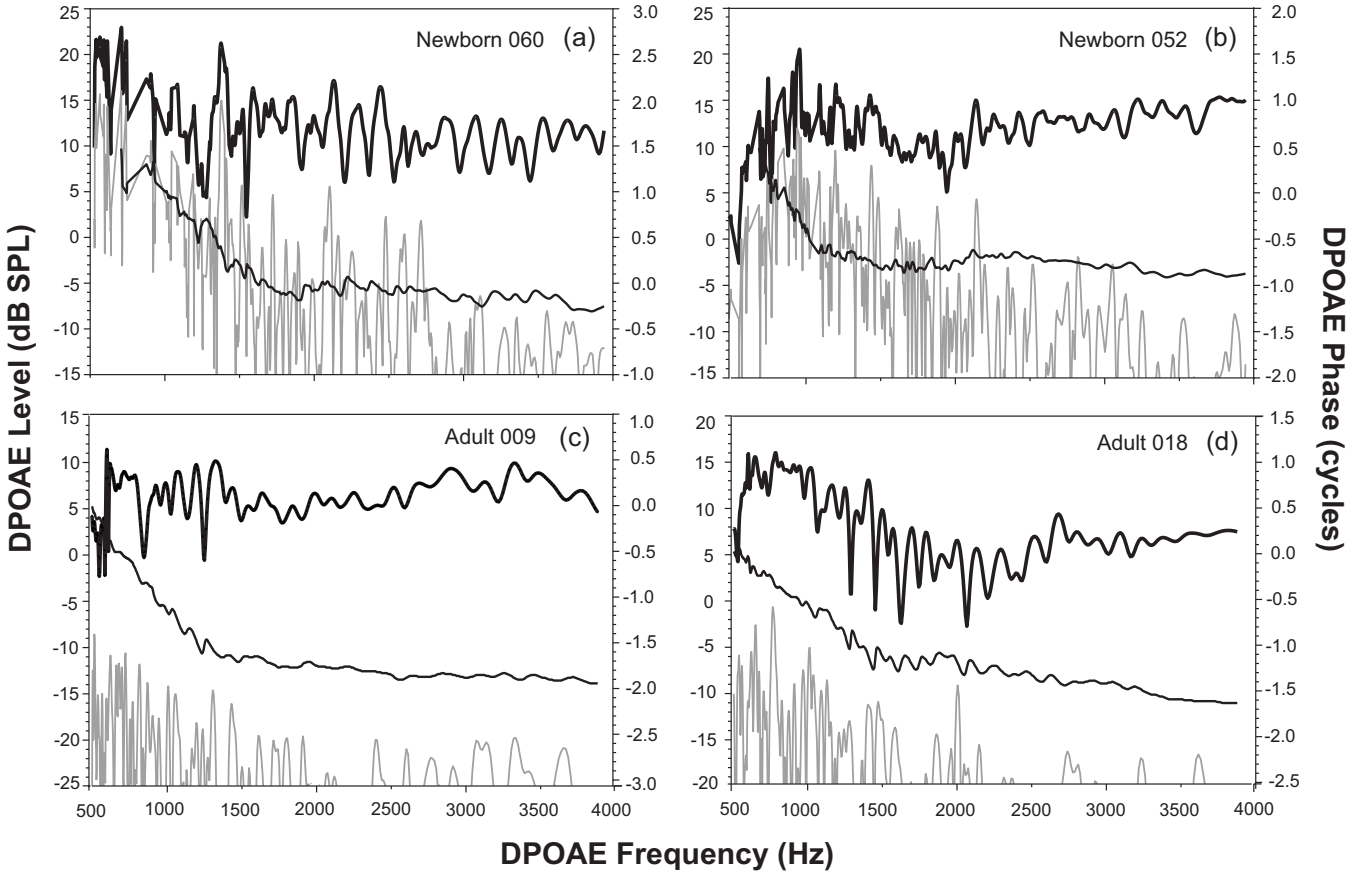


FIG. 2. [(a)–(d)] DPOAE fine structure (thick line), phase (thin line), and noise floor (gray line) from two newborn and two adult subjects. The range of values is the same in each graph, though absolute values vary. Note that DPOAE phase is referenced to the right vertical axis.

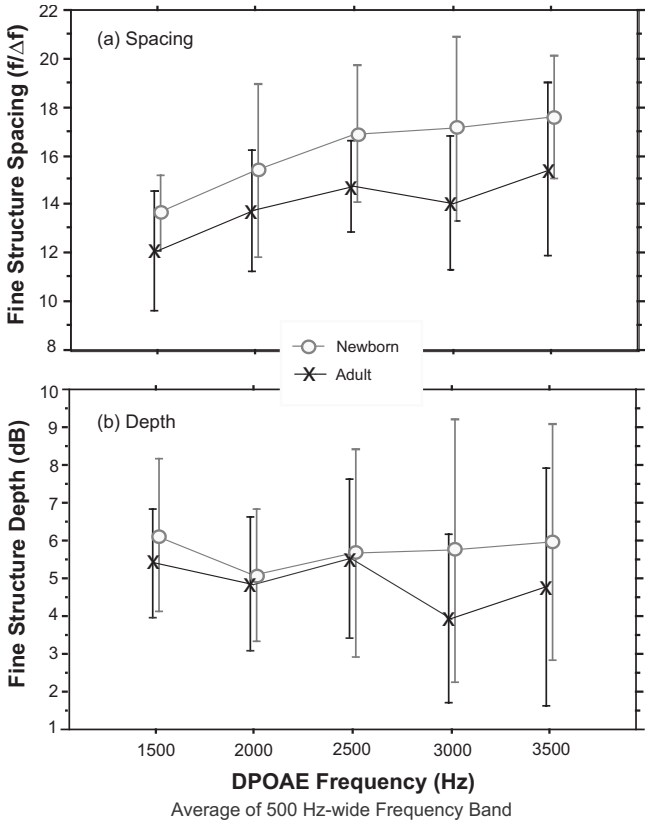


FIG. 3. Mean DPOAE fine structure features from 15 adults and 14 newborns: (a) frequency spacing between fine structure periods and (b) depth of fine structure periods. Data were averaged into 500-Hz-wide frequency bands. The frequency displayed represents the upper limit of this band. Error bars= ± 1 SD.

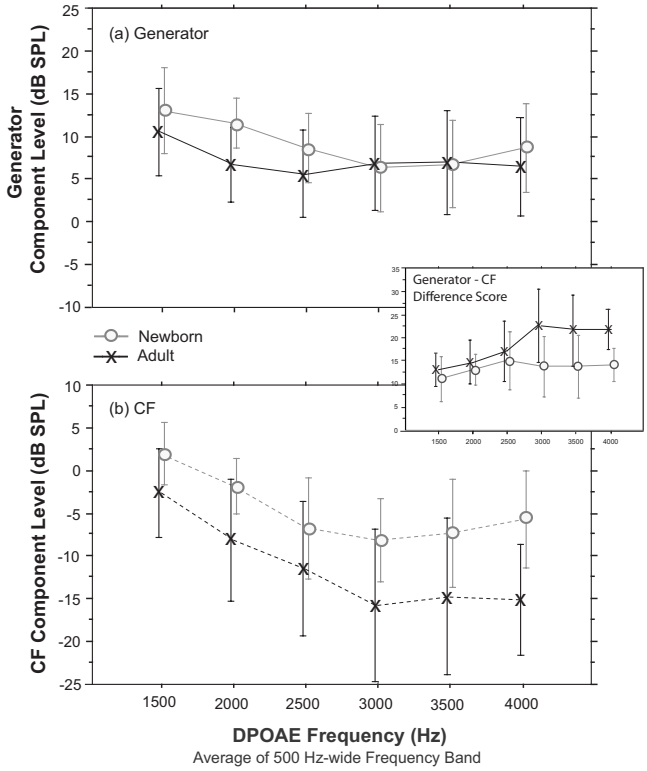


FIG. 5. [(a) and (b)] DPOAE level as a function of frequency for each component and age group separately. The smaller inset graph shows the level difference between components (generator level subtracted from CF level) for each age group. Data were averaged into 500-Hz-wide frequency bands. The frequency displayed represents the upper limit of these bands. Error bars= ± 1 SD.

double the adult value (1.91 versus 97). This finding indicates a steeper DPOAE phase gradient in the newborns at the lowest frequencies.

C. DPOAE components

1. Component magnitude

Once the DPOAE data were analyzed with an IFFT, the magnitude and phase of the two components comprising the ear canal response were examined separately. The generator component showed much higher levels than the CF component, which is typical in normal-hearing adults (Fig. 5, $f=433; p<0.0001$). There was also an interaction between age and component ($f=9.19; p=0.0053$). *Posthoc* analysis showed that newborns had a larger CF component than adults ($f=12.1; p=0.001$), but that generator component levels were comparable between age groups. The generator component averaged across many subjects should approximate the composite DPOAE level. This correspondence can be observed when comparing the data in Fig. 5(a) to data in Fig. 1.

The difference between component levels was further analyzed by calculating a difference score: CF component level subtracted from generator component level (Fig. 5, inset). Adults had a greater difference score than newborns ($f=10.52; p=0.0031$), but only for frequencies >2500 Hz. Since the generator component level is comparable between age groups, the reduction in adult CF level produces the larger difference scores noted in Fig. 5.

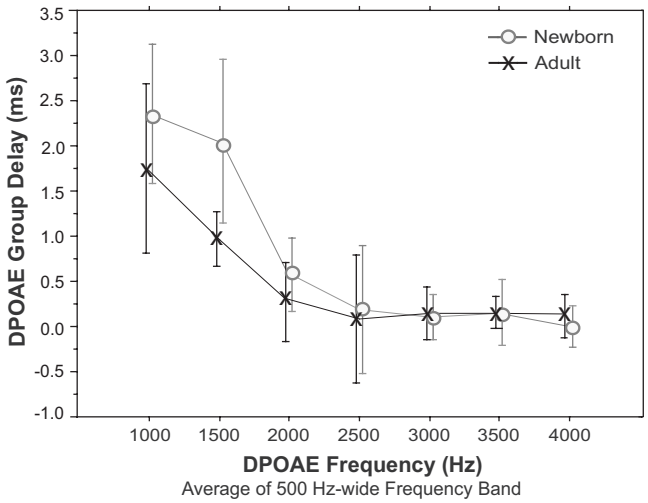


FIG. 4. Mean DPOAE group delay (negative of the slope of the phase) as a function of DPOAE frequency for 15 adults and 14 newborns. Data were averaged into 500-Hz-wide frequency bands. The frequency displayed represents the upper limit of this band. Error bars= ± 1 SD.

2. Phase

Component phase was analyzed by fitting a linear regression to the phase × frequency function for generator and CF component separately. Results shown in Fig. 6 manifest the expected constant phase (i.e., shallow slope) for the distortion component and rapidly cycling phase (i.e., steep slope) for the CF component. This figure shows individual

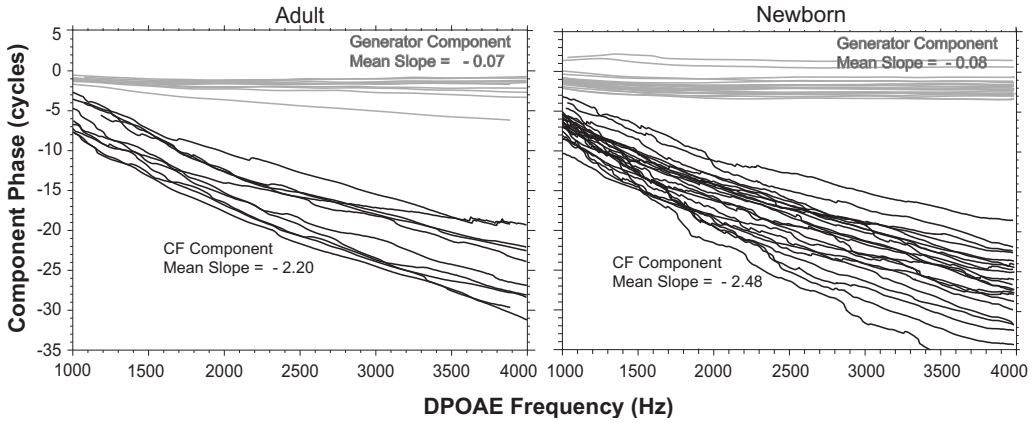


FIG. 6. DPOAE phase for the generator (gray) and CF component (black) as a function of DPOAE frequency for both age groups. Lines represent phase from one adult subject, whereas each newborn subject has two such traces displayed. The mean slope of phase values, calculated with a linear regression function, are shown for each age and component.

phase × frequency functions for each newborn and adult subject for both components. Clearly, there was a significant effect of component on slope of phase ($f=806; p<0.0001$) but there was not a significant effect of age on slope of phase for either component, although there was a trend for steeper slope of phase in infants for the CF component only ($p<0.08$).

IV. DISCUSSION

These data show that infants have more prevalent DPOAE fine structure than adults in high frequencies, narrower spacing between fine structure periods, and a steeper DPOAE phase gradient (longer group delay) than adults at frequencies below 2000 Hz. While the greater prevalence of fine structure in infants is consistent with our previous report (Dhar and Abdala, 2007), the narrower spacing in infants is not. Possible reasons for this and other inconsistencies among studies of DPOAE fine structure are further discussed in Sec. IV B.

A. Cochlear physiology

1. Steep phase gradient

Spacing of DPOAE fine structure is primarily determined by the slope of CF component phase since the phase of the generator component is largely invariant as a function of frequency (the special case of frequencies below 2000 Hz is discussed separately). Given similarly flat slopes of phase for the generator component in adults and newborns (Fig. 6), the narrower fine structure spacing found in infants implies a steeper phase gradient for the newborn CF component. Past work from the first author's laboratory has, in fact, found steeper CF component slope of phase for infants when frequencies between 500 and 1000 Hz were included in analysis; however, in the present study, slope of phase calculations for data lower than 1000 Hz were eliminated because the number of observations in this interval (500–1000 Hz) varied widely among newborn subjects due to signal-to-noise ratio (SNR) issues. Thus, the steepest part of the phase × frequency function was eliminated in the present analysis and may have contributed to the finding of borderline steeper slope in the phase of the DPOAE CF component measured from newborns.

Some have argued that DPOAE group delay is reflective of travel times in the cochlea (Moulin and Kemp, 1996a,1996b). Others have extended this interpretation to imply an underlying relationship with the sharpness of mechanical tuning in the cochlea (Bowman *et al.*, 1997). However, these interpretations are complicated by the dependence of the delay on the DPOAE recording conditions (Tubis *et al.*, 2000). Some have questioned the connection between estimates of cochlear travel times and DPOAE delays (Ruggero, 2004; Siegel *et al.*, 2005), while others have shown that group delays measured from SFOAEs are correlated with the sharpness of psychophysical tuning curves (Shera *et al.*, 2000). Slope of phase of the DPOAE CF component would be most closely related to the slope of phase of SFOAEs. Thus, the observation of narrower fine structure spacing in newborns could be interpreted to be indicative of sharper tuning. Consistent with this idea, DPOAE suppression tuning has been shown to be excessively sharp in human infants at high frequencies (Abdala *et al.*, 2007). However, the link between OAE suppression and mechanical tuning of the basilar membrane is not straightforward (Howard *et al.*, 2003).

2. Scaling symmetry

Consistent with previous observations, group delay was near 0 ms in the mid- to high frequencies with a typical break from this near-zero constant at frequencies below 2000 Hz. The invariance of DPOAE phase as a function of frequency above some mid-frequency boundary has been attributed to cochlear *scaling symmetry*. Since the most sensitive region for high to low frequencies is arranged from the base to the apex of the cochlea, vibrations caused by signals of different frequency "travel" different distances before reaching their best or CF region. Scaling symmetry simply implies that the number of cycles to the CF site is relatively independent of frequency, and mechanical vibrations produced by signals of all frequencies accumulate approximately the same amount of total phase (cycles) at their unique CF site.

Evidence supporting scaling symmetry has been observed in laboratory animals and human adults for a relatively large frequency range, with deviations from perfect scale invariance at lower frequencies, i.e., nearer the cochlear apex (Dhar *et al.*, 2002; Shera and Guinan, 1999; Shera *et al.*, 2000; Talmadge *et al.*, 1999). This break from symmetry has not previously been shown in the *developing* human auditory system, but is clearly present in the newborn data presented here. For the lowest frequencies, the prolongation of group delay values in newborns was significantly greater than adult delays, possibly indicating more marked deviation from scaling invariance in the apex of the newborn cochlea.

Violations of perfect scaling invariance (defined by DPOAE data) may be due to multiple sources, some related to cochlear physiology and others related to cochlear roughness, wave reflection, and source interference (Dhar *et al.*, 2002; Shera *et al.*, 2000). Deviations from scaling symmetry can also be produced by shifting transfer functions of the outer/middle ear. For example, Moulin and Kemp (1996a, 1996b) demonstrated the influence of middle ear pressure on DPOAE phase at low frequencies. Clearly, the relationship between DPOAE group delay and cochlear mechanics must be interpreted while considering this complexity and the varied sources contributing to scale invariance.

If newborns show a greater break from scaling symmetry than adults (i.e., a more prolonged group delay only at low frequencies), it could indicate that any of the above-named factors are immature around the time of birth. Morphologically, the mammalian cochlear apex is the last region to develop and segments of the apex remain non-adult-like well beyond the basal cochlea (Bredberg, 1968; Lavigne-Rebillard and Pujol, 1986, 1987; Pujol *et al.*, 1998). Such immaturity in the newborn cochlea, regardless of its exact nature, could be responsible for the comparatively larger deviation from scaling symmetry. The larger CF component observed in the newborns in the present study points to the possibility of increased cochlear roughness, which could either be the cause or simply a result of the deviation from scale invariance. Moreover, the observed break from scaling symmetry might suggest immaturities in frequency representation in the apical region of the newborn cochlea. These questions are of significant interest to appropriately define the source of infant immaturities in peripheral auditory system function and warrant further research.

3. CF component level

Newborns show higher relative levels of the CF component than adults, suggesting a larger contribution from the $2f_1-f_2$ site. This finding could be interpreted in two ways: The newborn cochlea is immature and/or the adult cochlea is compromised by normal aging and exposure. The low-level CF component (reflection) is more sensitive to integrity of the cochlear amplifier and thus would be quite robust and optimally functional in the pristine neonatal cochlea. Several models of DPOAE generation including coherent reflection filtering (Shera and Zweig, 1993b; Talmadge *et al.*, 1998; Zweig and Shera, 1995) depend on wave scattering from random, physical irregularities along the basilar membrane. Irrespective of the source of the roughness, the output of all such models depends critically on the height and breadth of the traveling wave. A more active cochlear amplifier would result in greater magnitude of vibration near the CF (a "taller" traveling wave), thereby leading to greater reflection and, subsequently, a larger DPOAE component from the CF region. Consistent with this hypothesis, newborns have larger (Burns *et al.*, 1992) and greater numbers of spontaneous OAEs (SOAEs) per ear (Abdala, 1996). SOAEs are thought to be amplitude-stabilized cochlear standing waves originating from the same reflection process as the CF component of DPOAEs (Shera, 2003). The combination of the pristine infant cochlea and natural adult exposure to damaging noise and ototoxins may explain the larger CF component in newborns, arguing for early "auditory aging" in the adults rather than immaturity in the newborns.

A second possibility to explain a larger relative contribution from the CF component in newborns involves the medial olivocochlear (MOC) effect. Although cochlear micromechanics may be mature in the infant, modulation of OHC function by the medial efferent tract may remain immature around the time of term birth. MOC fibers are present early in gestational life (Moore *et al.*, 1999), but it is not clear when MOC *function* becomes adult-like. There are limited reports of the MOC reflex in infants and most suggest the reflex is immature in prematurely born neonates and is approaching maturity around the time of term birth (Abdala *et al.*, 1999; Ryan and Piron, 1994; Morlet *et al.*, 1993). However, methods for probing the medial efferent system have changed significantly since these reports and they may not provide the most complete picture. It is possible that residual MOC immaturities in newborns influence the relative contribution of the CF component to the ear canal DPOAE. Recent work has, in fact, shown that the MOC suppresses the CF component more than the generator component in the adult cochlea (Abdala *et al.*, 2008). Its immaturity, then, might be expected to influence the CF component disproportionately.

Finally, an immature middle ear/cochlear junction would produce robust, multiple reflections within the cochlea, which may, in turn, contribute to an increase in the level of the CF component (Dhar *et al.*, 2002; Dhar and Abdala, 2007). Additionally, there is ample evidence to indicate that middle ear transmission is immature in newborns and affects both stimulus and response level (Abdala and Keefe, 2006; Keefe and Abdala, 2007). Lower effective stimulus levels could produce a relatively enhanced CF component in newborns, consistent with deeper fine structure for lower level stimuli in adult ears (Dhar *et al.*, 2005; He and Schmiedt, 1993; Konrad-Martin *et al.*, 2001, 2002). While it is true that deeper DPOAE fine structure reflects relatively equal component levels and could well be achieved by a reduced generator component, we did not observe significantly deeper fine structure in newborns, nor was the newborn generator component reduced, arguing against middle ear immaturities playing a major role in our findings.

The larger CF component in newborns and the similarity of the generator component between the two age groups suggest different maturational time lines for generator and CF

components and the physiological mechanisms responsible for their generation. Differences in the composition of the ear canal DPOAE between newborns and adults may have clinical implications. If the CF component (largely generated by a reflection mechanism) is relatively prominent in newborns, should either transient or stimulus frequency OAEs, also generated by the reflection mechanism, be more applicable in this population? If DPOAEs are then to be used, should the stimulus parameters be altered to bias the ear canal DPOAE toward the CF component? Of course, the oft-speculated greater sensitivity of reflection-source OAEs to cochlear pathology will have to be definitively demonstrated to support these types of clinical modifications.

B. Variability among reports

In a past report (Dhar and Abdala, 2007), we found slightly broader frequency spacing in the DPOAE fine structure of newborns compared to adults, which is in contrast to the present findings. Other aspects of the studies, such as the increased fine structure prevalence in newborns and the higher DPOAE levels, were in good agreement. There are some significant methodological differences between the two studies that might help explain these discrepancies, as well as inconsistencies in all reports of DPOAE fine structure among laboratories.

1. Range of test frequencies

In the present report, there is a marked difference in the number of observations available in the three highest-frequency bands tested (2500–3000, 3000–3500, and 3500–4000 Hz) between adults and infants. There is roughly twice the number of data points for newborns in the higher-frequency intervals, where the mean spacing ratio ($f/\Delta f$) is 17.2. Thus, the newborn data in the current study have a biased distribution toward the high frequencies and narrower spacing relative to adults. These high-frequency bands were not evaluated in our previous report, which was tested only through 2500 Hz. When the high-frequency data are removed, there is no age difference in DPOAE fine structure frequency spacing between ages in this data set.

2. Analysis scheme

Algorithms and analysis strategies for the classification of DPOAE fine structure vary from laboratory to laboratory and report to report. Clearly, the type of analysis routine applied can influence the measurements of fine structure prevalence, spacing, and depth significantly. Some fine structure features, such as prevalence and spacing, have been relatively consistent across laboratories for human adults (Abdala *et al.*, 2008; Dhar and Abdala, 2007; Engdahl and Kemp, 1996; He and Schmiedt, 1993; Heitmann *et al.*, 1998; Reuter and Hammershoi, 2006). However, there are too few reports of *infant* fine structure to know whether this same stability is to be expected. The higher and more erratic noise floors associated with infant testing may increase the likelihood that fine structure features are less stable in newborns across reports.

More importantly, since the early report from our joint laboratories (Dhar and Abdala, 2007), we have continued to streamline and strengthen our analysis techniques. The classification of DPOAE fine structure appears to be an evolving effort, as noted by reports dedicated solely to this task (Heise *et al.*, 2008). Whereas we processed all DPOAE magnitude and phase data through an automatic algorithm and did not treat the data further in our previous study, our current protocol includes additional steps of analysis (a visual check and the elimination of outliers). Any of these changes in our analysis routine may have contributed to differences among reports. Until there is a universally accepted strategy and algorithm for quantifying the features of DPOAE fine structure, these differences among studies will be inevitable.

3. Other methodological factors

A newer, swept-tone methodology was employed with infants in the current study giving 2–6 Hz frequency resolution, whereas discrete frequency recording was conducted with infants in the 2007 report. Although these two methods have been shown to be comparable in adults (Long *et al.*, 2008), such equivalence has not been demonstrated in infants. Lastly, an ER10C probe microphone was used for neonates only by Dhar and Abdala (2007), whereas an ER10B + was used for both adults and infants in this report. Although anecdotal data from our laboratories suggest that this probe difference will not affect fine structure spacing calculations, differences between the ER10B+ and ER10C have been documented in our laboratories with respect to several DPOAE features such as fine structure prevalence and DPOAE level.

C. Separating DPOAE components

Various methods of separating DPOAE components have been proposed and reported in literature. None of these methods are perfect mostly due to the complexity of DPOAE generation and its variability among individuals and across frequency. For example, the suppressor method, where a suppressor tone proximal in frequency to $2f_1-f_2$ is used to eliminate the CF component and isolate the generator component, is critically dependent on finding the optimal suppressor level for each individual and each frequency within individuals (Dhar and Shaffer, 2004; Johnson *et al.*, 2006, 2007). The method of time windowing used in this report relies on the differences in phase behavior between the two DPOAE components, with a steep phase gradient linked to the CF component. Scaling symmetry ensures that the phase of the generator component is relatively invariant as a function of frequency and, hence, can be distinguished from the CF component. A critical complication may arise when scaling symmetry breaks down, as is the case below 2000 Hz in our data set. Note from Fig. 4 that the phase gradient of the ear canal DPOAE is steep in low frequencies, even when the generator component is dominant in the ear canal. Consequently, at the output of an inverse FFT, some undefined portion of the generator component would be grouped with the CF component. This could produce underestimation of the generator and overestimation of the CF component. It

could also produce a more shallow CF phase gradient. Given this complication, the suppressor method of separating the DPOAE components might be better suited in the low-frequency range. However, this would require stimulus sweeps with and without the suppressor in newborns, making a challenging test protocol even more difficult. It appears warranted to consider alternative means of component separation in infants where the low-frequency break from scaling symmetry may be more marked.

V. CONCLUSIONS

Newborns show narrower fine structure spacing, a steeper DPOAE phase gradient (prolonged group delay) in the more apical regions of the cochlea, and enhanced contribution from the CF component relative to adults. These results strongly suggest an immaturity in the peripheral auditory system of newborns, although its source is not clear. It is likely that partly it can be explained by immaturity in the conductive system of newborns. Findings may also suggest immaturity in passive vibration of the basilar membrane and frequency representation in the cochlear apex of newborns. Future work would do well to scrutinize the infant apical region to further define these non-adult-like features. Additionally, it will be important to continue investigating the relative source contribution to infant DPOAEs and possibly refine OAE protocol for more effective evaluation in this age group. Finally, DPOAE fine structure is not being classified in a consistent manner across studies and laboratories. Although the features of fine structure will continue to be of some interest, it may be more productive to separate and analyze components contributing to the DPOAE, rather than focus on the interference pattern produced by their interaction. However, methods of component separation in newborns should be carefully considered before application, as complexities associated with DPOAE generation make their utility less than straightforward.

ACKNOWLEDGMENTS

This work was supported by a grant from the National Institutes of Health, NIDCD (Grant No. DC003552), the House Ear Institute, and Northwestern University. Authors would like to thank Tracy Williams and Srikanta Mishra for their assistance with data collection and management, Dr. Rangasamy Ramanathan for his continued support of neonatal hearing research at the University of Southern California Medical Center and Ping Luo for his technical assistance. Authors would also like to thank Chris Shera for his contributions with respect to phase calibration and correction. Data collection was conducted using software developed by Carrick L. Talmadge.

Abdala, C. (**1996**). "Distortion product otoacoustic emission ($2f_1-f_2$) amplitude as a function of f_2/f_1 frequency ratio and primary tone level separation in human adults and neonates," J. Acoust. Soc. Am. **100**, 3726–3740.

Abdala, C., and Keefe, D. H. (**2006**). "Effects of middle-ear immaturity on distortion product otoacoustic emission suppression tuning in infant ears," J. Acoust. Soc. Am. **120**, 3832–3842.

Abdala, C., Keefe, D. H., and Oba, S. (**2007**). "A longitudinal study of DPOAE suppression tuning and acoustic admittance in human infants from birth through six months of age," J. Acoust. Soc. Am. **121**, 3617–3627.

Abdala, C., Ma, E., and Sininger, Y. (**1999**). "Maturation of medial efferent system function in humans," J. Acoust. Soc. Am. **105**, 2392–2402.

Abdala, C., Mishra, S. K., and Williams, T. L. (**2009**). "Considering distortion product otoacoustic emission fine structure in measurements of the medial olivocochlear reflex," J. Acoust. Soc. Am. **125**, 1584–1594.

Bowman, D. M., Brown, D. K., Eggermont, J. J., and Kimberley, B. P. (**1997**). "The effect of sound intensity on f_1-sweep and f_2-sweep distortion product otoacoustic emissions phase delay estimates in human adults," J. Acoust. Soc. Am. **101**, 1550–1559.

Bredberg, G. (**1968**). "Cellular pattern and nerve supply of the human organ of Corti," Acta Oto-Laryngol., Suppl. **236**, 1–135.

Burns, E. M., Arehart, K. H., and Campbell, S. L. (**1992**). "Prevalence of spontaneous otoacoustic emissions in neonates," J. Acoust. Soc. Am. **91**, 1571–1575.

Dhar, S., and Abdala, C. (**2007**). "A comparative study of DPOAE fine structure in human newborns and adults with normal hearing," J. Acoust. Soc. Am. **122**, 2191–2202.

Dhar, S., Long, G. R., Talmadge, C. L., and Tubis, A. (**2005**). "The effect of stimulus frequency ratio on distortion product otoacoustic emission components," J. Acoust. Soc. Am. **117**, 3766–3776.

Dhar, S., and Shaffer, L. A. (**2004**). "Effects of a suppressor tone on distortion product otoacoustic emissions fine structure: Why a universal suppressor level is not a practical solution to obtaining single-generator DP-grams," Ear Hear. **25**, 573–585.

Dhar, S., Talmadge, C. L., Long, G. R., and Tubis, A. (**2002**). "Multiple internal reflections in the cochlea and their effect on DPOAE fine structure," J. Acoust. Soc. Am. **112**, 2882–2897.

Engdahl, B., and Kemp, D. T. (**1996**). "The effect of noise exposure on the details of distortion product otoacoustic emissions in humans," J. Acoust. Soc. Am. **99**, 1573–1587.

He, J., and Schmiedt, R. A. (**1993**). "Fine structure of the $2f_1-f_2$ acoustic distortion product: Changes with primary level," J. Acoust. Soc. Am. **94**, 2659–2669.

Heise, S. J., Verhey, J. L., and Mauermann, M. (**2008**). "Automatic screening and detection of threshold fine structure," Int. J. Audiol. **47**, 520–532.

Heitmann, J., Waldmann, B., Schnitzler, H. U., Plinkert, P. K., and Zenner, H. P. (**1998**). "Suppression of distortion product otoacoustic emissions (DPOAE) near f_1-f_2 removes DP-gram fine structure—Evidence for a secondary generator," J. Acoust. Soc. Am. **103**, 1527–1531.

Howard, M. A., Stagner, B. B., Foster, P. K., Lonsbury-Martin, B. L., and Martin, G. K. (**2003**). "Suppression tuning in noise-exposed rabbits," J. Acoust. Soc. Am. **114**, 279–293.

Johnson, T. A., Neely, S. T., Kopun, J. G., Dierking, D. M., Tan, H., Converse, C., Kennedy, E., and Gorga, M. P. (**2007**). "Distortion product otoacoustic emissions: Cochlear-source contributions and clinical test performance," J. Acoust. Soc. Am. **122**, 3539–3553.

Johnson, T. A., Neely, S. T., Kopun, J. G., and Gorga, M. P. (**2006**). "Reducing reflected contributions to ear-canal distortion product otoacoustic emissions in humans," J. Acoust. Soc. Am. **119**, 3896–3907.

Keefe, D. E., and Abdala, C. (**2007**). "Theory of forward and reverse middle-ear transmission applied to otoacoustic emissions in infant and adult ears," J. Acoust. Soc. Am. **121**, 978–993.

Konrad-Martin, D., Neely, S. T., Keefe, D. H., Dorn, P. A., Cyr, E., and Gorga, M. P. (**2002**). "Sources of DPOAEs revealed by suppression experiments, inverse fast Fourier transforms, and SFOAEs in impaired ears," J. Acoust. Soc. Am. **111**, 1800–1809.

Konrad-Martin, D., Neely, S. T., Keefe, D. H., Dorn, P. A., and Gorga, M. P. (**2001**). "Sources of distortion product otoacoustic emissions revealed by suppression experiments and inverse fast Fourier transforms in normal ears," J. Acoust. Soc. Am. **109**, 2862–2879.

Lavigne-Rebillard, M., and Pujol, R. (**1986**). "Development of the auditory hair cell surface in human fetuses. A scanning electron microscopy study," Anat. Embryol. (Berl) **174**, 369–377.

Lavigne-Rebillard, M., and Pujol, R. (**1987**). "Surface aspects of the developing human organ of Corti," Acta Oto-Laryngol., Suppl. **436**, 43–50.

Lavigne-Rebillard, M., and Pujol, R. (**1988**). "Hair cell innervation in the fetal human cochlea," Acta Oto-Laryngol. **105**, 398–402.

Long, G. R., Talmadge, C. L., and Lee, J. (**2008**). "Measuring distortion product otoacoustic emissions using continuously sweeping primaries," J. Acoust. Soc. Am. **124**, 1613–1626.

Mauermann, M., Uppenkamp, S., van Hengel, P. W. J., and Kollmeier, B. (**1999a**). "Evidence for the distortion product frequency place as a source

of distortion product otoacoustic emission (DPOAE) fine structure in humans. II. Fine structure for different shapes of cochlear hearing loss," J. Acoust. Soc. Am. **106**, 3484–3491.

Mauermann, M., Uppenkamp, S., van Hengel, P. W. J., and Kollmeier, B. (**1999b**). "Evidence for the distortion product frequency place as a source of distortion product otoacoustic emission (DPOAE) fine structure in humans. I. Fine structure and higher-order DPOAE as a function of the frequency ratio f_2/f_1 ," J. Acoust. Soc. Am. **106**, 3473–3483.

Mills, D. M., and Rubel, E. W. (**1996**). "Development of the cochlear amplifier," J. Acoust. Soc. Am. **100**, 428–441.

Moore, J. K., Simmons, D. D., and Guan, Y. (**1999**). "The human olivocochlear system: Organization and development," Audiol. Neuro-Otol. **4**, 311–325.

Morlet, T., Collet, L., Salle, B., and Morgon, A. (**1993**). "Functional maturation of cochlear active mechanisms and of the medial olivocochlear system in humans," Acta Oto-Laryngol. **113**, 271–277.

Moulin, A., and Kemp, D. T. (**1996a**). "Multicomponent acoustic distortion product otoacoustic emission phase in humans. I. General characteristics," J. Acoust. Soc. Am. **100**, 1617–1639.

Moulin, A., and Kemp, D. T. (**1996b**). "Multicomponent acoustic distortion product otoacoustic emission phase in humans. II. Implications for distortion product otoacoustic emissions generation," J. Acoust. Soc. Am. **100**, 1640–1662.

Overstreet, E. H., III, Temchin, A. N., and Ruggero, M. A. (**2002**). "Passive basilar membrane vibrations in gerbil neonates: Mechanical bases of cochlear maturation," J. Physiol. **545**, 279–288.

Pujol, R., Lavigne-Rebillard, M., and Lenoir, M. (**1998**). "Development of sensory and neural structures in the mammalian cochlea," in *Development of the Auditory System*, edited by E. Rubel, A. Popper, and R. Fay (Springer, New York), pp. 146–192.

Reuter, K., and Hammershoi, D. (**2006**). "Distortion product otoacoustic emission fine structure analysis of 50 normal-hearing humans," J. Acoust. Soc. Am. **120**, 270–279.

Ruggero, M. A. (**2004**). "Comparison of group delays of $2f_1-f_2$ distortion product otoacoustic emissions and cochlear travel times," ARLO **5**, 143–147.

Ryan, S., and Piron, J. (**1994**). "Functional maturation of the medial efferent olivocochlear system in human neonates," Acta Oto-Laryngol. **114**, 485–489.

Shera, C. A. (**2003**). "Mammalian spontaneous otoacoustic emissions are amplitude-stabilized cochlear standing waves," J. Acoust. Soc. Am. **114**, 244–262.

Shera, C. A., and Guinan, J. J. (**1999**). "Evoked otoacoustic emissions arise by two fundamentally different mechanisms: A taxonomy for mammalian OAEs," J. Acoust. Soc. Am. **105**, 782–798.

Shera, C. A., Talmadge, C., and Tubis, A. (**2000**). "Interrelations among distortion-product phase-gradient delays: Their connection to scaling symmetry and its breaking," J. Acoust. Soc. Am. **108**, 2933–2948.

Shera, C. A., and Zweig, G. (**1993b**). "Order from chaos: Resolving the paradox of periodicity in evoked otoacoustic emission," in *Biophysics of Hair Cell Sensory Systems*, edited by H. Duifhuis, J. W. Horst, P. van Dijk, and S. M. van Netten (World Scientific, Singapore), pp. 54–63.

Siegel, J. H., Cerka, A. J., Recio, A., Temchin, A. N., 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**, 2434–43.

Talmadge, C. L., Long, G. R., Tubis, A., and Dhar, S. (**1999**). "Experimental confirmation of the two-source interference model for the fine structure of distortion product otoacoustic emissions," J. Acoust. Soc. Am. **105**, 275–292.

Talmadge, C. L., Tubis, A., Long, G. R., and Piskorski, P. (**1998**). "Modeling otoacoustic emission and hearing threshold fine structures," J. Acoust. Soc. Am. **104**, 1517–1543.

Tubis, A., Talmadge, C. L., Tong, C., and Dhar, S. (**2000**). "On the relationships between the fixed-f_1, fixed-f_2, and fixed-ratio phase derivatives of the $2f_1-f_2$ distortion product otoacoustic emission," J. Acoust. Soc. Am. **108**, 1772–1785.

Withnell, R. H., Shaffer, L. A., and Talmadge, C. L. (**2003**). "Generation of DPOAEs in the guinea pig," Hear. Res. **178**, 106–117.

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.

Forward-masking patterns produced by symmetric and asymmetric pulse shapes in electric hearing

Olivier Macherey[a)]
ExpORL, Department of Neurosciences, K.U. Leuven, Herestraat 49, Bus 721, 3000 Leuven, Belgium and MRC Cognition and Brain Sciences Unit, 15 Chaucer Road, Cambridge CB2 7EF, United Kingdom

Astrid van Wieringen
ExpORL, Department of Neurosciences, K.U. Leuven, Herestraat 49, Bus 721, 3000 Leuven, Belgium

Robert P. Carlyon
MRC Cognition and Brain Sciences Unit, 15 Chaucer Road, Cambridge CB2 7EF, United Kingdom

Ingeborg Dhooge
Department of ORL, Gent University, De Pintelaan 185, 9000 Gent, Belgium

Jan Wouters
ExpORL, Department of Neurosciences, K.U. Leuven, Herestraat 49, Bus 721, 3000 Leuven, Belgium

(Received 30 August 2008; revised 30 September 2009; accepted 1 October 2009)

Two forward-masking experiments were conducted with six cochlear implant listeners to test whether asymmetric pulse shapes would improve the place-specificity of stimulation compared to symmetric ones. The maskers were either cathodic-first symmetric biphasic, pseudomonophasic (i.e., with a second anodic phase longer and lower in amplitude than the first phase), or "delayed pseudomonophasic" (identical to pseudomonophasic but with an inter-phase gap) stimuli. In experiment 1, forward-masking patterns for monopolar maskers were obtained by keeping each masker fixed on a middle electrode of the array and measuring the masked thresholds of a monopolar signal presented on several other electrodes. The results were very variable, and no difference between pulse shapes was found. In experiment 2, six maskers were used in a wide bipolar (bipolar+9) configuration: the same three pulse shapes as in experiment 1, either cathodic-first relative to the most apical or relative to the most basal electrode of the bipolar channel. The pseudomonophasic masker showed a stronger excitation proximal to the electrode of the bipolar pair for which the short, high-amplitude phase was anodic. However, no difference was obtained with the symmetric and, more surprisingly, with the delayed pseudomonophasic maskers. Implications for cochlear implant design are discussed.
 [DOI: 10.1121/1.3257231]

PACS number(s): 43.64.Me, 43.66.Dc, 43.66.Cb, 43.66.Hg [RYL] Pages: 326–338

I. INTRODUCTION

Multichannel cochlear implants (CIs) attempt to mimic the natural tonotopic encoding of the cochlea by stimulating different populations of auditory nerve fibers along the implanted electrode array. One potential limitation of contemporary devices lies in the so-called "channel interaction" phenomenon. Due to the conductive properties of the perilymph in the scala tympani, the current generated at one electrode site spreads out widely over the cochlea. As a consequence, the different stimulation channels excite overlapping neural populations and presumably degrade the spectral resolution of the sounds transmitted to the brain. Although existing CIs possess up to 22 intracochlear electrodes, little or no improvement in speech recognition is observed when the number of active electrodes is increased above approximately 8 (Fishman *et al.*, 1997; Friesen *et al.*, 2001, 2005). This may not impose a severe limitation for the perception of speech in quiet, where only a few channels are needed (Shannon *et al.*, 1995) and where many CI users perform well. However, it may impair performance in listening situations where more independent spectral channels are needed (e.g., for the perception of speech in noise or for the perception of music) and where CI users have been shown to perform poorly (Friesen *et al.*, 2001; McDermott, 2004). Several approaches have been followed to reduce channel interactions in CIs, and the present study forms part of this effort.

One approach is to place the electrodes closer to the excitable neural elements (Shepherd *et al.*, 1993). Some benefits of perimodiolar over outer-wall placements have been reported, including lower thresholds, better electrode discrimination, and narrower forward-masking profiles (Cohen *et al.*, 2001, 2006). To even further minimize the distance between the electrodes and the nerve, an alternative for future CI devices may be to implant the electrode array directly within the modiolus (Badi *et al.*, 2002; Hillman *et al.*, 2003; Middlebrooks and Snyder, 2007).

[a)]Author to whom correspondence should be addressed. Electronic mail: olivier.macherey@mrc-cbu.cam.ac.uk

Another approach is to manipulate the electrode configuration. Most clinical strategies use a monopolar configuration, which consists of stimulating each intracochlear electrode with reference to a remote (usually extracochlear) electrode. Theoretically, configurations that involve stimulation between closer *intracochlear* contacts, such as bipolar or quadrupolar, should produce a more spatially focused excitation and therefore reduce channel interactions compared to monopolar (van den Honert and Stypulkowski, 1987; Jolly *et al.*, 1996). However, behavioral measures in CI users failed to demonstrate sharper tuning or clear speech-perception benefits with bipolar or quadrupolar configurations compared to the usual monopolar one (Pfingst *et al.*, 1997; Mens and Berenstein, 2005; Kwon and van den Honert, 2006). One reason for this arises from the fact that in masking experiments, different masker types are presented at equal loudnesses. As Kwon and van den Honert (2006) pointed out, the current used for bipolar or quadrupolar configurations may need to be increased to recruit sufficient auditory nerve fibers, thereby undermining the putative increases in selectivity. For speech perception, the potential advantages of bipolar or quadrupolar stimulation may be further compromised by irregularities in neural survival patterns (Kwon and van den Honert, 2006; Bierer, 2007) and with the possible occurrence of a bimodal spatial excitation pattern in bipolar configuration (Chatterjee *et al.*, 2006; Snyder *et al.*, 2008).

A third possibility, which is the focus of the present study, is to modify the electrical waveform. It has been shown that pulse duration and pulse polarity are determinant factors of spatial selectivity in functional electrical stimulation (Grill and Mortimer, 1996; McIntyre and Grill, 1999, 2000, 2002). In contemporary CIs, electrical signals are trains of amplitude-modulated symmetric biphasic pulses consisting of two opposite-polarity phases (anodic and cathodic), which may both stimulate the nerve. Physiological studies in animals have shown that monophasic anodic and cathodic pulses initiate spikes at different locations along the nerve fibers, as shown by the longer latencies obtained with cathodic than with anodic stimulation (Miller *et al.*, 1999, 2004). This latency difference is the result of anodic pulses initiating spikes at a more central locus on the fibers than cathodic pulses. The fact that anodic stimuli initiate action potentials at a more remote site than cathodic stimuli is not specific to auditory nerve stimulation and has been observed in other extracellular neural stimulation studies (see Ranck, 1975).

To optimize the place-specificity of stimulation, action potential initiation should ideally be restrained to neural sites proximal to the electrode. The predictions of a computational model of the human cochlea developed by Rattay *et al.* (2001) suggest that this might be better achieved with cathodic stimuli. These authors first calculated the extracellular potentials produced by a monopolar electrode along 16 modeled neurons spanning the entire cochlea. They found that the across-neuron difference in extracellular voltage along their axis was larger at their peripheral end and became smaller at more central locations. This trend is due to the specific geometry of cochlear neurons, which get closer to each other as they extend from their peripheral terminal to the modiolus. The model of Rattay *et al.* (2001) further predicts that cathodic stimuli would produce a large depolarization (and therefore elicit spikes) at the level of the peripheral processes of most neurons. Because these neurons receive different contributions from the electrode (large contribution for neurons right next to the electrode and smaller contributions for more distant ones), even two neighboring neurons have very different thresholds. In contrast, for anodic stimuli, the largest depolarization is obtained at a level central to the cell body, where fibers are more tightly packed. The resulting effect is that neighboring neurons have much more similar thresholds and may be equally excited by the same stimulus current (see Table 2 in Rattay *et al.*, 2001). The predicted threshold differences between the several modeled neurons and the neuron closest to the electrode (neuron 7) are illustrated in Fig. 1 for both monophasic anodic and cathodic stimuli. It can be seen that the cathodic pulse does produce the more progressive spatial recruitment, i.e., for x dB above the threshold of neuron 7, the stimulus level is suprathreshold for fewer neurons with a cathodic than with an anodic pulse. Based on these predictions, we expect cathodic stimulation to produce a more place-specific excitation than anodic stimulation. Note that this model of the human auditory nerve assumes that the neurons' peripheral processes are not degenerated and also predicts that the threshold of the lowest-threshold neuron (number 7) is lower for a cathodic than for an anodic pulse. Our recent findings on polarity sensitivity in human CI users contrast with these predictions as we found the anodic phase to produce a stronger masking effect than the cathodic phase when stimulating at the same current level (Macherey *et al.*, 2008). Although the reason for this trend remains unknown, one explanation could be a substantial loss of peripheral processes (see Macherey *et al.*, 2008 for discussion). In such a "degenerated scenario," the place-specificity difference between anodic and cathodic stimuli discussed above may not hold.

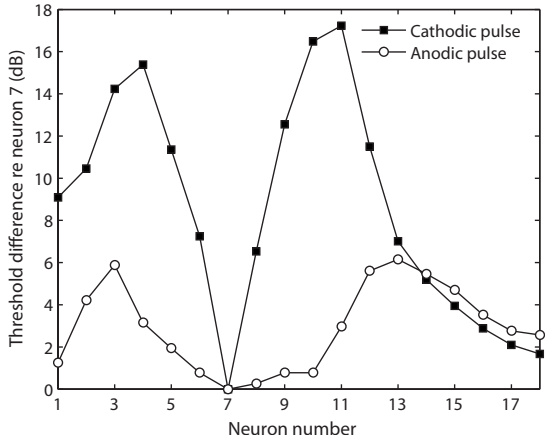


FIG. 1. Predicted threshold difference between several modeled neurons and neuron 7 for 100-microseconds anodic and cathodic monophasic pulses. The monopolar electrode is located below neuron 7 (insertion of 180°) and neurons are numbered from base to apex in 30° steps. These simulations correspond to the "short dendrite" case, replotted from Rattay *et al.* (2001, Table II p. 73).

TABLE I. Biographical data of the six Clarion CI subjects (S1–S6) with type of electrode array, age, duration of deafness (DD) and duration of CI use (CI) in years (y), and etiology.

Subject	Electrode array	Age (y)	DD (y)	CI (y)	Etiology
S1	HiFocus II	41	<0.5	3	Meningitis
S2	HiFocus II	60	3	2	Progressive
S3	HiFocus II	54	7	1	Progressive
S4	Helix	42	4	1	Progressive
S5	HiFocus II	55	9	2	Unknown
S6	HiFocus II	75	0.5	2	Progressive

The spatial selectivity of monophasic cathodic pulses cannot be directly evaluated in CI users because this would impose the delivery of a net dc charge to the nerve, which can cause damage to the tissue (Shepherd *et al.*, 1999). One way to approach the situation encountered with cathodic monophasic stimulation may be to use asymmetric biphasic pulses for which the contribution of the anodic phase is reduced by making it longer and lower in amplitude (Miller *et al.*, 2001; van Wieringen *et al.*, 2005; Macherey *et al.*, 2006). If a monophasic cathodic pulse excites a more spatially restricted neural population than an anodic pulse, we expect such cathodic asymmetric pulses to provide a more spatially focused excitation than symmetric ones when presented in monopolar configuration. This hypothesis will be tested in experiment 1.

Another different reason for using asymmetric pulses was suggested by a computational model of the guinea pig cochlea (Frijns *et al.*, 1996). Frijns *et al.* argued that the use of asymmetric rather than symmetric pulses would approximately double the number of independent channels of a CI with longitudinal *bipolar* electrodes. Their model predicted that the fibers proximal to the electrode for which the short, high-amplitude phase of the pulse is cathodic would be more effectively excited than the fibers proximal to the other electrode of the pair. In contrast, with biphasic stimulation, there would be a bimodal pattern of excitation, reflecting equal stimulation at sites near each electrode. These predictions are consistent with more recent physiological recordings by Bonham *et al.* (2003), who measured neural activity along the tonotopic axis of the inferior colliculus of guinea pigs following intracochlear bipolar stimulation with asymmetric pulses. They found that the place of the peak of excitation shifted when inverting the polarity of the asymmetric pulse. This shift was also consistent with a more effective excitation of fibers near the electrode for which the short, high-amplitude phase was cathodic. This finding probably arises from the fact that for a given current level, cathodic pulses are more effective in exciting nerve fibers than anodic pulses, as found in most animal physiological studies (e.g., Miller *et al.*, 1999). However, as previously mentioned, our recent work on polarity sensitivity in human CI users showed the opposite trend, i.e., that anodic phases have a stronger masking effect than cathodic phases (Macherey *et al.*, 2008). We may therefore expect the opposite effect to the one predicted by Frijns' model and obtained by Bonham *et al.* (2003), i.e., that fibers proximal to the electrode for which the short, high-amplitude phase is *anodic* are more effectively activated. This hypothesis will be tested in experiment 2.

Channel interactions can be assessed using several psychophysical (reviewed in Shannon *et al.*, 2004) or electrophysiological techniques (Cohen *et al.*, 2003; Abbas *et al.*, 2004). In the two experiments reported here, we investigated non-simultaneous interactions using a psychophysical forward-masking paradigm (Lim *et al.*, 1989; Shannon, 1990; Cohen *et al.*, 1996; Chatterjee and Shannon, 1998; Chatterjee *et al.*, 2006; Kwon and van den Honert, 2006). Forward masking refers to the threshold shift of a signal presented after a masker stimulus. It is related to the degree to which auditory nerve fibers that are important for the detection of the signal also respond to the masker, and whose response to the signal therefore is affected by neural refractoriness at the auditory nerve and more centrally. By keeping the masker on a fixed electrode and measuring the threshold shift of the signal on different electrodes spanning the entire electrode array, it is possible to measure the spatial spread of excitation produced by a given masker.

II. GENERAL METHODS

A. Subjects

Six postlingually deafened CI users (S1–S6) participated in a series of two psychophysical forward-masking experiments. All six subjects performed experiment 1, while only four (S1–S4) performed experiment 2. All subjects had been implanted with a perimodiolar electrode array (HiFocus II for all subjects except S4, who has a Helix) manufactured by Advanced Bionics and consisting of 16 intracochlear electrodes. Each electrode contact has a rectangular shape (0.4×0.5 mm^2) and is made of platinum and iridium. Electrodes are numbered from 1 to 16, from the most apical to the most basal one. The distance between two adjacent contacts is 1.1 mm for the HiFocus II and 0.85 mm for the Helix. Table I summarizes the information for each subject, including electrode array, age, duration of deafness, duration of CI use, and etiology of deafness. Testing was approved by the K.U. Leuven Medical Ethical Committee and was in accordance with the Declaration of Helsinki. Subjects were paid for participating.

B. Stimuli

Forward-masked thresholds were measured for three different masker shapes in both monopolar (experiment 1) and bipolar (experiment 2) configurations (Fig. 2). The three

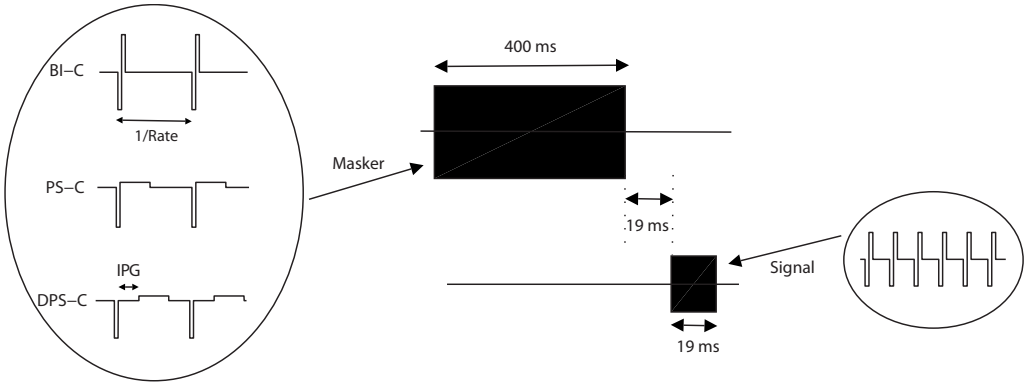


FIG. 2. Schematic representation of the stimuli used in the forward-masking experiments.

shapes included a symmetric biphasic cathodic-first (BI-C), a pseudomonophasic cathodic-first (PS-C), and a delayed pseudomonophasic cathodic-first (DPS-C) stimulus. Each of these three stimuli had a total duration of 400 ms and a rate of 104 pps, and the duration of the cathodic phase was always 97 μs. The duration of the second (anodic) phase was also 97 μs for the BI-C stimulus, whereas for the PS-C and DPS-C pulses, it was eight times longer, with an amplitude reduced by the same factor in order to maintain charge-balancing. The inter-phase gap (IPG) of DPS-C was set to 4.3 ms to present the second phase approximately midway between the first phases of two consecutive pulses. Each current-source of the implant is coupled to a dc-blocking capacitor, which induces a current flow during the inter-phase and inter-pulse gaps of DPS-C. The amplitude of this current flow was similar to that previously measured by Macherey *et al.* (2006), i.e., about 50 dB lower than the pulse amplitude.

The three stimuli were used as maskers of a 19-ms BI-C signal presented 19 ms after the end of each masker (i.e., after the end of the second phase of the last masker pulse). The signal was always presented in monopolar mode (for both monopolar and bipolar maskers), had the same phase duration as the masker (97 μs), and had a much higher rate of 937 pulses/s. We used a monopolar signal to avoid the marked variations in absolute sensitivity that can occur across the electrode array with bipolar stimulation (e.g., Bierer, 2007). Unlike previous forward-masking studies with CIs, which used identical masker and signal, we used different stimulation rates to avoid possible confusion effects, as suggested by psychoacoustical data (Moore and Glasberg, 1982; Neff, 1985). Specifically, if the signal is presented on the same electrode as the masker and has the same pulse rate, it is possible that it will be perceived as a mere continuation of the masker. This could prevent the signal from being detected, even when it provides substantial stimulation of the auditory nerve. Furthermore, when the signal and masker are presented on different electrodes, these confusion effects would be expected to decrease, leading to an over-estimate of the degree of tonotopic selectivity (Moore *et al.*, 1984).

The stimuli were presented through direct stimulation of the implant via the clinical programming interface provided by the implant manufacturer. The psychophysical tests were performed using the APEX software platform (Laneau *et al.*, 2005) and the BEDCS software developed by Advanced Bionics (Litvak, 2003). As previously described in Macherey *et al.* (2006), a finite number of current-level values can be accessed using this equipment. Specifically, the current level is coded on 8 bits, and the minimal current-step that can be used depends on the current range that is selected. The following experiments were performed either in the 0–255 μA or in the 0–510 μA current range. These two ranges have different minimal current-step sizes of 1 and 2 μA, respectively. For asymmetric pulses the minimal current-step was increased by a factor of 8, leading to current-step sizes of 8 and 16 μA for current ranges of 0–255 and 0–510 μA, respectively (see Macherey *et al.*, 2006).

III. EXPERIMENTS

A. Experiment 1: Forward masking in monopolar configuration

1. Design and methods

In this experiment, we measured forward-masking patterns produced by the three maskers described in Fig. 2 (BI-C, PS-C, and DPS-C). The maskers were presented in monopolar configuration on an electrode located in the middle of the array (electrode 9) with reference to the case (extracochlear) electrode of the implant. If the cathodic phase produces a more spatially compact neural excitation than the anodic phase, then asymmetric pulses (PS-C and DPS-C) should yield a narrower forward-masking pattern than BI-C. We used DPS-C in addition to PS-C because DPS-C has the advantage of requiring much less current than BI-C or PS-C to evoke the same loudness (Macherey *et al.*, 2006). In the following paragraphs, the term "channel" will refer to the active electrode of the monopolar channel.

A preliminary aim of this experiment was to equate the effectiveness of the three masker types for a signal presented on the same channel, so that it is possible to directly compare the masked threshold differences produced by moving the signal to more distant locations. The experimental design was therefore divided into six steps:

(i) The MCL was first determined for each masker type.
(ii) Each masker was then set to its MCL, and we determined the masked threshold of the BI-C signal when presented on the same channel as the masker. This

TABLE II. Current level (in dB re 1 mA) used in experiments 1 and 2. For experiment 1 (left part), the levels were determined so that the three masker shapes would mask as effectively a signal presented on the masker channel (channel 9). For experiment 2, the levels were loudness-balanced at MCL. S5 and S6 did not participate in experiment 2. The last two rows show the mean levels of subjects S1–S4 who participated in both experiments and of subjects S1–S6 only for experiment 1.

	Experiment 1			Experiment 2					
Subject	Masker BI-C	Masker PS-C	Masker DPS-C	BI-A	BI-C	PS-A	PS-C	DPS-A	DPS-C
S1	−19.6	−19	−28	−16.5	−16.9	−17.3	−18.4	−22.9	−22.9
S2	−12	−10.3	−13.6	−12	−12	−11.3	−12.1	−12.7	−13.3
S3	−17.7	−15.1	−17.3	−16.8	−17.1	−17.9	−16.8	−22.9	−22.9
S4	−7.5	−9.5	−13	−13.6	−13.6	−13.6	−14.3	−17.3	−17.3
S5	−14.7	−14.3	−19.7	…	…	…	…	…	…
S6	−15.5	−15.1	−22.9	…	…	…	…	…	…
Mean (S1–S4)	−14.2	−13.5	−18	−14.7	−14.9	−15	−15.4	−19	−19.1
Mean (S1–S6)	−14.5	−13.9	−19.1	…	…	…	…	…	…

was done using a two alternative forced choice, 2-down, 1-up adaptive procedure (Levitt, 1971). The starting level of the signal was set to be clearly audible, usually at the most comfortable level (MCL) of the subject, and was subsequently *decreased* by steps of four (for S1, S3, S5, and S6) or eight (for S2 and S4) current-steps following correct responses of the subject. After two reversals, the adjustment step size was decreased to one current-step for all subjects. Each run contained eight reversals, and the mean of the last six reversals was calculated as the estimate.

(iii) The signal level was then set to the minimum of the three masked threshold values obtained in (ii), and the level of each *masker* was adjusted so that it just masked the signal. An adaptive procedure identical to (ii) was used, except that, here, the level of the masker *increased* following correct responses of the subject. The starting level was chosen to be soft enough not to mask the signal. The initial step size was one current-step for asymmetric pulses and eight current-steps for symmetric pulses. After two reversals, it was further reduced or maintained to one current-step for both symmetric and asymmetric pulses.

(iv) The resulting masker levels were averaged over four estimates, and we further checked that all of them would achieve approximately the same amount of masking by repeating step (ii) using the masker levels found in (iii).

(v) Once the three maskers were equated, we measured the masked thresholds of the signal on eight different channels spanning the entire electrode array (channels 2, 4, 6, 8, 10, 12, 14, and 16), the masker channel (9) being fixed. Each data point represents the average of three (for S5 and S6) to six (S1–S4) estimates using the same adaptive procedure as in (ii).

(vi) Finally, the absolute thresholds of the signal were also measured on the nine channels and averaged over three repetitions using a procedure identical to (v), except that no masker was present.

The advantages and limitations of this method over those previously used will be discussed further in Sec. IV A. For the meantime, it is worth noting that the maskers are equated not for loudness, but for the amount of masking produced at a single site. This avoids the situation of having a masker that produces a narrow spread being increased in level so that its loudness matches that of the others and of this loudness increase being bought at the expense of a broader spread resulting from the increased current.

2. Results

The masker equating procedure resulted in the masker levels displayed in Table II. Consistent with previous reports (Macherey *et al.*, 2006), these are usually lowest for the DPS-C stimulus. Figure 3 shows the absolute and masked thresholds for the six subjects and the three different masker types. As expected, the masked thresholds on the masker channel (9) are similar for the three masker types. Because the pattern of absolute thresholds differed across subjects, the same data are replotted in terms of amount of masking in Fig. 4. The functions show, in general, a maximum of excitation on the masker channel and a decrease as the signal channel is more spatially remote from the masker channel. For S4, the maximum of excitation is located at a more basal region (channel 10 or 12; see Figure 4). This may reflect the particular threshold trend of this subject, who shows lower absolute thresholds near the base (Fig. 3). The shift in masking peak may therefore indicate that fibers at a more basal location than electrode 9 are more effectively masked than fibers proximal to electrode 9 due to, e.g., a specific pattern of neural survival or a "kink" in the electrode array. Furthermore, the absolute amount of threshold shift was variable across subjects. For example, the difference in excitation between channels 9 and 16 was only about 3 dB for S1 and 9 dB for S2 (Fig. 3).

Two-way repeated-measure analyses of variance (ANOVAs) were conducted on the mean data of the six subjects. As we aimed to compare the BI-C excitation pattern with the other two, we performed two different analyses. One analysis compared the masked thresholds of BI-C and PS-C, and the other compared the masked thresholds of BI-C and DPS-C. The treatment factors were the masker type and the signal channel. If the hypothesis that asymmetric pulses are more selective than symmetric pulses is confirmed, this should be reflected by a significant interaction between masker type

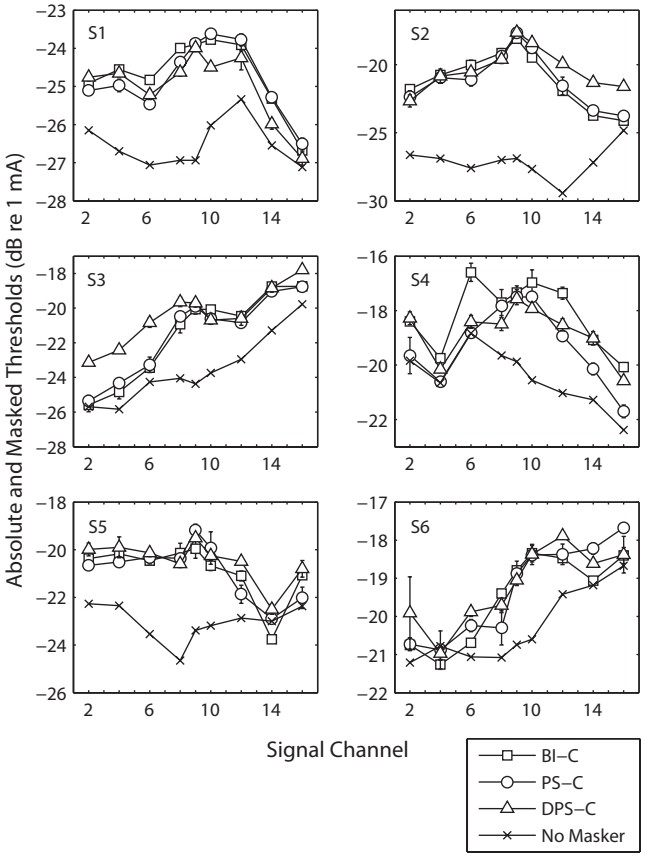


FIG. 3. Absolute and masked thresholds of a signal presented on different channels and after the three different monopolar maskers (BI-C, PS-C, and DPS-C). The error bars represent the standard errors, as in the following figures.

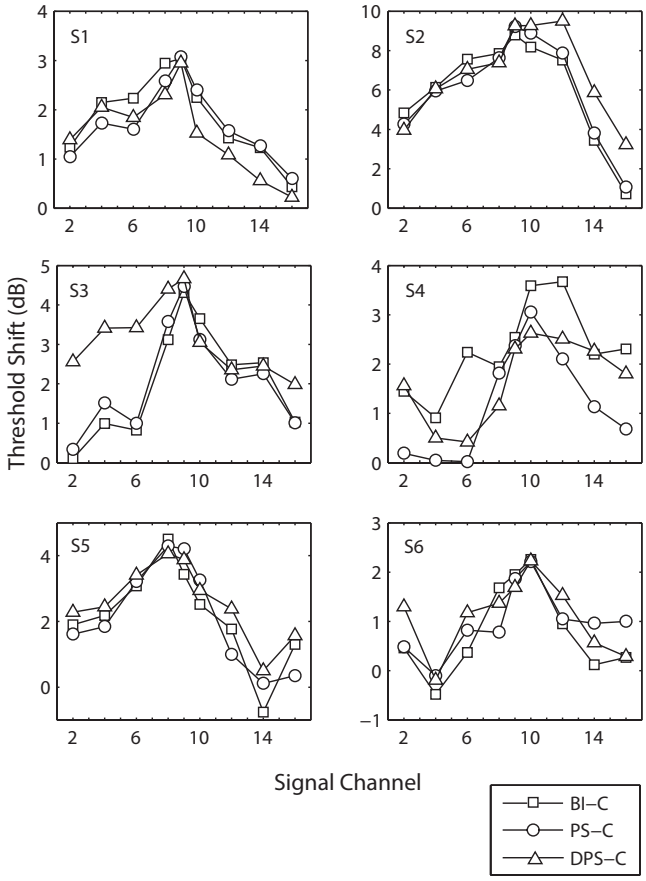


FIG. 4. Mean threshold shift (difference between masked and absolute thresholds) of a monopolar BI-C signal presented on different channels and after the three different monopolar maskers.

and signal channel as the amount of masking produced by symmetric and asymmetric maskers should differ more and more when the signal channel is moved further from the masker channel. For both analyses, there was a significant between-subjects effect [$F(1,5)=11.2$, $p=0.02$ for BI-C/PS-C and $F(1,5)=11.6$, $p=0.02$ for BI-C/DPS-C]. However, there was no effect of masker type [$F(1,5)=0.78$, $p=0.42$ for BI-C/PS-C and $F(1,5)=0.78$, $p=0.42$ for BI-C/DPS-C], nor of the interaction between masker type and signal channel [$F(1,5)=1.26$, $p=0.31$ for BI-C/PS-C and $F(1,5)=1.26$, $p=0.31$]. The only significant contributor was the signal channel, reflecting the obvious fact that there is usually more masking near the masker channel than at other cochlear locations, whatever the masker type is. Therefore, the hypothesis that asymmetric pulses with a long, low-amplitude anodic phase would improve spatial focusing compared to symmetric pulses in monopolar configuration is not supported by these data.

Because of the large inter-subject variability, the data were also analyzed statistically for each subject separately, again using two-way repeated-measure ANOVAs (with each masked threshold estimate counting as a repetition). When the interaction between masker type and signal channel was not significant, subsequent paired-sample *t*-tests were performed on individual channels using the Bonferroni correction. A statistical summary is given in Table III.

S1 and S4 demonstrated a sharper excitation pattern for PS-C than for BI-C, as demonstrated by a significant effect of masker type. Although no significant interaction between masker type and channel was found, paired-sample *t*-tests (df=5) revealed significant differences on channel 6 ($p=0.012$) for S1 and on channels 4 ($p=0.009$), 6 ($p=0.009$), 12 ($p=0.018$), 14 ($p=0.009$), and 16 ($p=0.018$) for S4. For the other four subjects (S2, S3, S5, and S6), the BI-C and PS-C patterns were very similar, and the effect of masker type was not significant.

The BI-C and DPS-C maskers produced significantly different patterns in five of the six subjects (all except S5). For S6, none of the differences observed on individual channels (6 and 14) survived the Bonferroni correction. DPS-C led to significantly less masking on channels 10 ($p<0.001$) and 14 ($p=0.045$) for S1 and on channel 6 ($p=0.045$) for S4. However, for S2 and S3, the opposite was true; i.e., the DPS-C pattern was significantly broader (showing more masking) than the BI-C pattern. For these two subjects, one side of the DPS-C pattern (the basal side for S2 and the apical side for S3) had a shallower slope than the BI-C pattern. This observation was corroborated by the statistical tests, which revealed a significant effect of the interaction between masker type and channel for the two subjects S2 and S3. Moreover, paired-sample *t*-tests revealed significant differences on channels 9 ($p=0.008$), 10 ($p=0.013$),

TABLE III. Summary of the statistical analyses performed on individual subjects' data in experiment 1. The statistical tests were two-way repeated-measure ANOVAs. For S1–S4, six repetitions of each measure had been performed, and degrees of freedom were (1,5). For S5 and S6, some measures only had three repetitions, and only the last three repetitions of each condition were used. Degrees of freedom were (1,2). Statistically significant effects are shown in bold type.

		Comparison BI/PS		Comparison BI/DPS	
Subject	Treatment factor	*F*	*p*	*F*	*p*
S1	Masker type	6.747	**0.048**	16.724	**0.009**
	Channel	99.793	**<0.001**	82.908	**<0.001**
	Masker type * channel	2.318	0.188	2.162	0.201
S2	Masker type	0.007	0.935	20.466	**0.006**
	Channel	71.112	**<0.001**	72.134	**<0.001**
	Masker type * channel	2.323	0.188	14.113	**0.013**
S3	Masker type	0.104	0.76	42.837	**0.001**
	Channel	178.274	**<0.001**	140.71	**<0.001**
	Masker type * channel	1.433	0.285	11.889	**0.018**
S4	Masker type	118.577	**<0.001**	16.289	**0.01**
	Channel	40.105	**0.001**	38.838	**0.002**
	Masker type * channel	3.711	0.112	3.461	0.122
S5	Masker type	0.049	0.845	4.942	0.156
	Channel	35.989	**0.027**	22.062	**0.042**
	Masker type * channel	2.597	0.248	2.313	0.268
S6	Masker type	3.307	0.211	51.377	**0.019**
	Channel	40.028	**0.024**	25.159	**0.038**
	Masker type * channel	4.746	0.161	1.082	0.408

12 ($p=0.002$), 14 ($p<0.001$), and 16 ($p<0.001$) for S2 and on channels 2 ($p=0.002$), 4 ($p=0.001$), 6 ($p<0.001$), 9 ($p=0.022$), and 16 ($p=0.025$) for S3.[1]

In summary, there appeared to be no consistent advantages of using asymmetric rather than symmetric pulses in monopolar configuration. Although some improvements were sometimes obtained on a few channels for some of the subjects using PS-C instead of BI-C, it is worth noting that asymmetric pulses could also *degrade* spatial selectivity, as shown by the trends of S2 and S3 for the DPS-C masker.

B. Experiment 2: Forward masking in bipolar configuration

1. Design and methods

As already mentioned in Sec. I, Frijns *et al.* (1996) suggested that improved spatial selectivity may be obtained using pseudomonophasic pulses in longitudinal bipolar configuration because such stimuli should focus the stimulation near one electrode of the pair. To investigate this hypothesis, we performed a second forward-masking experiment with four subjects (S1–S4) using maskers presented in a relatively wide bipolar configuration (bipolar+9). We chose such a large separation to maximize the probability of obtaining a bimodal excitation pattern with symmetric biphasic pulses (Chatterjee *et al.*, 2006; Snyder *et al.*, 2008), i.e., two peaks of excitation near each electrode of the bipolar pair. The two electrodes (4, 14) were selected because they showed similar absolute threshold values of the signal for three of the subjects (S1, S2, and S4). Given the particular threshold trend of subject S3 (Fig. 3), it was not possible to find two electrodes widely separated with similar absolute thresholds, and we used the same electrode pair as for the others. Six maskers were evaluated: BI-C, PS-C, and DPS-C (cathodic-first relative to electrode 4), and the same three stimuli reversed in polarity, which are named BI-A, PS-A, and DPS-A (anodic-first relative to electrode 4, i.e., cathodic-first relative to electrode 14).

The six masker stimuli were loudness-balanced at MCL. This was done by first determining the MCL for the BI-A masker and further balancing the other five maskers relative to BI-A (see Table II). For each loudness adjustment, two stimuli were continuously presented to the subjects, the first being the reference, constant across the experiment, the second being the target to balance, with a gap of approximately 500 ms in between. Subjects could adjust the level of the second stimulus by pressing one of six buttons (three to increase and three to decrease), each button corresponding to a different step size (one, two, or three current-steps), until they perceived the levels to sound equal. For each pair of stimuli that needed to be compared, at least four loudness-balanced measures were performed.

We subsequently measured the masked thresholds of the same monopolar BI-C signal as used in experiment 1 presented on each of the two masker electrodes and for the six different maskers. For each condition, six repetitions were performed. If the anodic phase is more effective than the

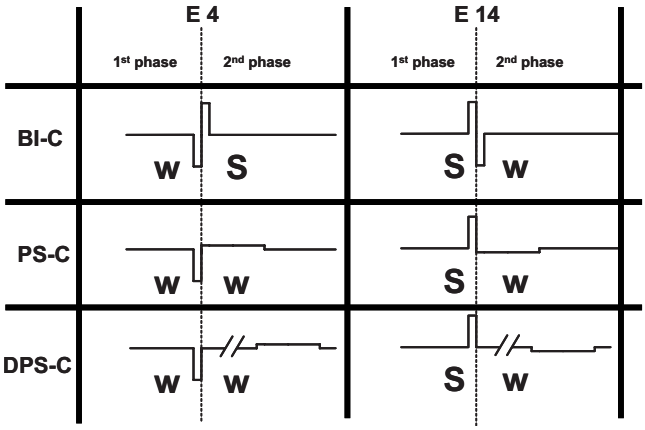


FIG. 5. Diagram showing the expected contributions of each phase of the BI-C, PS-C, and DPS-C maskers to neural excitation of fibers proximal to electrodes 4 (left panel) and 14 (right panel). The maskers are in bipolar configuration, and the leading polarity refers to the polarity relative to electrode 4. "S" and "w" correspond to expected strong and weak contributions, respectively. The predictions for BI-A, PS-A, and DPS-A are not shown as they are symmetrical to the ones for BI-C, PS-C, and DPS-C (by inverting electrodes 4 and 14).

cathodic phase, we expect the asymmetric maskers to predominantly excite neurons proximal to the electrode of the bipolar channel for which the short, high-amplitude phase is anodic. Note that, here, the maskers were not equated to give the same amount of masking at a single site (as in experiment 1) because we expected two remote masking peaks (proximal to each electrode) for the symmetric pulse shape (see Chatterjee *et al.*, 2006). Furthermore, we were not interested in the sharpness of the pattern per se but rather in the relative masking differences between the two electrode sites for different masker polarities.

In addition to the masking experiment, the pitches evoked by the six bipolar maskers were also compared, as we expect different places of excitation to elicit different pitch percepts. The six stimuli were pitch ranked using the optimally efficient "midpoint comparison procedure" that was originally developed to assist the fitting of auditory brainstem implants (Long *et al.*, 2005). The algorithm involves making pitch comparisons between pairs of stimuli, with the provisional pitch-ordering being updated as more comparisons are made. Each new stimulus is initially compared to the middle-ranking one in the provisional list, and the list is successively "bisected" as more comparisons are made. This procedure was repeated at least ten times and produced a mean rank and a standard error associated with each of the six stimuli.

2. Results

Figure 5 shows a diagram with the expected contributions of each phase of the maskers on the excitation of fibers near electrodes 4 (left part of the panel) and 14 (right part). The letters "S" and "w" correspond to expected strong and weak masking effects, respectively. The hypotheses are that a short, high-amplitude phase would produce a stronger excitation than a long, low-amplitude phase (Moon *et al.*, 1993; Macherey *et al.*, 2006) and that an anodic phase would produce a stronger excitation than a cathodic phase (Macherey *et al.*, 2008). For a BI-C stimulus, the first phase is anodic relative to electrode 14 and should predominantly excite fibers near electrode 14, while the second phase is anodic relative to electrode 4 and should excite fibers near electrode 4. We therefore expect equal excitation at both electrode sites. A similar effect is expected for BI-A. The situation should, however, be different for PS-C where the second phase has a long, low-amplitude phase and should have a weak masking effect. PS-C should therefore primarily excite fibers near electrode 14. Symmetrically, the excitation should be larger near electrode 4 for PS-A. A trend similar to that for PS is expected for the DPS pulse shape because the two phases are also asymmetric.

The results of the forward-masking experiment are illustrated in Fig. 6. Each panel corresponds to one of the three pulse shapes (left for BI, center for PS, and right for DPS) and shows, for each subject, the threshold shift obtained with an anodic-first masker (left part of the panels) and with a cathodic-first masker (right part). Solid lines connecting filled symbols represent the threshold shifts of a signal presented on channel 4, whereas dotted lines connecting open symbols are for a signal on channel 14. Thick lines show the mean data across subjects. For each pulse shape (BI, PS, and DPS), we performed a two-way repeated-measure ANOVA on the mean threshold shifts of the four subjects. The treatment factors were the signal channel and the leading polarity.

The masker polarity did not have any effect on the threshold shift of the signal for the BI and DPS pulse shapes, as suggested by a lack of significance of the polarity [$F(1,3)=0.59$, $p=0.5$ for BI and $F(1,3)=1.6$, $p=0.3$ for DPS] and interaction factors [$F(1,3)=2.38$, $p=0.2$ for BI and $F(1,3)=0.014$, $p=0.9$ for DPS]. However, the PS pulse shape showed consistent differences in threshold shifts that were dependent on the pulse polarity, as illustrated by the opposite trend of the solid and dotted thick lines in the middle panel of Fig. 6. This result was consistent with what was expected: For a signal presented on channel 4, the threshold shift was larger when the leading (short, high-amplitude) phase of the masker was anodic relative to channel 4 (i.e., using PS-A). Similarly, for a signal presented on channel 14, the threshold shift was larger when the leading phase of the masker was cathodic relative to channel 4 (i.e., using PS-C; except for S1 who showed no difference between the two polarities for this specific channel). This finding, which is supported by a significant signal channel X polarity interaction [$F(1,3)=11.9$, $p=0.041$], demonstrates that a PS stimulus presented to a wide bipolar channel can more effectively excite one region of the cochlea over another depending on its leading polarity. As expected from our previous study (Macherey *et al.*, 2008) and contrary to Frijns' model (1996), the region of the cochlea located near the electrode for which the short, high-amplitude phase is anodic was more effectively stimulated.

Figure 7(A) shows the results of the pitch ranking task. Three of the four subjects (S1, S2, and S4) ranked the PS-C stimulus as higher in pitch than the PS-A stimulus, consistent with a stronger activation of the basal part of the cochlea due to the PS-C pulse being anodic relative to electrode 14. S3 was, however, unable to rank the BI-A, BI-C, PS-A, and

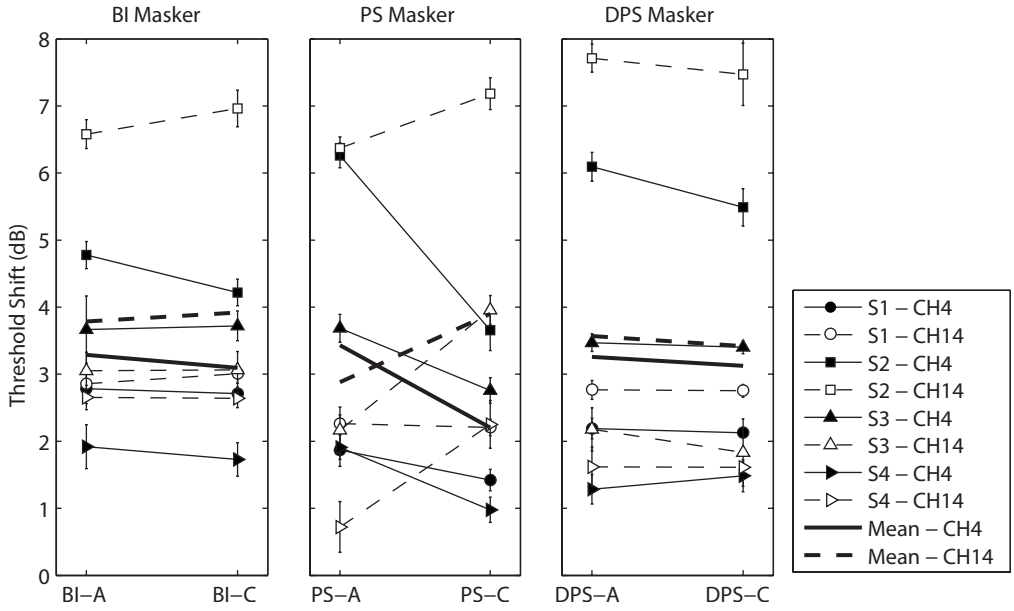


FIG. 6. Each panel illustrates the threshold shift of a monopolar BI-C signal for different bipolar masker shapes (left for BI, center for PS, and right for DPS). Four conditions were performed for each masker shape: The signal was either presented on channel 4 (filled symbols) or 14 (open symbols), and the masker was either anodic-first (right part of each panel) or cathodic-first (left part) with reference to the most apical electrode (4) of the pair.

PS-C stimuli. It is worth noting that the three subjects who showed a difference in pitch between PS-A and PS-C were not the same three subjects who showed a clear interaction in the masking experiment. The DPS-A and DPS-C stimuli evoked similar pitch percepts and, surprisingly, were in most cases ranked as lower in pitch than the other four stimuli. Finally, we performed paired-sample t-tests on the individual ranks obtained for BI-A and BI-C for the four subjects and found the pitch of BI-C to be significantly higher than that of BI-A for S1 and S2 ($p<0.02$) despite the fact that there was no masking difference between electrodes 4 and 14 for these two pulse shapes. It is therefore possible that the order of phases also affects pitch perception.

The mean pitch ranks were correlated with the differences in threshold shift between electrodes 14 and 4 for the six maskers. We expect the pitch to be higher when this threshold shift difference is large than when it is small or negative, as it would indicate a relatively stronger excitation of the basal region. Figure 7(B) shows the masking difference as a function of the pitch rank for the four subjects. Prior to performing the correlation, the masking data were standardized: For each subject, the mean threshold shift difference (averaged across maskers) was subtracted from each individual data. The within-subjects correlation turned out to be significant ($r=0.64$, $p=0.001$).

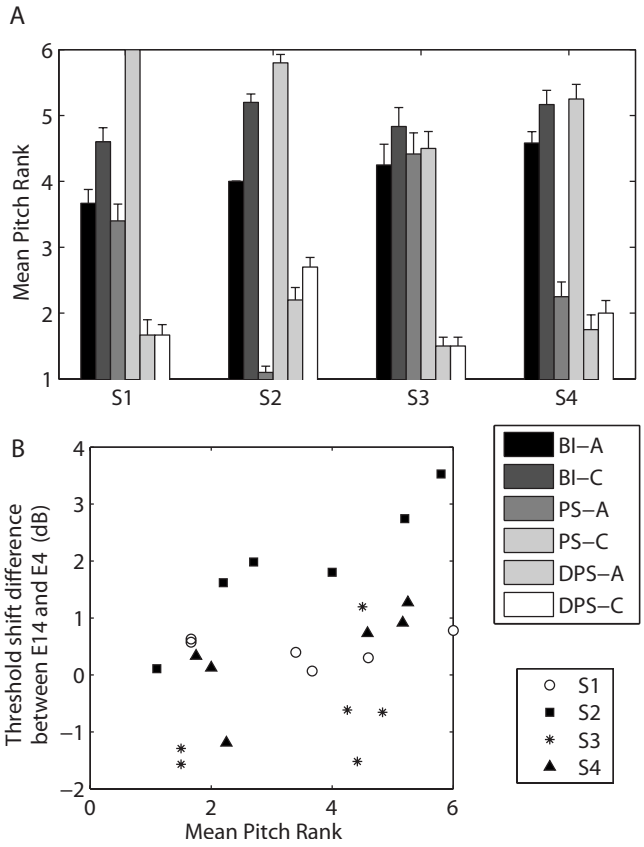


FIG. 7. (A) Mean pitch ranks and standard errors obtained by running the midpoint comparison with the six masker stimuli of experiment 2. (B) Threshold shift difference between electrodes 14 and 4 as a function of the mean pitch rank for the six maskers of experiment 2.

IV. DISCUSSION

A. Comparison of forward-masking functions: Methodological considerations

The improvement of future CI devices will probably require an increase in the number of independent stimulation channels. The implantation of a higher number of electrodes may achieve this goal only if each electrode is able to stimulate a spatially restricted portion of the cochlea. Therefore, it is necessary to develop techniques that selectively stimulate fibers proximal to an active electrode while producing as little excitation as possible at the other electrode sites.

In Sec. II, we have presented the results of a forward-

masking experiment that compared the spatial profiles produced by three different maskers. A complication in comparing the spatial spread of excitation for different stimuli (e.g., for different electrode configurations or different wave forms) comes from the fact that the masker levels are usually set to achieve a particular loudness percept (Chatterjee *et al.*, 2006; Kwon and van den Honert, 2006). This can lead to masking functions that can be offset vertically from each other by substantial amounts and which further need to be normalized in order to compare their sharpness. However, the normalization process may only be valid if the growth of masking is linear and also similar for all channels of the implant as it is equivalent to multiplying all masked thresholds of a given pattern by the same arbitrary factor. The work of Nelson and Donaldson (2001, 2002) provides an insight into the validity of this hypothesis. They measured the amount of masking produced on a masker channel as a function of masker level in single-pulse and pulse-train forward-masking experiments. They showed the amount of masking to increase linearly with masker level (expressed in μA). However, the slope of this increase was different across subjects and, more importantly, across electrodes measured within the same subjects. Extrapolating the results of their studies, the normalization of forward-masked excitation patterns relative to a given channel may well lead to an under- or an over-estimate of the degree of tonotopic selectivity in CI stimulation.

In the present study, we have adopted an alternative method to compare the masking patterns produced by different stimuli by equating the amount of masking on a given channel. This procedure has the advantage of avoiding normalization: If two functions are equal at the peak, then the one that is lower at the edge will always reflect better selectivity. Nevertheless, it also has its drawbacks. First, the masker equating part may be difficult to achieve. There is some variability in the measures that may prevent determination of the exact masker levels that mask the signal by the same amount. Second, the fact that the maskers are not compared at the same loudness may not be appropriate to identify the most spatially focused stimulus to be used in contemporary commercial implant systems. This is because existing clinical fitting procedures usually involve setting all individual channels to a particular loudness (i.e., MCL).

Our methodology may, however, have a potential clinical relevance for future implant designs employing high-density electrode arrays. In such a configuration, it is likely that the stimulation levels of each electrode will not be set to achieve a particular loudness percept but rather to stimulate a restricted group of fibers. Ultimately, one would want to be able to stimulate independently every surviving nerve fiber (or, more realistically, as many distinct groups of surviving nerve fibers as possible). Consequently, there is a need to identify the stimulus parameters that will optimally achieve this goal, and we think that the method described here constitutes a step forward. Furthermore, it should provide a more direct way to compare human data with animal or theoretical modeling data because it does not involve loudness judgments of the subject. In physiological and modeling experiments, spatial selectivity usually refers to the ability to stimulate a discrete group of neurons without stimulating neighboring ones. In practice, this is assessed by measuring the spatial extent of stimulation for a number of dB above the threshold of the lowest-threshold neuron, which is usually the one closest to the electrode (e.g., Snyder *et al.*, 2008). The approach in this paper is similar in that we determine how much "unwanted" stimulation there is for a fixed masking effect at the masker site. Another potentially important factor that we have taken into account is the confusion effect between masker and signal (Moore and Glasberg, 1982; Neff, 1985). We have proposed a way to overcome it by using different pulse rates for masker and signal.

B. Asymmetric pulses in monopolar configuration

The forward-masking patterns measured in experiment 1 showed some variability across subjects and pulse shapes. The BI-C and PS-C maskers produced very similar patterns in four of the subjects. Only S1 and S4 demonstrated sharper tuning for a PS-C compared to a BI-C masker. This sharpening was particularly remarkable for S4. The fact that this subject is the only one with a Helix electrode array may only be a coincidence as the main difference with the HiFocus II array (implanted in all other subjects) is the smaller distance between intracochlear contacts. It is unlikely that this would affect place-specificity since maskers and signals were monopolar stimuli in experiment 1.

Although there was no overall interaction between target electrode and masker type, some subjects did show a difference between the BI-C and DPS-C patterns. DPS-C clearly produced a broader excitation than BI-C in S2 and S3. Several studies have shown that less current is required to evoke the same loudness percept when the IPG is increased (McKay and Henshall, 2003; Carlyon *et al.*, 2005; van Wieringen *et al.*, 2005, 2006). Prado-Guitierrez *et al.* (2006) demonstrated that the effect of IPG on threshold was correlated with neural survival. One explanation that could account for their observation was that neural degeneration is usually accompanied by a partial demyelination of neurons. This tends to increase the time constants and could possibly reduce the effect of IPG on neural excitation. It is therefore possible that the across-channel differences between BI-C and DPS-C obtained in experiment 1 for some subjects relate to differences in neural survival. In regions of particularly high neural survival, we expect the IPG of the DPS-C masker to be more effective and therefore to produce relatively more masking than BI-C than in regions of poor neural survival.

The fact that we did not observe consistent results across subjects may be related to individual differences such as the position of the electrode relative to the neural elements and the specific degree of degeneration and of demyelination of the nerve fibers. All these factors are known to affect polarity sensitivity of neurons (Ranck, 1975; Rattay *et al.*, 2001). Also, we made the implicit assumption in experiment 1 that the the short, high-amplitude cathodic phase would be the most effective phase of PS-C and DPS-C although this may not necessarily be the case. The effects of phase duration on threshold and loudness have previously been studied using

symmetric biphasic pulses (Shannon, 1985; Moon *et al.*, 1993). Both of these studies showed that long-phase duration pulses are usually less effective (needing more charge) than shorter ones to reach threshold or to obtain the same loudness sensation, at least for phase durations similar to those used here (97 and 776 μs). However, the exact mechanisms of interactions between phase duration and polarity in pseudomonophasic pulses are, at present, unclear, and it is possible that the long, low-amplitude anodic phase was still a significant contributor to neural excitation, as suggested by recent masking data (Macherey *et al.*, 2008).

C. Asymmetric pulses in bipolar configuration

In experiment 2, we demonstrated that a pseudomonophasic stimulus with no IPG presented in a relatively wide bipolar configuration (bipolar+9) could more specifically excite the fibers proximal to either electrode of the bipolar pair, depending on its polarity. Consistent with our previous work on polarity sensitivity (Macherey *et al.*, 2008), we showed that fibers proximal to the electrode for which the short, high-amplitude phase was anodic were more effectively excited than fibers proximal to the other electrode of the bipolar channel. This reduced bimodality was, however, not observed with a DPS masker, which is identical to PS except that it has a long IPG. There may be two different explanations for this.

First, for our asymmetric maskers, the stimulus ended with the long, low-amplitude phase. Although we expect this phase to produce less neural excitation than the short, high-amplitude phase, it is, in the case of DPS-C, 4.4 ms closer to the signal. Hence, it is possible that the greatest effectiveness of the anodic, high-amplitude phase was counteracted by it being temporally further from the signal. However, this explanation fails to account for the finding that leading polarity affects pitch for PS but not for DPS pulses [Fig. 7(A)].

A second possible explanation relates to the current levels that were used. We showed that polarity sensitivity is prominent at comfortable listening level (e.g., Macherey *et al.*, 2006), i.e., that less current is needed for PS-A than for PS-C (in monopolar mode) to evoke the same loudness. However, these two pulse shapes do not show any difference at threshold. As DPS pulses need overall less current than PS pulses to reach the same loudness (see Table II), it is possible that the levels used for these stimuli are below the level at which polarity starts to matter. This would also be consistent with the fact that we observe a MCL difference between monopolar DPS-A and DPS-C for a 22-μs but not for a 97-μs phase duration in our previous publications (because short-phase durations require higher current levels than longer ones to evoke the same loudness; Macherey *et al.*, 2006, 2008).

It still remains to be demonstrated whether the polarity effect that we observed using this wide bipolar configuration (bipolar+9) holds for narrower configurations. Snyder *et al.* (2008) noted that bimodal excitation patterns using bipolar, symmetric biphasic pulses have been observed in CI users only when the distance between the two electrodes is large (see Chatterjee *et al.*, 2006). This probably arises from the fact that the two regions of activation overlap more and more as the distance between the electrodes is reduced. In a *narrow* bipolar configuration, we may still expect a PS pulse to sharpen one side of the spatial pattern compared to a BI pulse and focus the excitation near one electrode of the pair, as suggested by recordings in the inferior colliculus of guinea pigs (Bonham *et al.*, 2003). Clearly, some more work is needed to confirm those predictions. Moreover, it is likely that the place-specificity of asymmetric pulses will depend on the neural survival patterns of individual patients. For example, if one electrode of the bipolar pair is in a "dead" region and the other is in a "live" region, the polarity of the pulse will probably not affect the place of stimulation: Only the same live region will be stimulated for both PS-A and PS-C pulses. Concerning the implementation of such pseudomonophasic pulses in a realistic speech processing strategy (e.g., with a pulse rate higher than 800 pulses/s per channel), some compromises will have to be made regarding the rate of stimulation, the duration of the long, low-amplitude phase, the bipolar separation, and the nonsimultaneity of stimulation (van Wieringen *et al.*, 2008).

An additional potential advantage of the pseudomonophasic shape arises from the fact that the electrode array of a CI is usually not inserted all the way up to the apex. Consequently, some nerve fibers lying near the apex may not be excited by existing stimulation strategies. When symmetric pulses are presented in bipolar mode, the perceived pitch ("place" pitch) may correspond to the average of the two peaks of excitation corresponding to each electrode. The use of bipolar configuration combined with pseudomonophasic pulses may provide an access to more apical (low-frequency) regions of the cochlea by presenting the effective short, high-amplitude anodic phase of the pulse to the most apical electrode of the implant.

ACKNOWLEDGMENTS

We would like to thank the subjects for their enthusiasm, their patience, and the long time they have devoted to this research project. We also thank Leo Litvak for providing information on the electrode geometry, Dr. Frank Rattay for helpful discussions, and three anonymous reviewers for critical comments on a previous version of this manuscript. This research was supported by the Fonds voor Wetenschappelijk Onderzoek FWO-Vlaanderen (Grant No. FWO-G.0233.01) and the Research Council of the Katholieke Universiteit Leuven (Grant No. OT/03/58).

[1]The differences between BI-C and DPS-C on channel 9, although being small relative to the across-channel differences (0.48 and 0.35 dB for S2 and S3, respectively), were statistically significant and might have biased the results. This reveals a certain difficulty in optimally equating the maskers due to some variability in the results of step (iii) of the experimental design. For these two subjects, the statistical analyses were reiterated with the forward-masking data of DPS-C shifted to give the same mean excitation than the BI-C data on channel 9. The three factors of the two-way repeated-measure ANOVA, including masker type, channel, and interaction term, were still statistically significant. Furthermore, the paired-sample *t*-tests showed that the differences between BI-C and DPS-C on the three most apical channels of S2 and on the three most basal channels of S3 remain significant.

Abbas, P. J., Hughes, M. L., Brown, C. J., Miller, C. A., and South, H. (**2004**). "Channel interaction in cochlear implant users evaluated using the electrically evoked compound action potential," Audiol. Neuro-Otol. **9**, 203–213.

Badi, A. N., Hillman, T., Shelton, C., and Normann, R. A. (**2002**). "A technique for implantation of a 3-dimensional penetrating electrode array in the modiolar nerve of cats and humans," Arch. Otolaryngol. Head Neck Surg. **128**, 1019–1025.

Bierer, J. A. (**2007**). "Threshold and channel interaction in cochlear implant users: Evaluation of the tripolar electrode configuration," J. Acoust. Soc. Am. **121**, 1642–1653.

Bonham, B. H., Middlebrooks, J. C., Rebscher, S. J., and Snyder R. L. (**2003**), "The neurophysiological effects of simulated auditory prosthesis stimulation," Third Quarterly Progress Report No. N01-DC-02–1006, NIH. http://www.nidcd.nih.gov/staticresources/funding/programs/npp/pdf/N01–DC-2–1006QPR03.pdf.

Carlyon, R. P., van Wieringen, A., Deeks, J. M., Long, C. J., Lyzenga, J., and Wouters, J. (**2005**). "Effect of inter-phase gap on the sensitivity of cochlear implant users to electrical stimulation," Hear. Res. **205**, 210–224.

Chatterjee, M., Galvin, J. J. III, Fu, Q. J., and Shannon, R. V. (**2006**). "Effects of stimulation mode, level and location on forward-masked excitation patterns in cochlear implant patients," J. Assoc. Res. Otolaryngol. **7**, 15–25.

Chatterjee, M., and Shannon, R. V. (**1998**). "Forward masked excitation patterns in multielectrode electrical stimulation," J. Acoust. Soc. Am. **103**, 2565–2572.

Cohen, L. T., Busby, P. A., and Clark, G. M. (**1996**). "Cochlear implant place psychophysics. 2. Comparison of forward masking and pitch estimation data," Audiol. Neuro-Otol. **1**(5), 278–292.

Cohen, L. T., Richardson, L. M., Saunders, E., and Cowan, R. S. (**2003**). "Spatial spread of neural excitation in cochlear implant recipients: Comparison of improved ECAP method and psychophysical forward masking," Hear. Res. **179**, 72–87.

Cohen, L. T., Saunders, E., and Clark, G. M. (**2001**). "Psychophysics of a prototype peri-modiolar cochlear implant electrode array," Hear. Res. **155**, 63–81.

Cohen, L. T., Saunders, E., Knight, M. R., and Cowan, R. S. (**2006**). "Psychophysical measures in patients fitted with Contour and straight Nucleus electrode arrays," Hear. Res. **212**, 160–175.

Fishman, K. E., Shannon, R. V., and Slattery, W. H. (**1997**). "Speech recognition as a function of the number of electrodes used in the SPEAK cochlear implant speech processor," J. Speech Lang. Hear. Res. **40**, 1201–1215.

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.

Friesen, L. M., Shannon, R. V., and Cruz, R. J. (**2005**). "Effects of stimulation rate on speech recognition with cochlear implants," Audiol. Neuro-Otol. **10**, 169–184.

Frijns, J. H., de Snoo, S. L., and ten Kate, J. H. (**1996**). "Spatial selectivity in a rotationally symmetric model of the electrically stimulated cochlea," Hear. Res. **95**, 33–48.

Grill, W. M. Jr., and Mortimer, J. T. (**1996**). "The effect of stimulus pulse duration on selectivity of neural stimulation," IEEE Trans. Biomed. Eng. **43**, 161–166.

Hillman, T., Badi, A. N., Normann, R. A., Kertesz, T., and Shelton, C. (**2003**). "Cochlear nerve stimulation with a 3-dimensional penetrating electrode array," Otol. Neurotol. **24**, 764–768.

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.

Kwon, B. J., and van den Honert, C. (**2006**). "Effect of electrode configuration on psychophysical forward masking in cochlear implant listeners," J. Acoust. Soc. Am. **119**, 2994–3002.

Laneau, J., Boets, B., Moonen, M., van Wieringen, A., and Wouters, J. (**2005**). "A flexible auditory research platform using acoustic or electric stimuli for adults and young children," J. Neurosci. Methods **142**, 131–136.

Levitt, H. (**1971**). "Transformed up-down methods in psychoacoustics," J. Acoust. Soc. Am. **49**, 467–477.

Lim, H. H., Tong, Y. C., and Clark, G. M. (**1989**). "Forward masking patterns produced by intracochlear electrical stimulation of one and two electrode pairs in the human cochlea," J. Acoust. Soc. Am. **86**, 971–980.

Litvak, L. (**2003**). BEDCS Bionic Ear Data Collection System, Version 1.16, User Manual.

Long, C. J., Nimmo-Smith, I., Baguley, D. M., O'Driscoll, M., Ramsden, R., Otto, S. R., Axon, P. R., and Carlyon, R. P. (**2005**). "Optimizing the clinical fit of auditory brain stem implants," Ear Hear. **26**, 251–262.

Macherey, O., Carlyon, R. P., van Wieringen, A., and Wouters, J. (**2008**). "Higher sensitivity of human auditory nerve fibers to positive electrical currents," J. Assoc. Res. Otolaryngol. **9**, 241–251.

Macherey, O., van Wieringen, A., Carlyon, R. P., Deeks, J. M., and Wouters, J. (**2006**). "Asymmetric pulses in cochlear implants: Effects of pulse shape, polarity, and rate," J. Assoc. Res. Otolaryngol. **7**, 253–266.

McDermott, H. J. (**2004**). "Music perception with cochlear implants: A review," Trends Amplif. **8**, 49–82.

McIntyre, C. C., and Grill, W. M. (**1999**). "Excitation of central nervous system neurons by nonuniform electric fields," Biophys. J. **76**, 878–888.

McIntyre, C. C., and Grill, W. M. (**2000**). "Selective microstimulation of central nervous system neurons," Ann. Biomed. Eng. **28**, 219–233.

McIntyre, C. C., and Grill, W. M. (**2002**). "Extracellular stimulation of central neurons: Influence of stimulus waveform and frequency on neuronal output," J. Neurophysiol. **88**, 1592–1604.

McKay, C. M., and Henshall, K. R. (**2003**). "The perceptual effects of interphase gap duration in cochlear implant stimulation," Hear. Res. **181**, 94–99.

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., and Snyder, R. L. (**2007**). "Auditory prosthesis with a penetrating nerve array," J. Assoc. Res. Otolaryngol. **8**, 258–279.

Miller, C. A., Abbas, P. J., Hay-McCutcheon, M. J., Robinson, B. K., Nourski, K. V., and Jeng, F. C. (**2004**). "Intracochlear and extracochlear ECAPs suggest antidromic action potentials," Hear. Res. **198**, 75–86.

Miller, C. A., Abbas, P. J., Robinson, B. K., Rubinstein, J. T., and Matsuoka, A. J. (**1999**). "Electrically evoked single-fiber action potentials from cat: Responses to monopolar, monophasic stimulation," Hear. Res. **130**, 197–218.

Miller, C. A., Robinson, B. K., Rubinstein, J. T., Abbas, P. J., and Runge-Samuelson, C. L. (**2001**). "Auditory nerve responses to monophasic and biphasic electric stimuli," Hear. Res. **151**, 79–94.

Moon, A. K., Zwolan, T. A., and Pfingst, B. E. (**1993**). "Effects of phase duration on detection of electrical stimulation of the human cochlea," Hear. Res. **67**, 166–178.

Moore, B. C., and Glasberg, B. R. (**1982**). "Contralateral and ipsilateral cueing in forward masking," J. Acoust. Soc. Am. **71**, 942–945.

Moore, B. C., Glasberg, B. R., and Roberts, B. (**1984**). "Refining the measurement of psychophysical tuning curves," J. Acoust. Soc. Am. **76**, 1057–1066.

Neff, D. L. (**1985**). "Stimulus parameters governing confusion effects in forward masking," J. Acoust. Soc. Am. **78**, 1966–1976.

Nelson, D. A., and Donaldson, G. S. (**2001**). "Psychophysical recovery from single-pulse forward masking in electric hearing," J. Acoust. Soc. Am. **109**, 2921–2933.

Nelson, D. A., and Donaldson, G. S. (**2002**). "Psychophysical recovery from pulse-train forward masking in electric hearing," J. Acoust. Soc. Am. **112**, 2932–2947.

Pfingst, 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.

Prado-Guitierrez, P., Fewster, L. M., Heasman, J. M., McKay, C. M., and Shepherd, R. K. (**2006**). "Effect of interphase gap and pulse duration on electrically evoked potentials is correlated with auditory nerve survival," Hear. Res. **215**, 47–55.

Ranck, J. B. (**1975**). "Which elements are excited in electrical stimulation of mammalian central nervous system: A review," Brain Res. **93**(3), 417–440.

Rattay, F., Leao, R. N., and Felix, H. (**2001**). "A model of the electrically excited human cochlear neuron. II. Influence of the three-dimensional cochlear structure on neural excitability," Hear. Res. **153**, 64–79.

Shannon, R. V. (**1985**). "Threshold and loudness functions for pulsatile stimulation of cochlear implants," Hear. Res. **18**, 135–143.

Shannon, R. V. (**1990**). "Forward masking in patients with cochlear implants," J. Acoust. Soc. Am. **88**, 741–744.

Shannon, R. V., Fu, Q. J., Galvin, J., and Friesen, L. (**2004**). "Speech perception with cochlear implants," in *Cochlear Implants: Auditory Prostheses and Electric Hearing*, edited by F. G. Zeng, A. N. Popper, and R. R.

Fay (Springer-Verlag, New York), pp. 334–376.
Shannon, R. V., Zeng, F. G., Kamath, V., Wygonski, J., and Ekelid, M. (**1995**). "Speech recognition with primarily temporal cues," Science **270**, 303–304.
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.
Shepherd, R. K., Linahan, N., Xu, J., Clark, G. M., and Araki, S. (**1999**). "Chronic electrical stimulation of the auditory nerve using non-charge balanced stimuli," Acta Oto-Laryngol. **119**, 674–684.
Snyder, R. L., Middlebrooks, J. C., and Bonham, B. H. (**2008**). "Cochlear implant electrode configuration effects on activation threshold and tonotopic selectivity," Hear. Res. **235**, 23–38.
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.
van Wieringen, A., Carlyon, R. P., Laneau, J., and Wouters, J. (**2005**). "Effects of waveform shape on human sensitivity to electrical stimulation of the inner ear," Hear. Res. **200**, 73–86.
van Wieringen, A., Carlyon, R. P., Macherey, O., and Wouters, J. (**2006**). "Effects of pulse rate on thresholds and loudness of biphasic and alternating monophasic pulse trains in electrical hearing," Hear. Res. **220**, 49–60.
van Wieringen, A., Macherey, O., Carlyon, R. P., Deeks, J. M., and Wouters, J. (**2008**). "Alternative pulse shapes in electrical hearing," Hear. Res. **242**, 154–163.

Temporal pitch percepts elicited by dual-channel stimulation of a cochlear implant

Olivier Macherey[a] and Robert P. Carlyon
MRC Cognition and Brain Sciences Unit, 15 Chaucer Road, Cambridge CB2 7EF, United Kingdom

(Received 27 February 2009; revised 7 July 2009; accepted 6 November 2009)

McKay and McDermott [J. Acoust. Soc. Am. **100**, 1081–1092 (1996)] found that when two different amplitude-modulated pulse trains are presented to two channels separated by <1.5 mm, some cochlear implant (CI) listeners perceive the aggregate temporal pattern. The present study attempted to extend this general finding and to test whether dual-electrode stimulation would increase the upper limit of temporal pitch perception in CIs. Six subjects were asked to rank 12 dual-channel stimuli differing in their rate [ranging from 92 to 516 pps (pulses per second) on each individual channel] and in their inter-channel delay (pulses on the two channels being either nearly simultaneous or delayed by half the period). The data showed that, for an electrode separation of 0.75 or 1.1 mm, (a) the perceived pitch was on average slightly higher for the long-delay than for the short-delay stimuli but never matched the pitch corresponding to the aggregate temporal pattern, (b) the upper limit of temporal pitch did not increase using long-delay stimuli, and (c) the pitch differences between short- and long-delay stimuli were largely insensitive to channel order and to electrode configuration. These results suggest that there may be more independence between CI channels than previously thought.
 [DOI: 10.1121/1.3269042]

PACS number(s): 43.64.Me, 43.66.Hg, 43.66.Ts [RYL] Pages: 339–349

I. INTRODUCTION

In cochlear implants (CIs), sound information is transmitted through amplitude-modulated high-rate electrical pulse trains delivered to multiple locations along the cochlea. Although existing CIs use carrier rates up to several thousands of pps (pulses per second) per channel, there is some evidence that most implanted subjects are insensitive to temporal fine structure at such high rates. For example, several reports showed little or no improvement in speech recognition for increases in pulse rate above 400–800 pps per channel (e.g., Friesen *et al.*, 2005). It has also been suggested that very high rates may increase non-simultaneous channel interactions and therefore impair the coding of temporal modulations (Middlebrooks, 2004).

The limitation in the extraction of temporal fine structure experienced by CI listeners may partly relate to their inability to discriminate between single-channel pulse trains of different rates when the standard rate is higher than approximately 300 pps or to perceive an increase in pitch above such a rate. The reason for this upper limit of temporal pitch is unclear given that subjects show a great amount of variability with some of them being able to discriminate between rates as high as 800 pps (Townshend *et al.*, 1987; Kong *et al.*, 2009). Several possible reasons for this limitation have been put forward, including the absence of a place-rate match in the cochlea (e.g., Oxenham *et al.*, 2004), the fact that CIs do not stimulate very apical neurons (Middlebrooks and Snyder, 2009), or the possibility that more central processes have lost their ability to follow high rates due to prolonged deafness. Moreover, normal-hearing listeners show a higher upper limit of temporal pitch than the majority of CI listeners, being able to discriminate between unresolved harmonic complexes filtered in a fixed frequency region for rates as high as at least 700 pps (Carlyon and Deeks, 2002). Previous studies on the perception of temporal pitch by CI users have usually involved single-channel stimulation. In this study, we will investigate the temporal pitch percepts elicited by unmodulated pulse trains presented to two neighboring channels of a CI and we will study how these percepts are altered by changing the inter-channel delay.

The sensitivity of CI users to inter-channel delay has been mainly investigated in discrimination studies (Tong and Clark, 1986; McKay and McDermott, 1996, 1999; Carlyon *et al.*, 2000). Tong and Clark (1986) presented 100-pps pulse trains to two bipolar channels and measured the just-discriminable inter-channel delay. Their reference stimulus had a 5-ms inter-channel delay (half-a-period) and that of the signal was varied in steps of ±0.1 ms. Delay difference limens (DLs) were measured for several distances of the two bipolar channels, including a reference condition where both pulse trains were presented to the same channel. They obtained the smallest DLs (less than 0.5 ms) in the "same channel" condition and found the DLs to increase as the distance between the two bipolar pairs increased.

Similarly, McKay and McDermott (1996) performed a discrimination experiment using 500-pps pulse trains, also presented to two bipolar channels. Each pulse train was amplitude modulated at 100 Hz and consisted of one "big" and four "smaller" pulses (with the amplitude of the small pulses reduced to 50% of the dynamic range). They found that subjects could detect the inter-channel delay when the two bipolar channels were separated by less than approximately 3–4 mm but not for larger separations.

[a]Author to whom correspondence should be addressed. Electronic mail: olivier.macherey@mrc-cbu.cam.ac.uk

The results of these two studies may reflect the amount of overlap between the neural populations excited by each channel. If the two stimulating channels are close to each other, some auditory nerve fibers will respond to both of them and will convey a temporal code highly dependent on the inter-channel delay. As the spatial distance between channels increases, each channel will excite a more discrete neural population and the amount of overlap will decrease. In another study, McKay and McDermott (1999) used dual-channel stimulation as a way to assess spatial spread of excitation as a function of pulse duration. They found that subjects could discriminate between stimuli with different inter-channel delays at larger channel separations for long (266 μs) than for shorter (less than 100 μs) phase durations, consistent with a broader spatial spread for the longer-phase stimuli.

In addition, there is some evidence that CI subjects are sensitive to across-channel timing differences even when the overlap between neural populations is presumably small or absent. Carlyon *et al.* (2000) showed that cochlear implantees could reliably discriminate between inter-channel delays of 0.1 and 2 ms when *remote* bipolar channels were stimulated. The fact that this ability persisted when the pulse polarity was inverted or when an additional "masker" pulse train was presented on a channel in between the two target channels argued against their results being due to overlapping neural populations. One possible interpretation was that a more central mechanism (e.g., co-incidence detectors) could detect fine timing differences between the nerve firings of distinct populations of auditory nerve fibers. It is also worth noting that normal-hearing listeners can detect asynchronies between pairs of concurrent pulse trains that have been bandpass filtered into separate frequency regions, even when potential within-channel interactions are masked by noise (Carlyon, 1992; Carlyon and Shamma, 2003).

Recently, much attention has been given to the effects of simultaneous or quasi-simultaneous dual-channel stimulation on place-pitch perception (e.g., Donaldson *et al.*, 2005). However, little is known about the temporal pitch information that is conveyed when more than one channel is stimulated. To our knowledge, this has only been studied by McKay and McDermott (1996). They compared the pitches evoked by several dual-channel stimuli (pulse trains modulated at 100 Hz) in a pitch scaling experiment. For channel separations smaller than 1.5 mm, pitch increased with increases in inter-channel delay for three of the five subjects tested. The magnitude of this increase suggested that the aggregate temporal pattern rather than the individual channel patterns could be conveyed to higher levels of the auditory system. McKay and McDermott (1996) further proposed two alternative interpretations for this observation. First, subjects might have perceived the aggregate pattern because some auditory nerve fibers were responding to both stimulating channels and were therefore conveying a different temporal pattern when the inter-channel delay changed. Second, it was possible that each channel excited a different set of auditory nerve fibers, but that a more central neural mechanism was able to combine information arising from these two intracochlear sites.

The present study aims to disentangle these two hypotheses and to quantify the pitch variations as a function of inter-channel delay in dual-channel stimulation. If the second hypothesis of McKay and McDermott (1996) is true, then, we may expect the upper limit of temporal pitch to *increase* when two channels are stimulated with a delay of half-a-period compared to when only one channel is stimulated. For example, if neural populations proximal to two CI electrodes are able to follow a rate of 300 pps but not higher due to a limitation at the auditory nerve level, more central neurons may receive inputs from these two populations and convey a temporal code of 600 pps. The existence of such a "summing" central process is also suggested by physiological studies of acoustic stimulation. When primary auditory neurons are excited by a single tone with a frequency less than approximately 4–6 kHz, their responses are phase-locked to the stimulus. However, a given neuron usually does not fire on every cycle of the stimulus, especially as the frequency is increased beyond 200 Hz (Rose *et al.*, 1967). The encoding of the frequency of the waveform therefore requires a combination of the firings of several neurons (volley principle).

In the present study, we will compare the pitches elicited by unmodulated dual-channel stimuli differing in their inter-channel delay (quasi-simultaneous or delayed by half-a-period) for pulse rates ranging from 92 to 516 pps per channel. In Sec. II, we will show that the perceived pitch of a half-period delay dual-channel stimulus is only slightly higher than that of a quasi-simultaneous stimulus and that the upper limit of temporal pitch does not increase using half-period delay stimuli. In Sec. III, we will demonstrate that these results are robust to modifications of the electrode configuration (monopolar or bipolar) and of the channel order of stimulation (apical or basal-first). Finally, we will focus on the discriminability of dual-channel stimuli with different inter-channel delays in Sec. IV.

II. EXPERIMENT 1: PITCH RANKING OF DUAL-CHANNEL STIMULI

A. Rationale and methods

1. Subjects and stimuli

Four Advanced Bionics CII/HiRes90k (S1–S4) and two Nucleus CI24 (S5 and S6) subjects participated in experiment 1. Their biographical and clinical details are shown in Table I. The stimuli consisted of 400-ms duration, unmodulated pulse trains presented to two neighboring bipolar channels of the implant. The pulses were anodic-first, symmetric biphasic with a short inter-phase gap (0 or 8 μs). Polarity refers to the most apical electrode of the bipolar pair. The 12 stimuli differed in their rate (92, 129, 184, 258, 368, or 516 pps per channel) and in their inter-channel delay as follows. For each rate, there was a short- and long-delay condition [see Fig. 1(a)]. The short-delay was set to 388 μs and was constant across the different rates. It allowed the pulses on both channels to be nearly synchronous while limiting interaction effects due to residual polarization (de Balthasar *et al.*, 2003). The long-delay was set so that the second channel

TABLE I. Biographical and clinical details of the subjects who took part in the study. The last three columns show the stimulus parameters used in the bipolar dual-channel conditions.

Subject	Age (yr)	Etiology	Duration of deafness (yr)	Duration of implant use	Implant type	Phase duration (μs)	Basal channel	Apical channel
S1	64	Progressive	>12	6 months	HR90K	97	(8,10)	(7,9)
S2	73	Unknown sudden	<2	6 yr	CII	97	(8,10)	(7,9)
S3	61	Unknown progressive	34	5 yr	HR90K	97	(8,10)	(7,9)
S4	54	Meningitis	1	6 yr	CII	97	(8,10)	(7,9)
S5	60	Sudden viral	>30	3 yr	CI24R (CA)	150	(12,9)	(13,10)
S6	63	CSOM	10	10 yr	CI24M	150	(12,10)	(13,11)

would deliver a pulse delayed by half-a-period relative to the first channel and its value consequently co-varied with stimulation rate.

As we aimed to extend the results of McKay and McDermott (1996), we used similar parameters as in their study, i.e., a narrow bipolar separation ("BP+1" for S1–S4, and S6, and "BP+2" for S5), a phase duration ranging from 97 μs (for S1–S4) to 150 μs (for S5 and S6) and the smallest distance possible between the two channels (one electrode separation). BP+*X* refers to the separation between the active and return electrodes of the bipolar pair (e.g., BP+1 means that the two electrodes of the pair are separated by two electrode distances). The distance between two neighboring electrodes is 1.1 mm for the Advanced Bionics and 0.75 mm for the Nucleus. The electrodes tested for each subject are indicated in Table I.

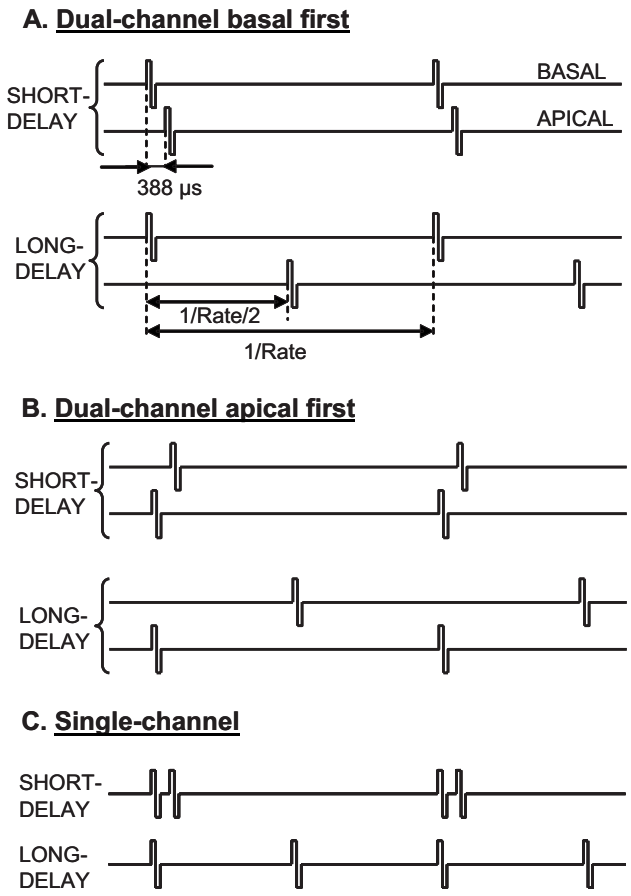


FIG. 1. Schematic representation of the stimuli used in the study: (A) dual-channel stimuli with a basal-leading channel for the short-delay (388 μs) and long-delay (half-a-period) conditions; (B) same as (A) with an apical-leading channel (C) same as (A) and (B) for single-channel stimulation. The two pulse trains are merged on the same channel.

In this experiment, the leading channel was the most basal electrode pair. This was to avoid any temporal effect being confounded by a place of excitation cue. In the short-delay condition, we may indeed expect the first pulse (leading channel) to partially mask the effect of the second pulse (delayed channel). Consequently, if the first pulse is presented on the basal channel, we may expect the short-delay stimulus to have a more basal place of excitation, and therefore to elicit a higher pitch sensation than the long-delay stimulus if the pitch difference was only due to a place cue. We anticipate that, if the results show that the long-delay stimulus sounds higher in pitch than the short-delay one, then this will strongly suggest that the pitch difference reflects a change in the temporal pattern of stimulation. Note that the effect of channel order will be specifically addressed in Sec. III.

All data were collected using the APEX experimental software platform (Laneau *et al.*, 2005), which acted as an interface to the BEDCS and NIC2 software routines provided by Advanced Bionics (Sylmar, CA) and Cochlear Corporation (Sydney, Australia), respectively. Impedances on all electrodes were checked prior to performing the experiments in order to ensure that the implants were driven below compliance.

2. Loudness balancing

Prior to performing the pitch comparisons, the dual-channel stimuli were equated in loudness using an adjustment paradigm. Each loudness-balancing trial consisted of two sounds presented consecutively with a 500-ms gap between them. The first sound was the reference and was fixed across the adjustment; the second sound was the one to match and subjects did so by pressing one of six virtual buttons labeled as "−," "− −," and "− − −" to make the sound softer and "+," "++," and "+++" to make it louder. The different buttons corresponded to different current steps: 8, 16, and 24 μA for Advanced Bionics subjects, and 1, 2, and 3 current units (approximately 0.176 dB per current unit) for CI24 subjects. Each time a button was pressed, the two stimuli were presented again with the new level of the target and this was repeated until the subject perceived them as equally loud.

For each pair of sounds that had to be loudness-balanced, two adjustments were performed. In the first one, the reference was fixed to a specific level and the target was adjusted to match its loudness. In the second one, the initial target became the reference and its level was set at the balanced level obtained in the first adjustment. The initial reference stimulus became the target and was the one to adjust. This was done to counterbalance any potential response bias due to the order of presentation. Once these two adjustments were performed, the balanced level of the initial target was obtained by averaging the level differences (in decibels) obtained in the two adjustments and subtracting it from the level of the initial reference sound.

The loudness-balancing of the 12 dual-channel stimuli was done in four steps.

(i) The most comfortable level (MCL) of the 92-pps pulse train presented on the basal channel was obtained by first increasing its level until the subject found it slightly too loud and then decreasing it to reach MCL. The 129-pps pulse train on the basal channel was further loudness-balanced to the 92-pps stimulus at MCL. Similarly, the 258-pps stimulus was balanced to the 129 pps and the 516 pps to the 258 pps. The levels of the 184-pps and 368-pps stimuli were obtained by logarithmic interpolation of the balanced levels found for the other rates.

(ii) Each rate-stimulus on the apical channel was then loudness-balanced to the same rate-stimulus presented on the basal channel at MCL [as obtained in (i)].

(iii) The dual-channel, short-delay stimuli were then constructed and the MCL for the 92-pps stimulus was determined by increasing the levels on both channels at the same time, keeping their level difference in decibels [as found in (ii)] constant. Let d be the difference in decibels between the MCLs of the 92-pps stimulus when presented alone on the basal channel and when presented in the dual-channel, short-delay condition. The higher-rate, short-delay stimulus levels were either inferred from the 92-pps dual-channel stimulus by reducing, for each rate, the levels obtained in the single-channel case by d or, if time allowed, were formally loudness-balanced to the 92-pps short-delay stimulus. In the latter case, the results were consistent with d being constant at all rates. This is also consistent with the loudness model of McKay *et al.* (2003), which first performs a temporal integration on each individual channel and then sums the neural activity obtained at different cochlear locations. The value of d averaged across subjects was 1.5 dB.

(iv) Each of the six long-delay stimuli was finally balanced to the corresponding short-delay stimulus at the same rate.

The loudness-balancing results of S1 at step (i) stage showed a large amount of variation. His results seemed to be particularly influenced by the starting level of the signal. The MCLs of all stimuli were therefore determined using a loudness chart and the subject was asked to estimate the loudness of the sounds. The same four steps were repeated as for the other subjects except that there was no loudness-balancing involved. It was noticed that this subject would rate an unusually wide range of levels as "most comfortable."

3. Pitch ranking

The aim of this experiment was to pitch-rank the 12 dual-channel stimuli previously described. This was done using the midpoint comparison procedure developed by Long *et al.* (2005) originally designed to optimize the fitting of auditory brainstem implants. This procedure consists of making pitch comparisons between pairs of sounds with the provisional pitch ordering being updated as more comparisons are made. To illustrate its functioning, assume a list of stimuli labeled A–Z that need to be ranked in pitch and that at one point of the procedure, the provisional order is [F B G C A E D]. The next stimulus to be ranked (H) will first be compared to the middle-ranked stimulus of the list (C). If it is higher, the list will be bisected and H will be compared to the middle-ranked stimulus (E) of the higher part of the provisional list ([A E D]). Subsequently, H will be compared to D or to A, depending on whether it was higher or lower than E. This will eventually lead to a new provisional ranking with more stimuli in the list.

The procedure was repeated at least 15 times with each subject and, for each of the 12 dual-channel stimuli, we calculated the mean and standard error of the rank. The order of presentation of the stimuli was randomized across the different repetitions.

B. Results

Figure 2 shows, for each subject, the mean rank as a function of pulse rate for both short-delay (filled squares) and long-delay (open circles) stimuli. Although the data of the two delay conditions are connected using different lines and symbols, all stimuli were mixed within the same block. The short-delay function shows that pitch increased as a function of pulse rate up to about 258 pps for S2–S4, after which it started to asymptote. The three other subjects (S1, S5, and S6) perceived the pitch to increase up to the highest rate tested (516 pps).

It was initially expected that at least at a low pulse rate [e.g., 92 pps, similar to the modulation rate used by McKay and McDermott (1996)], the pitch of the long-delay stimulus would be more similar to the pitch of a short-delay stimulus at twice the rate (184 pps), which would correspond to an exact doubling of the perceived pitch. At no instance did the results show this pattern. Overall, the pitch of the long-delay stimulus was only slightly higher than that of the short-delay stimulus when compared at the same rate. The data were analyzed in a two-way repeated-measures analysis of variance (ANOVA) with rate and delay as treatment factors. The analysis shows the effects of pulse rate [$F(1,5)=99.9$, $p<0.001$] and delay [$F(1,5)=37.1$, $p=0.002$] to be highly significant. However, there was no significant interaction between the two factors [$F(1,5)=0.63$, $p=0.46$]. The mean rank difference between short- and long-delay stimuli averaged across rate was relatively small (only 0.6 rank). As an

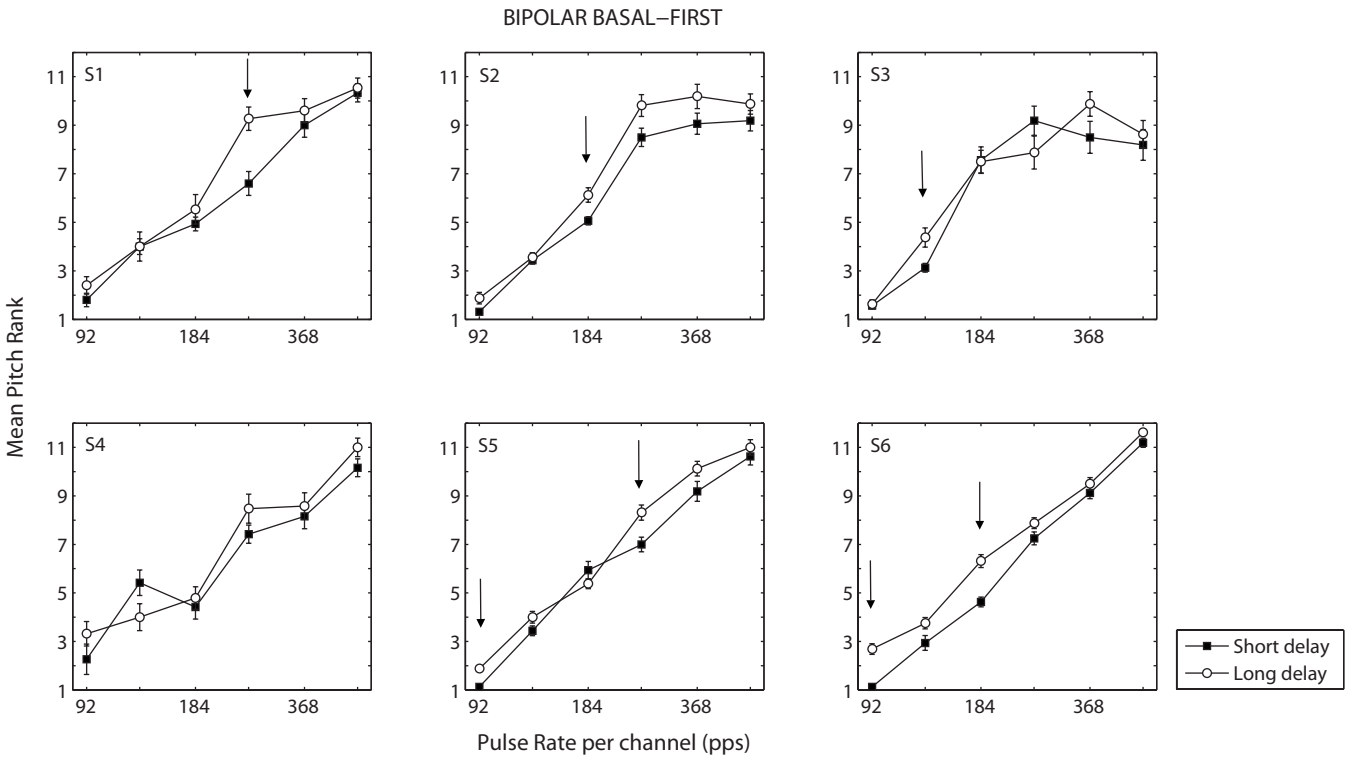


FIG. 2. Results of experiment 1 (bipolar, basal-leading dual-channel condition). Mean pitch-rank as a function of pulse rate per channel for the six subjects. Error bars indicate ± standard error.

example, the pitch of the long-delay stimulus at 92 pps was always lower than the pitch of the short-delay stimulus at 129 pps. Because consecutive rates are separated by approximately 40%, this means that the increase in pitch from short- to long-delay at 92 pps was always less than 40%.

The arrows in Fig. 2 point to the rates for which the rank difference between short- and long-delay stimuli was significant ($p<0.05$, Bonferroni-corrected). For the two highest rates tested (368 and 516 pps), the ranks were very similar for both delay conditions. Furthermore, the two functions (long- and short-delay) followed the same pattern. For example, S2–S4 showed a similar plateau as in the short-delay condition for rates higher than about 258 pps. These two observations suggest that dual-channel stimulation may not be useful to increase the upper limit of temporal pitch in CIs or that if it is, this combination does not depend on inter-channel delay. The third, and somewhat puzzling, observation was that the individual rates where subjects showed a pitch difference between the two delay conditions were not necessary the lowest rates. For example, S1 could not differentiate the pitches of the short- and long-delay stimuli at 92 pps but could do so at 258 pps. A similar "nonmonotonic" pattern was observed for S2, S3, and S5. Possible reasons for this trend will be further investigated in Sec. IV.

III. EXPERIMENT 2: EFFECTS OF ELECTRODE CONFIGURATION AND OF CHANNEL ORDER ON THE PITCH OF DUAL-CHANNEL STIMULI

A. Rationale and methods

The absence of any pitch doubling in the long-delay condition of experiment 1 contrasted with the results of some of the subjects tested by McKay and McDermott (1996). Experiment 2 provides several additional conditions aimed to test the generality of our results.

First, it was checked that the results were not due to conflicting place and temporal cues because of the specific channel order that was used (basal-channel first). The same procedure was therefore repeated with a leading *apical* channel [see Fig. 1(b)]. The levels on both channels were the same as in the basal-leading condition.

Second, we implicitly assumed in Sec. II that nearly simultaneous pulses would produce a temporal pattern of neural activity similar to truly simultaneous pulses. However, this needed to be confirmed. The same procedure was therefore repeated with both pulse trains presented on the same bipolar channel [see Fig. 1(c)]. In this case, we would expect the second pulse of the short-delay stimulus either to fall within the absolute refractory period of the nerve due to the first pulse or to evoke spikes shortly after the first pulse. The pitch evoked by the long-delay stimulus should therefore be more similar to that of the short-delay stimulus at twice the rate. The 12 single-channel stimuli were loudness-balanced in a similar way as in experiment 1. Five subjects performed this condition (all except S3). The electrodes tested were the same electrodes as those of the basal channel of experiment 1 ([8, 10] for S1, S2, and S4, [12, 9] for S5, and [12, 10] for S6).

Third, it is possible that the results of experiment 1 were due to the two bipolar channels producing very focused stimulation. The same procedure was therefore repeated in monopolar mode which (1) should theoretically produce a spatially broader excitation (e.g., Snyder *et al.*, 2008) and (2) is used in most contemporary CI speech-processing strate-

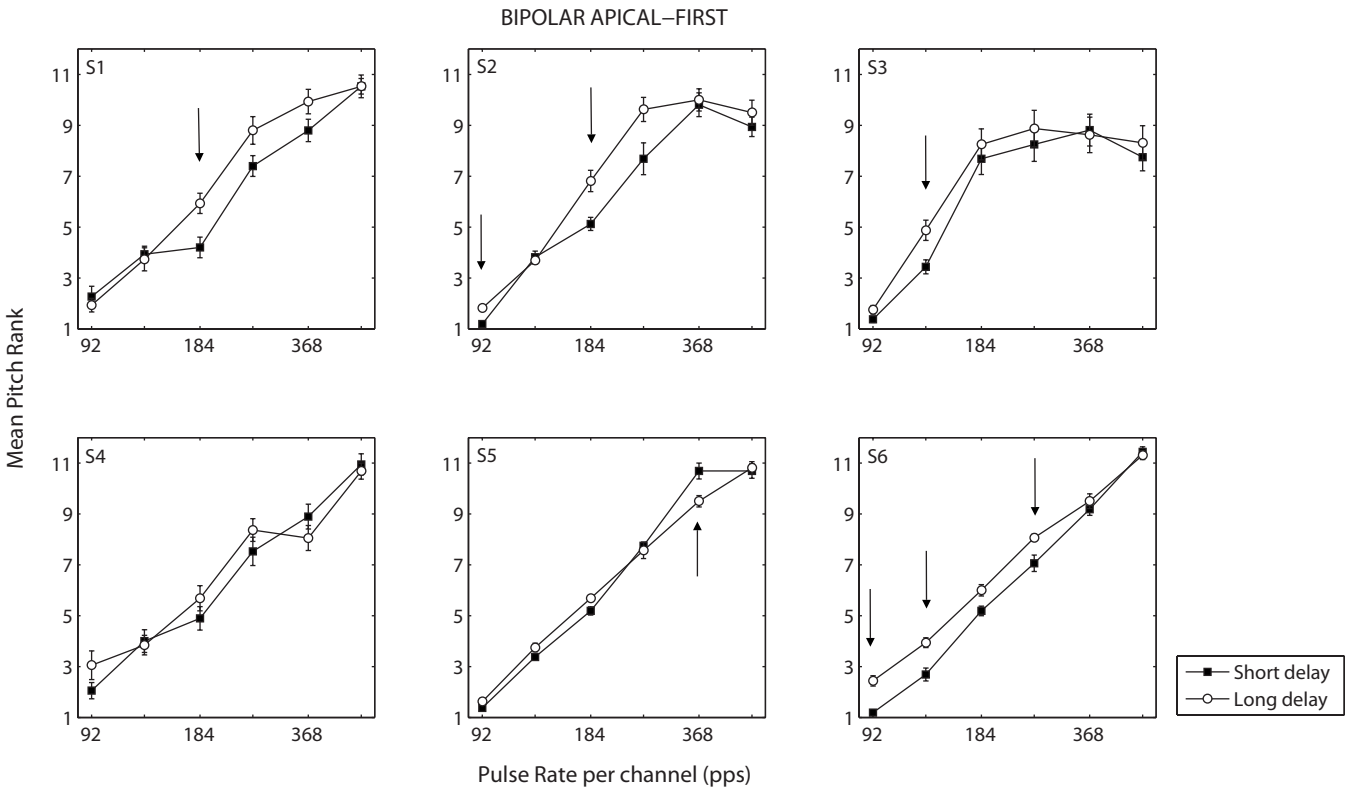


FIG. 3. Results of experiment 2 for the bipolar, apical-leading dual-channel condition. Mean pitch-rank as a function of pulse rate per channel for the six subjects. Error bars indicate ± standard error.

gies. Four subjects (S1, S2, S5, and S6) performed this condition with basal- and apical-leading channels [see Figs. 1(a) and 1(b)]. The levels were determined in the same way as described in Sec. II A. The active basal and apical electrodes were numbers 8 and 7 for the Advanced Bionics subjects (S1 and S2) and numbers 12 and 13 for the CI24 subjects (S5 and S6), respectively. The return electrode was the extracochlear case electrode for the Advanced Bionics subjects and the two extracochlear reference electrodes for the CI24 subjects (so-called "MP 1+2 mode").

B. Results

1. Apical-leading bipolar condition

Figure 3 shows the results of the apical-leading condition in bipolar mode. The results were very similar to those of experiment 1. A two-way repeated-measures ANOVA showed the effects of rate [$F(1,5)=89.5$, $p<0.001$] and delay [$F(1,5)=13.9$, $p=0.014$] to be significant, while the interaction factor was not [$F(1,5)=3.0$, $p=0.14$]. To check for any interaction between channel order and delay, an additional three-way repeated-measures ANOVA was performed on the mean data with pulse rate, delay, and channel order as treatment factors. Of course, no main effect of channel order was expected as the two conditions (apical- and basal-leading) were performed in separate blocks and therefore had the same mean pitch-rank. The analysis showed the effects of pulse rate [$F(1,5)=98.1$, $p<0.001$] and delay [$F(1,5)=28.9$, $p=0.003$] to remain highly significant. However, the interaction between delay and channel order [$F(1,5)=1.4$, $p=0.3$] was not significant nor were any of the other interaction factors. This lack of an interaction between delay and channel order combined with the similarity of the results in the basal- and apical-leading conditions strongly suggest that the pitch difference between short- and long-delay stimuli reflects a difference in the temporal patterns of neural discharges and is not strongly influenced by place of excitation cues.

2. Single-channel condition

Figure 4 shows the results of the single-channel condition. The long-delay stimulus is, here, equivalent to a regular, unmodulated pulse train [see Fig. 1(c), bottom] at twice the nominal rate. The data of the long-delay condition are replotted as a function of the "true" pulse rate using open triangles (long-delay "shifted" function). For three subjects (S1, S5, and S6), the triangles and filled squares perfectly overlap, consistent with a doubling in the perceived pitch of the long-delay stimulus compared to that of the short-delay stimulus. For S2, the trends of the short- and long-delay functions were similar (non-monotonic) although at 184 pps, the pitch of the long-delay stimulus was significantly higher than that of the short-delay condition (when compared at the same "true" rate). The results of S4 were very variable (showing large standard deviations) and the subject seemed to have been confused by the task. This may relate to the presence of several very high rate stimuli that were not discriminable. Another explanation could be that the short-delay stimuli had a different sound quality due to spikes being elicited by both of the pulses presented in each period and that this difference would prevent the subjects from comparing the stimuli based on the pitch. Overall, the large contrast between the single-channel (Fig. 4) and the dual-channel (Figs. 2 and 3) results

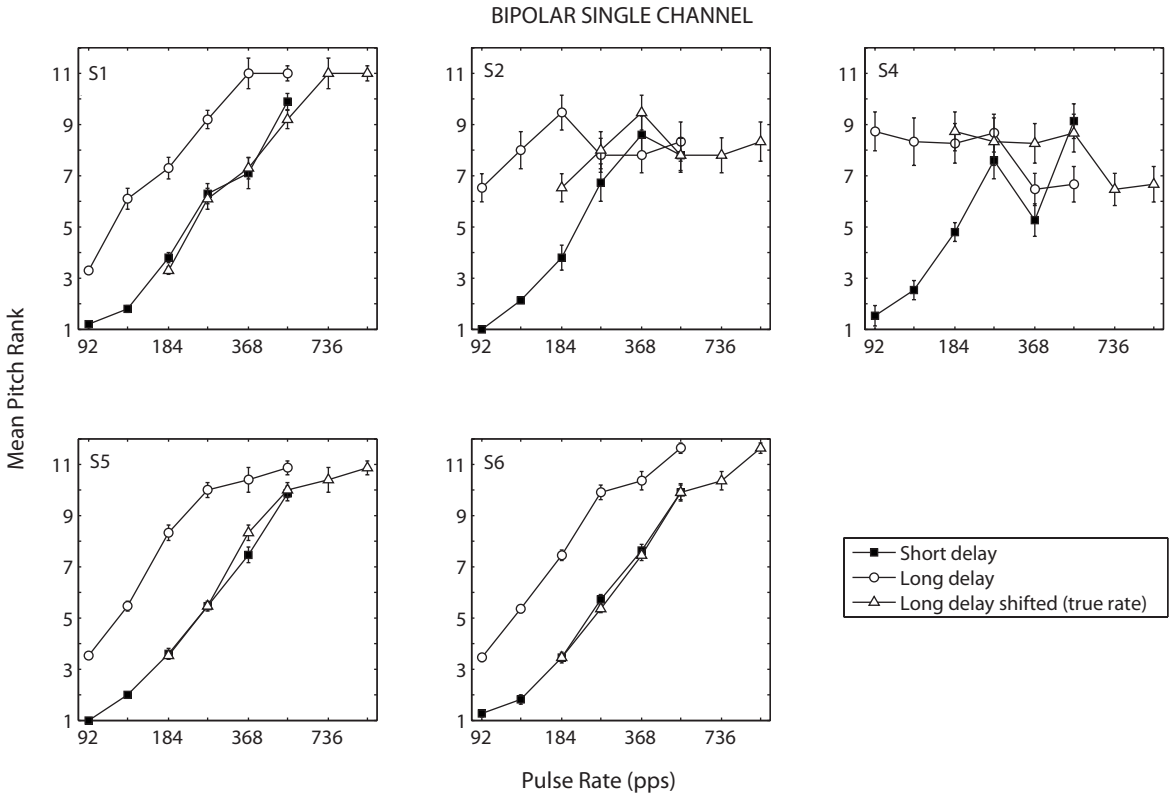


FIG. 4. Results of experiment 2 for the single-channel condition. Mean pitch-rank as a function of pulse rate for five subjects. Error bars indicate ± standard error. The open circles show the rank of the long-delay stimuli as a function of rate as defined in Fig. 1(a). Open triangles show the same data plotted as a function of the "true" rate on the single channel.

reinforces our finding that two neighboring channels separated by only 0.75 or 1.1 mm do not convey the aggregate temporal code to higher levels of the auditory system.

3. Basal- and apical-leading monopolar conditions

Figure 5 shows the results of the basal- and apical-leading conditions in monopolar mode. Here, again, the pitch differences between short- and long-delay stimuli were small. It is worth noting that the number of rates at which the pitches of the long- and short-delay stimuli differed appeared to be larger for the CI24 (S5 and S6) than for the Advanced Bionics subjects. This may relate to the channel separation that was used (one electrode), which is smaller for the CI24 implant (0.75 vs 1.1 mm.).

As in experiment 1, two-way repeated-measures ANOVAs were performed separately for the basal- and apical-leading conditions. While the pulse rate still had a highly significant contribution on the perceived pitch, the effect of delay failed to reach significance in both cases [$F(1,3)=3.36$, $p=0.16$ for basal-first; $F(1,3)=3.68$, $p=0.15$ for apical-first]. To check for any interaction effect between delay and electrode configurations, an additional four-way repeated-measures ANOVA was performed on the mean data of the four subjects (S1, S2, S5, and S6) who took part in this sub-experiment. The treatment factors were the delay, pulse rate, stimulation mode (bipolar or monopolar), and channel order (basal- or apical-first). Here, again, no main effect of stimulation mode or channel order was expected as the four different conditions were run in different blocks and therefore had the same mean rank. The effects of pulse rate [$F(1,3)=266.2$, $p=0.001$] and delay [$F(1,3)=13.2$, $p=0.036$] were significant but there was no interaction between any of the treatment factors. Only the interaction between delay and channel order approached significance [$F(1,3)=7.9$, $p=0.067$]. Furthermore, for a given subject, the rates showing a significant difference were often the same for the apical- and basal-first conditions. These data suggest that the pitch difference between short- and long-delay stimuli is largely independent of the stimulation mode and of the channel order of stimulation.

IV. EXPERIMENT 3: DISCRIMINATION OF INTER-CHANNEL DELAY

A. Rationale and methods

An intriguing result of experiments 1 and 2 was that the long-delay stimuli were, for some subjects, perceived as higher in pitch than the short-delay ones only at intermediate rates. In this experiment, we investigated the discriminability of short- and long-delay stimuli using two different methods.

First, the stimuli of experiment 1 (basal-leading, bipolar mode) were used in an odd-man out task. The main difference was that subjects could use any cues to perform the task, which consisted of a three-interval, two-alternative, forced-choice procedure. The first interval was fixed and always contained the short-delay stimulus. The second and third intervals were randomly assigned the short- or long-delay stimulus. Subjects were asked to indicate which of the second or third interval was different from the other two, i.e., which one contained the long-delay stimulus. Performance

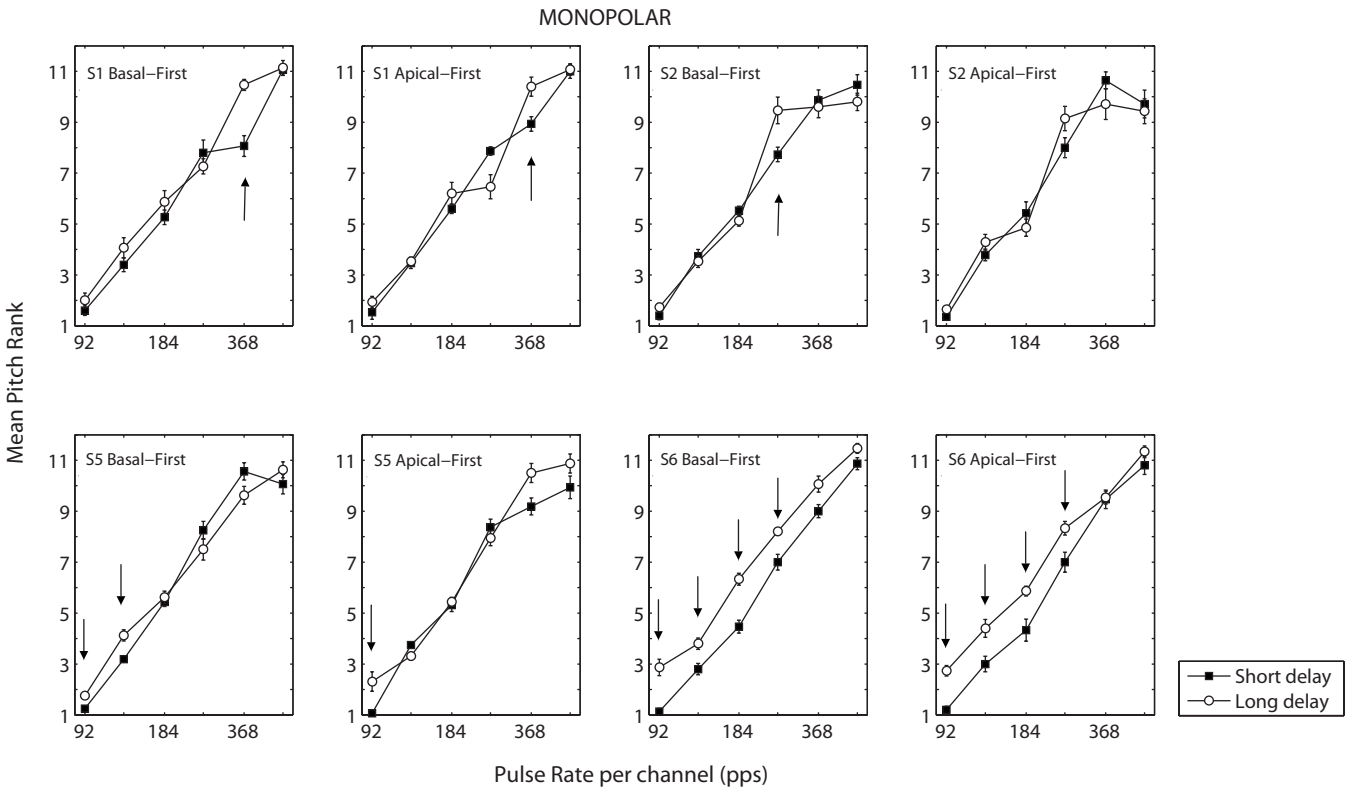


FIG. 5. Results of experiment 2 for the monopolar, basal- and apical-leading dual-channel conditions. Mean pitch-rank as a function of pulse rate for four subjects. Error bars indicate ± standard error.

was measured at the two extreme rates used in the pitch ranking experiments (92 and 516 pps per channel) and at a third intermediate rate corresponding to the highest rate for which there was a significant pitch difference in Sec. II. The value of this intermediate rate differed across subjects (129 pps for S3, 184 pps for S2 and S6, 258 pps for S1 and S5, and none for S4). The experiment consisted of blocks of 20 trials of the same rate condition, which were repeated between two and five times each. The different pulse rate conditions alternated from block to block.

Second, regarding the initial hypothesis that more central neurons are able to combine the temporal information conveyed by nearby auditory nerve fibers, a possibility would be that these central neurons can only do so at some specific rates. To investigate this, we measured inter-channel delay DLs at the highest rate for which there was a significant pitch difference in experiment 1 and compared it to inter-pulse interval DLs using a single bipolar channel at the same pulse rate. The hypothesis was that if the DLs are smaller in the dual-channel than in the single-channel case, then this would necessary imply either that a more central process is able to combine the inputs from two distinct populations of auditory nerve (AN) fibers in order to enhance the inter-channel delay sensitivity or, alternatively, that one channel conveys more accurate timing information than the other. However, if the DLs are smaller in the single-channel case, a sufficient explanation would be that a small subset of AN fibers is being stimulated by both channels. The task was a three-interval, two-alternative, forced-choice, three-down, one-up adaptive procedure with feedback (Levitt, 1971). The procedure stopped after eight reversals and each estimate was the average of the last six reversals. The standard was the short-delay stimulus and the delay of the signal was varied in steps of ± 194 μs. The signal had an initial delay of half-a-period (equivalent to the long-delay stimulus in Sec. II). Each adaptive procedure was repeated at least three times. Only four subjects participated (S1, S2, S5, and S6) because S3 decided to withdraw from the experiment and S4 did not show any pitch difference at any of the rates tested in Sec. II. The electrodes tested in the single-channel case corresponded to the electrodes of the basal channel of the dual-channel stimuli ([8, 10] for S1 and S2, [12, 9] for S5, and [12, 10] for S6). The current levels for the dual- and single-channel stimuli were the same as those used in the dual- and single-channel pitch experiments, respectively, and were kept constant during the procedure.

B. Results

1. Discriminability of short- and long-delay stimuli

The results of the odd-man out task are shown in Fig. 6(A) for rates of 92 pps (black), 516 pps (white), and for an intermediate rate whose value differed across subjects (gray). Discrimination performance was better at 92 pps than at 516 pps for all subjects. At 92 pps, four subjects (S2 and S4–S6) showed scores higher than 95%. Although this could have been expected for S5 and S6 who showed a significant pitch difference in experiment 1, it is worth noting that S2 and S4 did not consistently pitch-rank the two delay stimuli but could easily discriminate them. At 516 pps, most subjects could not discriminate between the two different delays, showing scores close to chance. In the intermediate rate con-

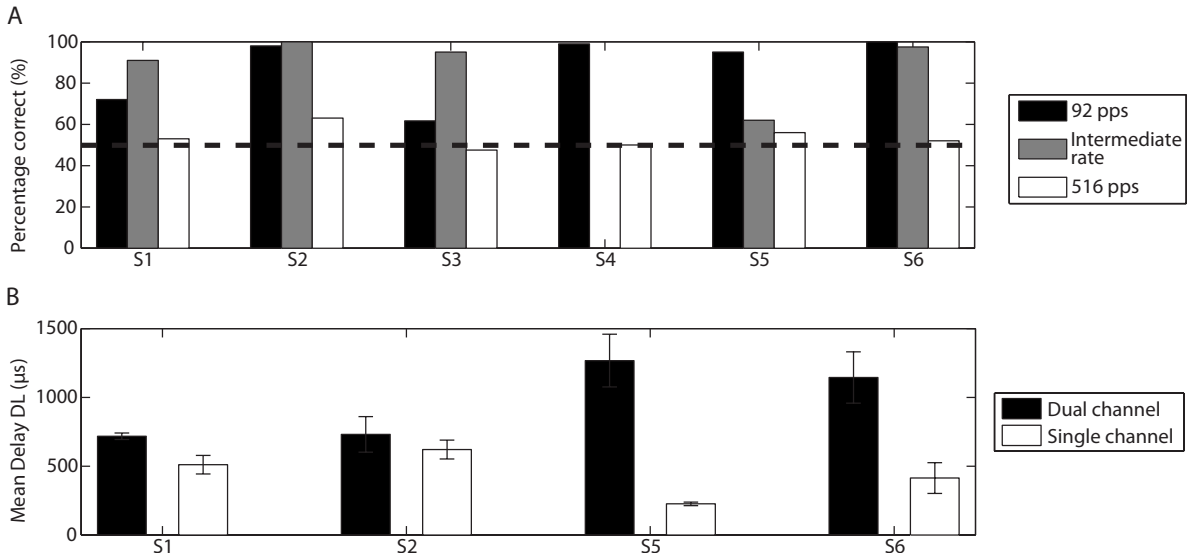


FIG. 6. Results of experiment 3: (A) percentage correct in the odd-man out task at three different rates for the six subjects. Only S4 did not perform the intermediate rate condition because he could not reliably pitch-rank short- and long-delay stimuli at any of the rates (see Fig. 2); (B) smallest detectable delay in μs at the intermediate rate for dual- and single-channel stimuli. Errorbars indicate $\pm$ standard error.

dition, all subjects except S5 showed high levels of discrimination. An interesting observation was that S1 and S3 were better at the intermediate rate than at 92 pps. This result is consistent with their ranking data, showing a larger pitch difference at the intermediate rate than at 92 pps.

2. Inter-channel difference limens

The results of the delay discrimination experiments are illustrated in Fig. 6(B). Black and white bars show the smallest detectable delay for the dual- and single-channel cases, respectively. For all subjects, this delay was either the same or smaller in the single-channel condition. This suggests that the sensitivity to long-delay stimuli in the dual-channel case can be reasonably explained by the same process that occurs in the single-channel case, i.e., that a subset of auditory nerve fibers is being stimulated by both channels.

V. DISCUSSION

A. Main findings

We have investigated the effect of inter-channel delay on temporal pitch perception for dual-channel stimuli with a channel separation of 0.75 or 1.1 mm. We showed the following.

(1) The perceived pitch was, on average, slightly but significantly higher for the long-delay than for the short-delay stimuli. Nevertheless, it never matched the pitch corresponding to the aggregate temporal pattern.
(2) This result was independent of the mode of stimulation (bipolar or monopolar) and of the channel order (basal- or apical-first).
(3) The upper limit of temporal pitch was not improved using long delays. In particular, subjects could not discriminate between short- and long-delay stimuli at the highest rate tested (516 pps).
(4) The largest pitch differences between short- and long-delay stimuli were often obtained at intermediate and not at the lowest rate.
(5) In an odd-man out task, most subjects could discriminate between short- and long-delay stimuli at 92 pps, but performed at chance at 516 pps.

The primary aim of this series of experiment was to evaluate the hypotheses of McKay and McDermott (1996) and to test whether subjects would perceive the aggregate temporal pattern (1) because of auditory nerve fibers responding to both channels or (2) because of a more central process combining inputs originating from two distinct populations of auditory nerve fibers. Our results are not consistent with either of these hypotheses as the perceived pitch never matched the pitch expected from the aggregate temporal pattern. The fact that the pitch was slightly higher for the long-delay than for the short-delay stimuli is consistent with most neurons conveying the single-channel temporal code and only a small subset responding to both channels and conveying the aggregate temporal code.

The fact that the pitch difference was sometimes larger at intermediate rates than at the lowest rate tested is intriguing. This may relate to the salience of the pitch percept being clearer at some rates than others. For example, if at a given rate, the aggregate pattern was conveying a more salient pitch cue than the individual channel patterns, this could make the subject "listen" preferentially to the neurons conveying the aggregate pattern. Incidentally, Kong *et al.* (2009) showed that rate discrimination by CI subjects was sometimes better at 200–300 pps than at 100 pps, suggesting that the transmission of temporal cues does not always degrade monotonically with increases in pulse rate.

B. Comparison to previous studies

Previous studies on dual-channel stimulation have been performed at a pulse rate or a modulation rate of 100 pps per

channel (Tong and Clark, 1986; McKay and McDermott, 1996; Carlyon *et al.*, 2000). For small channel separations, these authors observed high levels of discrimination performance. The results of our discrimination experiment at 92 pps are consistent with these data and show that most of our subjects could reliably perceive a difference between short- and long-delay stimuli.

In their pitch scaling study, McKay and McDermott (1996) found that three out of five subjects could perceive a pitch consistent with the aggregate temporal pattern, whereas this was never the case in our experiment. There are, however, several differences between their study and ours. First, they compared the pitches of amplitude-modulated pulse trains having different inter-channel delays with the pitches of *unmodulated* pulse trains (identical to our "long-delay" condition) having different rates. It is possible that these two groups of stimuli have different sound qualities and were therefore difficult to compare. In the present experiment, we only used unmodulated pulse trains and compared their pitches by varying both the inter-channel delay and rate. Second, as suggested by the facts that our patients were implanted more recently than those tested by McKay and McDermott (1996), and that criteria for implantation have become more relaxed over time, our subjects may have had better neural survival. If this is true, we would expect our subjects to show less overlap in the neural spreads of excitation produced by nearby channels. Third, McKay and McDermott (1996) used a pitch scaling procedure where they asked subjects to give a numerical estimate of each stimulus individually. Such a method may be prone to several kinds of non-sensory response biases that can affect the accuracy of the pitch match (Poulton, 1979). In contrast, we performed direct pitch comparisons between the stimuli using a ranking task. Finally, and perhaps most importantly, all subjects tested by McKay and McDermott (1996) were users of the Mini System 22 implant. This device has banded electrode contacts assumed to lie along the outer wall of the scala tympani (so-called "straight" electrode). In contrast, all our subjects except S6 have an electrode array designed to have a perimodiolar placement (HiFocus electrode for S1–S4 and Contour electrode for S5). Perimodiolar electrodes presumably lie closer to the excitable neural elements and have been shown to yield lower thresholds and narrower forward-masking profiles (Cohen *et al.*, 2006). Therefore, it is possible that our long-delay stimuli did not correspond to the aggregate temporal pattern because each individual channel produced a more spatially selective stimulation than in the McKay and McDermott (1996) study. Incidentally, S6, who has a straight (outer-wall placement) electrode was the subject for whom the pitch difference between short- and long-delay stimuli was the largest at several rates.

C. Implications for cochlear implants

One of the most commonly cited reasons for the limitations experienced by CI subjects in a range of tasks is the spread of current between neighboring electrode channels. Large current spreads have been reported in physiological studies for both monopolar and bipolar configurations (Snyder *et al.*, 2008). In human CI experiments, performance on speech recognition tasks usually does not improve when the number of channels is increased above approximately 8 (Friesen *et al.*, 2001), suggesting that the number of functional channels is smaller than the number of implanted intracochlear electrodes. The fact that the perceived pitch of our dual-channel stimuli did not correspond to the pitch of the aggregate temporal pattern suggests that the temporal codes conveyed by neighboring channels are largely independent. In other terms, although the neural populations excited by neighboring channels may be overlapping, it is possible that the neurons conveying the meaningful temporal codes of each channel are located in spatially restricted regions that do not overlap much between channels. The fact that several subjects were able to discriminate between long- and short-delay stimuli but were not able to pitch-rank them would support such a hypothesis. Another alternative explanation could be that the response to one or to both of the channels spreads away, leading to "off-place listening" (see Dingemanse *et al.*, 2006).

ACKNOWLEDGMENTS

We thank Colette McKay for her helpful discussions. This study was supported by the Wellcome Trust (Grant No. 080216). This research was approved by the Cambridge Local Research Ethics Committee.

Carlyon, R. P. (**1992**). "Detecting F0 differences and pitch-pulse asynchronies," in *The Auditory Processing of Speech. From Sounds to Words*, edited by B. Schouten (Mouton-De Gruyter, Berlin), pp. 149–156.

Carlyon, R. P., and Deeks, J. M. (**2002**). "Limitations on rate discrimination," J. Acoust. Soc. Am. **112**, 1009–1025.

Carlyon, R. P., Geurts, L., and Wouters, J. (**2000**). "Detection of small across-channel timing differences by cochlear implantees," Hear. Res. **141**, 140–154.

Carlyon, R. P., and Shamma, S. (**2003**). "An account of monaural phase sensitivity," J. Acoust. Soc. Am. **114**, 333–348.

Cohen, L. T., Saunders, E., Knight, M. R., and Cowan, R. S. (**2006**). "Psychophysical measures in patients fitted with contour and straight nucleus electrode arrays," Hear. Res. **212**, 160–175.

de Balthasar, C., Boëx, C., Cosendai, G., Valentini, G., Sigrist, A., and Pelizzone, M. (**2003**). "Channel interactions with high-rate biphasic electrical stimulation in cochlear implant subjects," Hear. Res. **182**, 77–87.

Dingemanse, J. G., Frijns, J. H., and Briaire, J. J. (**2006**). "Psychophysical assessment of spatial spread of excitation in electrical hearing with single and dual electrode contact maskers," Ear Hear. **27**, 645–657.

Donaldson, G. S., Kreft, H. A., and Litvak, L. (**2005**). "Place-pitch discrimination of single- versus dual-electrode stimuli by cochlear implant users," J. Acoust. Soc. Am. **118**, 623–626.

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.

Friesen, L. M., Shannon, R. V., and Cruz, R. J. (**2005**). "Effects of stimulation rate on speech recognition with cochlear implants," Audiol. Neuro-Otol. **10**, 169–184.

Kong, Y. Y., Deeks, J. M., Axon, P. R., and Carlyon, R. P. (**2009**). "Limits of temporal pitch in cochlear implants," J. Acoust. Soc. Am. **125**, 1649–1657.

Laneau, J., Boets, B., Moonen, M., van Wieringen, A., and Wouters, J. (**2005**). "A flexible auditory research platform using acoustic or electric stimuli for adults and young children," J. Neurosci. Methods **142**, 131–136.

Levitt, H. (**1971**). "Transformed up-down methods in psychoacoustics," J. Acoust. Soc. Am. **49**, 467–477.

Long, C. J., Nimmo-Smith, I., Baguley, D. M., O'Driscoll, M., Ramsden, R., Otto, S. R., Axon, P. R., and Carlyon, R. P. (**2005**). "Optimizing the clinical fit of auditory brain stem implants," Ear Hear. **26**, 251–262.

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., and McDermott, H. J. (**1996**). "The perception of temporal patterns for electrical stimulation presented at one or two intracochlear sites," J. Acoust. Soc. Am. **100**, 1081–1092.
McKay, C. M., and McDermott, H. J. (**1999**). "The perceptual effects of current pulse duration in electrical stimulation of the auditory nerve," J. Acoust. Soc. Am. **106**, 998–1009.
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 Snyder, R. L. (**2009**). "Enhanced transmission of temporal fine structure using penetrating auditory nerve electrodes," in Proceedings of the 32nd Midwinter Research Meeting, Association for Research in Otolaryngology, Baltimore, MD.
Oxenham, A. J., Bernstein, J. G., and Penagos, H. (**2004**). "Correct tonotopic representation is necessary for complex pitch perception," Proc. Natl. Acad. Sci. U.S.A. **101**, 1421–1425.
Poulton, E. C. (**1979**). "Models for biases in judging sensory magnitude," Psychol. Bull. **86**, 777–803.
Rose, J. E., Brugge, J. F., Anderson, D. J., and Hind, J. E. (**1967**). "Phase-locked response to low-frequency tones in single auditory nerve fibers of the squirrel monkey," J. Neurophysiol. **30**, 769–93.
Snyder, R. L., Middlebrooks, J. C., and Bonham, B. H. (**2008**). "Cochlear implant electrode configuration effects on activation threshold and tonotopic selectivity," Hear. Res. **235**, 23–38.
Tong, Y. C., and Clark, G. M. (**1986**). "Loudness summation, masking, and temporal interaction for sensations produced by electric stimulation of two sites in the human cochlea," J. Acoust. Soc. Am. **79**, 1958–1966.
Townshend, B., Cotter, N., Van Compernolle, D., and White, R. L. (**1987**). "Pitch perception by cochlear implant subjects," J. Acoust. Soc. Am. **82**, 106–115.

Sensory constraints on auditory identification of the material and geometric properties of struck bars

Robert A. Lutfi and Christophe N. J. Stoelinga
Department of Communicative Disorders, Auditory Behavioral Research Laboratory, University of Wisconsin, Madison, Wisconsin 53706

(Received 28 May 2009; revised 8 October 2009; accepted 12 October 2009)

A computational formula is derived for estimating the constraints limited auditory sensitivity imposes on auditory identification of the material and geometric properties of struck bars. The formula combines a model of the transverse motion of the bar with empirical psychometric functions to map out "null" regions in the bar's physical parameter space where changes in the frequency, amplitude, and decay of partials are likely below threshold for detection. Parameters of the physical space include bar density, Young's modulus, fluid and viscoelastic damping factors, bar length, and bar cross-sectional area (as related to bar shape and hollowness). The formula is used to estimate the possible effect of limited sensitivity in past studies on the auditory identification of bar attributes. The results suggest that sensitivity may, indeed, have played a role in some studies, and that apparent discrepancies in results may be understood based on whether the predominant source of damping was internal or external to the bar. The formula identifies conditions representing an expected bound on identification performance and thereby may be used to aid in the design of future studies for which the struck bar is the sound source of choice.
 [DOI: 10.1121/1.3263606]

PACS number(s): 43.66.Ba, 43.66.Fe [BCM] Pages: 350–360

I. INTRODUCTION

Most of us at one time or another have tapped on an object to determine from the sound whether it is hollow or solid, or made of metal or wood. What information in the sound allows us to do this? It is a simple question, but one that is currently at the center of a debate concerning one of the most important challenges for contemporary psychoacoustics—understanding how we perceive the world through sound. Historically, the debate has involved competing theories as to whether the listeners have access to *invariant* acoustic information that would allow them to uniquely identify object attributes, such as material composition or hollowness (cf. Gibson, 1966; Ullman, 1980; Micheals and Carello, 1981; Neuhoff, 2004; and Lutfi, 2008). In the parlance of literature, invariant acoustic information for an object attribute changes only with that attribute; it is unchanged for variation in all others. Invariant acoustic information for material, for example, would not change with the size of the object, its shape, or the manner in which the object is suspended and struck. As such it would allow a listener to accurately identify material despite the many forms the sounding object or event may take.

Much effort, naturally, has been devoted to determining what might serve as invariant acoustic information for different source attributes. Conjecture, for the most part, has been based on physical-acoustic analyses of rudimentary, freely vibrating sources with few degrees of freedom (homogenous, isotropic bars, and plates). The basic attributes of bars and plates considered thus far include material composition (Wildes and Richards, 1988; Lutfi and Oh, 1997), hollowness (Lutfi, 2001), size (Gaver, 1988; Lutfi, 2008), shape (Kac, 1966; Lakatos *et al.*, 1997; Kunkler-Peck and Turvey, 2000), mallet hardness (Freed, 1990), and point of contact (Lutfi and Liu, 2007). The relevant acoustic relations identified in each case have been proposed as a likely basis for listener judgments regarding these attributes, but such speculation has so far received only spotty empirical support. In the identification of material and in the detection of hollowness, for example, listeners are often found to base their judgments on simple differences in the frequency or duration of sounds (Lutfi and Oh, 1997; Lutfi, 2001; Giordano and McAdams, 2006), or to be overly influenced by the highest-level partials in the sounds (Lutfi *et al.*, 2008). Individual differences in the cues listeners use to make decisions have been reported (Lutfi and Liu, 2007), and it is not uncommon for studies investigating very similar conditions to come to quite different conclusions regarding the role of different sources of acoustic information (cf. Lutfi and Oh, 1997; Klatzky *et al.*, 2000).

The apparent discrepancies in the outcomes of these studies are not well understood, but in attempts to explain them one consideration has been largely overlooked—it is whether the invariant acoustic information for identification is always above the listeners' threshold for detection. Human listeners, we know, are capable of detecting quite small changes in the frequency, amplitude, and decay of sounds (Jesteadt *et al.*, 1977; Wier *et al.*, 1977; Schlauch *et al.*, 2001), but sensitivity for the relations among these parameters, such as might constitute the invariant acoustic information for different source attributes, has not been widely studied (cf. Stoelinga and Lutfi, 2008). Lutfi (2001) considered the role of limited sensitivity in the auditory detection of the hollowness of struck bars. From an analysis of theoretical equations of bar motion, the invariant acoustic information was determined to be a specific relation between the

frequency and decay of partials. Sensitivity to this relation was gauged by adding small random perturbations to the frequency and decay of partials so as to simulate the effects of sensory (internal) noise. For the specific conditions of the study, the simulations suggested that the relation would not yield a significant detection advantage over simple differences in frequency alone. This outcome was corroborated by data from listeners showing trial-by-trial judgments of hollowness to be most strongly correlated with frequency. Further simulations by Lutfi (2000) suggested that there are likely other conditions involving judgments of material and size where invariant acoustic relations are similarly vulnerable to sensory noise. Giordano and McAdams (2006) considered the role of limited sensitivity in the judgment of material from the impact sounds of large suspended plates. A cochlear model was used to evaluate the audibility of changes in various descriptors of these sounds, including a specific relation between frequency and decay deemed to be invariant acoustic information for material (cf. Wildes and Richards, 1988). Listener judgments were found to be correlated with this relation, but only for judgments made across gross material categories (steel-glass vs wood-plexiglass). Judgments within material categories, where the acoustic differences were likely more subtle, were correlated more strongly with simple differences in frequency.

The acoustic information that is available about a source must be limited to some extent by the ear's finite ability to resolve changes in the frequency, amplitude, and decay of individual partials. What these limits are and how they might have influenced past results are not known, nor would it be practical for all cases of potential interest to attempt an empirical determination of these limits. On the other hand, human sensitivities to individual changes in frequency, amplitude, and decay are fairly well documented in studies. By considering these data within the context of a viable model of the resonant source, it should be possible to provide a reasonable estimate of the effects of sensitivity on the identification of source properties for a broad range of conditions of interest. In the present study, we take this approach to map out the regions of a large physical parameter space for which corresponding changes in frequency, amplitude, and decay are likely to be below a listener's threshold for detection. We refer to these as null regions (NRs) on the premise that a change not reliably detected is not likely to participate in a higher order relation constituting invariant acoustic information for the listener.

II. ANALYSIS

A. Model of the resonant source

As a first step in the analysis we require a viable model of the resonant source that can be used to identify the acoustic effects of manipulations in source parameters and their likely effect on sensitivity. We choose for this purpose a homogenous, isotropic bar, struck with a mallet. The struck bar has been the sound source of choice in many studies because its acoustic properties (at least prior to propagation) are fairly well understood and because it represents, in principle, a large class of sound-producing objects encountered in everyday listening.

The specific model of the bar described here is based on the work of Chaigne and Doutaut (1997). The Chaigne–Doutaut model is principally the same model used for sound synthesis in several of the studies cited in the Introduction, except that it provides a more refined account of viscoelastic losses in the bar and a more realistic representation of the bar-mallet interaction (for psychophysical evaluations, see Lutfi *et al.*, 2005; McAdams *et al.*, 2004). The model gives the transverse displacement w over time at each point x along the length bar in response to an applied force density $F(x,x_0,t)$ at point x_0. Only the transverse displacement is considered, as the longitudinal and torsional components of the motion typically radiate little sound. The displacement function is given by the fifth-order differential equation

$$\frac{\partial^2 w}{\partial t^2}(x,t) = \frac{1}{\rho S}\frac{\partial^2 M}{\partial x^2}(x,t) - \gamma_B \frac{\partial w}{\partial t}(x,t) + F(x,x_0,t), \tag{1a}$$

$$M(x,t) = -EI\left(1 + \eta\frac{\partial}{\partial t}\right)\frac{\partial^2 w}{\partial x^2}(x,t), \tag{1b}$$

where ρ is mass density of the bar, S is its cross-sectional area, M is the bending moment, γ_B is the fluid (air) damping coefficient, E is Young's modulus, I is the moment of inertia of the cross section, and η is the viscoelastic damping coefficient. The values of ρ, E, and η in this case are dictated by the particular material of the bar and are measured empirically. The force density term $F(x,x_0,t)$ is given by a model of the interaction between bar and mallet in which the impact force is governed by Hertz's law and the inertial force of the mallet by Newton's second law. Chaigne and Doutaut (1997) used the finite difference method to provide numerical time-domain solutions to the differential equations and showed these to be in excellent agreement with physical measurements on wood bars. As they noted, however, the model does not apply to most metal bars where other sources of damping, particularly those resulting from thermoelastic losses, play a complex role (cf. Cremer, 2005).

The Chaigne–Doutaut model is best suited for our purposes, but numerical solutions are computationally prohibitive for all points of interest in the physical space. Fortunately, we do not require complete solutions, only a few basic acoustic descriptors of these solutions. Expressions describing the dependence of these descriptors on the physical properties of the bar can be derived analytically by imposing a few minor constraints in the application of the model. To begin, we limit consideration to cases where the bar is lightly damped. The solution to Eq. (1) is then of the form

$$w = \sum_n A_n e^{-t/\tau_n} \sin(2\pi f_n t), \tag{2}$$

where f_n, A_n, and $\tau_n \gg 1/f_n$ denote, respectively, the frequencies, initial amplitudes, and decay moduli of individual partials making up the sound. Expressions for the frequencies in this case are derived in standard acoustic texts (cf. Morse and Ingard, 1968). For example, for the cylindrical bar clamped

at one end and struck at the other, the frequencies are

$$f_n = \frac{\pi r}{4L^2}\sqrt{\frac{E}{\rho}}\beta_n^2, \tag{3}$$

where $\beta_1 = 0.597$, $\beta_2 = 1.494$, and $\beta_{n>2} = n - \frac{1}{2}$ determine the frequency ratio of successive partials.

Consider next the amplitudes A_n. Their role in the identification of bar properties has been mostly neglected in literature because the changes in amplitude are generally small compared to those that result from differences in the manner of impact and other factors unrelated to bar properties. The changes are considered here, however, as we have found points in the physical parameter space where they may prove to be a useful cue. So as to constrain the behavior of the amplitudes the force density is approximated by a delta pulse applied to one end of the bar, $f(x, x_0 = L, t) = \delta$. Though far from a realistic representation of the complexities of the bar-mallet interaction, the constraint is considered justified inasmuch as the present focus is on the properties of the bar, not the impact, which is expected to convey little if any information about bar properties.[1] The derivation of the amplitudes for the cylindrical bar clamped at one end is given by Morse and Ingard (1968). For a total force P of the impulse applied to the free end of the bar,

$$A_n = (-1)^{n-1}\frac{PL}{\pi^3 r^3}\sqrt{\frac{32}{\rho E}}\beta_n^{-2}. \tag{4}$$

For the decay moduli τ_n, we consider separately the cases for which the predominant source of damping is external (fluid) and internal (viscoelastic). There are two reasons for this apart from expediency. First, of the two sources of damping in the model, only the viscoelastic component is directly related to the material of the bar. Later, in fact, we will discuss work in which a quantity related to the viscoelastic damping coefficient η is identified as a source of invariant acoustic information for material (Wildes and Richards, 1988). Second, empirical measurements by Chaigne and Doutaut (1997) on wood bars show the viscoelastic damping to be, by far, the predominant source of damping, fluid damping having but a small influence at the lowest frequencies, $f_n < 400$ Hz. Considering only the viscoelastic component the model gives, consistent with measurements,

$$\tau_n^{(v)} = \frac{1}{2\pi^2 \eta f_n^2}, \tag{5}$$

where (v) denotes the effect of viscoelastic losses on τ_n [cf. Eqs. 18 and 19 of Chaigne and Doutaut, 1997]. In the present application the values of η for different materials are taken from tables published by Cremer (2005).

For comparison to this model, we consider an alternative form of fluid damping proposed by Morse and Ingard (1968) in which the damping by air is given by a fixed force per length of the bar. Unlike Chaigne–Doutaut's model of fluid damping, this model yields an expression for the decay moduli, which depends critically on the material and geometric properties of the bar over a wide frequency range. The relation is

$$\tau_n^{(a)} = c\left(\frac{\pi\rho}{r^2 f_n^3}\right), \tag{6}$$

where (a) denotes the effect of air damping and $c = 16 \times 10^8$ is an air damping factor. This model has been used in the psychoacoustic studies of Lutfi and Oh (1997) and Lutfi (2001). It yields similar values of τ to the Chaigne and Doutaut (1997) model in some cases, but quite different values in others. Beyond increasing the generality of the analysis, the specific advantage of considering the two forms of damping separately is that it greatly simplifies the analysis. Inspection of Eqs. (3)–(6) reveals that the information in successive partials is redundant, the acoustic parameters being scaled in each case by a constant factor β_n raised to a constant power. This means that sensitivity to changes in individual partials of the sound, and thus sensitivity to change in the sound as a whole, can be estimated using a single computational formula,; only the values of the parameters of this formula need change with the different partials. Empirical support for this approach is later given in Sec. IV, and possible exceptions are considered in Sec. V.

B. Determination of null regions

Having specified a model of the resonant source, the next step is to determine regions of the physical parameter space given by the model for which changes are likely to be below the listener's threshold for detection; these are the NRs mentioned earlier. For this purpose, three threshold levels corresponding to three values of d' (=1, 2, or 3) are estimated from empirical psychometric functions (PFs) relating d' to changes x in the frequency, amplitude, and decay of pure tones (Abel, 1972; Dai, 1995; Buus and Florentine, 1991; Freyman and Nelson, 1986; Jesteadt and Bilger, 1974; Jesteadt and Sims, 1975; Jesteadt *et al.*, 1977; Wier *et al.*, 1977). The form of the PF typically used in these studies and the one that describes the data well is a power law

$$d' = ax^b, \tag{7}$$

where a and b are, respectively, the intercept and slope of the function fitted to the data in log-log coordinates (Green and Swets, 1966; Egan *et al.*, 1969; Buus and Florentine, 1991). For a given stimulus ϕ the magnitude of change has been variously expressed as $\Delta\phi$, $\Delta\phi/\phi$, and $\ln(1+\Delta\phi/\phi)$. However, where $0 < \Delta\phi/\phi \ll 1$, as it is for the discrimination of frequency, amplitude, and decay, these different representations can be shown to yield essentially equivalent fits to the data with an appropriate selection of a and b.[2] For consistency we choose a single measure of change to describe the PFs from all studies,

$$x = \ln\left(1 + \frac{\Delta\phi}{\phi}\right), \tag{8}$$

where $\Delta\phi/\phi$ is Weber's fraction. We also choose the slope of the PF for each ϕ to be constant across its range. For the discrimination of tone amplitude ($\phi = A$) the slope of the PF given by Eqs. (7) and (8) is found to be essentially constant over most of the audible range of A and just slightly less than 1, $b = 0.94$ (Buus and Florentine, 1991).[3] Similarly, for the

discrimination of frequency ($\phi=f$), a constant slope of $b=0.72$ appears to describe the data reasonably well (Jesteadt and Sims, 1975; Dai, 1995; Jesteadt and Bilger, 1974). Comparable data on PF slopes for the discrimination of decay are lacking; however, it seems likely that the slopes would follow closely those for the discrimination of overall duration (T), given that a change in duration accompanies any change in decay. Indeed, as we discuss shortly, the agreement between duration discrimination thresholds for signals with and without exponential decay is quite good. Abel (1972) showed the values of ΔT (for $d'=1$) derived from the PFs for the discrimination of tone duration to be well predicted by the corresponding change in energy over a range of $T=0.16-960$ ms. Based on this result we assume the PFs for the discrimination of tone decay to behave like those for the discrimination of tone amplitude described above.

For the intercepts of the PFs we use the differential thresholds ($\Delta\phi$) obtained in studies corresponding to a given value of d'. For the discrimination of frequency the most extensive data are provided by Wier *et al.* (1977). These authors showed that at moderate sensation levels (40 dB SL) the relation

$$\Delta f = 10^{0.026\sqrt{f}-0.533} \tag{9}$$

accurately describes the data over a wide range of frequencies (0.2–8.0 kHz), deviating only a few Hz at levels near the threshold of audibility. Jesteadt *et al.* (1977) provided complementary data for the discrimination of tone intensity (I). They showed threshold ΔI to be essentially independent of frequency and well described by the relation $\Delta I=0.463(I/I_0)^{0.928}$, where I_0 is the threshold for detection as reported in their study. Converting this to a differential threshold for amplitude yields

$$\Delta A = 0.680(A/A_0)^{0.928}, \tag{10}$$

where A_0 is the detection threshold in amplitude. Van Heuven and Van Den Broecke (1979) conducted the most extensive study involving the discrimination of pure-tone decay. However, they used a method of adjustment procedure, which precludes a measure of d'. We rely, instead, on the data of Schlauch *et al.* (2001), who used an objective three-alternative, forced-choice procedure to measure thresholds. The signals in the study of Schlauch *et al.* (2001) were noise bursts instead of tones; however, our informal analyses indicate that the threshold estimates from the two studies are in reasonably good agreement if one assumes the equivalent of a low false-alarm rate in the study of Van Heuven and Van Den Broecke (1979). The relevant data from Schlauch *et al.* (2001) are given in their Fig. 7 (panel 5, open symbols) where threshold $\Delta T/T$ is plotted against $T=10-100$ ms on log-log coordinates. A linear regression provides an excellent fit to these data yielding the relation $\Delta T=0.93T^{0.535}$. Schlauch *et al.* (2001) chose the time constants for decay to be one-fifth of the total duration T of their signals. Hence, converting this relation, we have

$$\Delta\tau = 0.439\tau^{0.535}, \quad 2 \le \tau \le 20 \text{ ms}. \tag{11a}$$

For longer duration decay constants we use the ΔT thresholds of Abel (1972). This seems justified inasmuch as the reported fits to her data agree remarkably well with those of Schlauch *et al.* (2001) over the same range of durations (cf. $\Delta T=1.09T^{0.449}$ and $T=0.16-80$ ms). The curves corresponding to these fits accounted for more than 97% of the variance in the data and are shown in Fig. 1. Note that two curves were fitted to the data to account for a break point near $T=100$ ms. The curve for the longer durations $T=100-1000$ is given by the relation $\Delta T=0.38T^{0.722}$. To estimate the corresponding relation for $\Delta\tau$ we take one-fifth the duration as before since this previously yielded excellent agreement to the Schlauch *et al.* (2001) data at the shorter durations in Fig. 1. We have

$$\Delta T_{1/5} \rightarrow \Delta\tau = 0.243\tau^{0.722}, \quad 20 < \tau \le 200 \text{ ms}. \tag{11b}$$

This and the other relations giving the slopes and intercepts of the PFs used in this analysis are summarized in Table I.

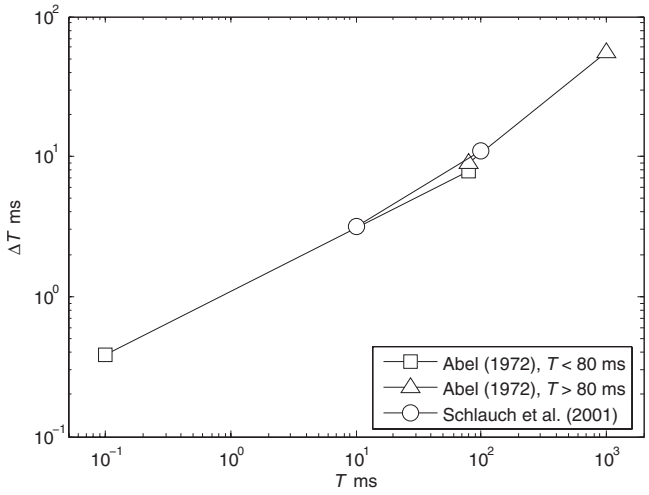


FIG. 1. Comparison of the best-fitting curves to the duration discrimination (ΔT) thresholds of Abel (1972) (triangles and squares) and Schlauch *et al.* (2001) (circles). In the Schlauch *et al.* (2001) study the signals decayed exponentially with decay constant equal to one-fifth the duration of the signal. In the Abel (1972) study the signals were gated on and off rectangularly.

TABLE I. Summary of the parameters of the PFs used in the analysis. Values of the intercepts a are computed from the data in columns 3, 4, and 6 according to the relation $a=d'[\ln(1+\Delta\phi/\phi)]^{-b}$. Columns 5 and 7 give, respectively, the range of values over which the intercepts a and slopes b were measured in the referenced studies.

ϕ	Studies	$\Delta\phi$	d'	Range	b	Range
f	[a–d]	$10^{0.026\sqrt{f}-0.533}$	0.78	0.2–8.0 kHz	0.72	0.2–1.0 kHz
A	[d–f]	$0.680(A/A_0)^{0.928}$	0.78	5–80 dB SL	0.94	30–90 dB SPL
τ	[g]	$0.439\tau^{0.535}$	1.35	2–20 ms	⋯	⋯
	[h*]	$-0.248\tau^{0.722}$	0.95	20–200 ms	0.94	0.03–200 ms

[a]Wier *et al.*, 1977.
[b]Dai, 1995.
[c]Jesteadt and Bilger, 1974.
[d]Jesteadt and Sims, 1975.
[e]Buus and Florentine, 1991.
[f]Jesteadt *et al.*, 1977.
[g]Schlauch *et al.*, 2001.
[h]Abel, 1972.
*Estimates based on $T_{1/5}$.

TABLE II. Values of exponents p_i from Eqs. (3)–(6) relating physical to acoustic parameters, used to compute NRs according to Eq. (15).

	E	ρ	η	L	r
f	1/2	−1/2	⋯	−2	1
A	−1/2	−1/2	⋯	1	−3
$\tau^{(v)}$	−1	1	−1	4	−2
$\tau^{(a)}$	−3/2	5/2	⋯	6	−5

We are now in a position to derive an analytic expression giving d' for the change in any one of the three acoustic parameters $\phi \in \{f, A, \tau\}$ corresponding to the change in any combination of the five physical parameters $\theta \in \{\rho, E, \eta, L, r\}$. For this we note that the general form of the relation between ϕ and θ given by Eqs. (3)–(6) is

$$\phi = \lambda \prod_i \theta_i^{p_i}, \tag{12}$$

where λ is a constant and the p_i are the exponents on the physical parameters corresponding to ratios of small integers (given in Table II). The values of θ_i in this case do not depend on one another; hence, for any given $\theta_{i=k}$,

$$\phi = \gamma \theta_k^{p_k}, \tag{13a}$$

$$\phi + \Delta\phi = \gamma(\theta_k + \Delta\theta_k)^{p_k}, \tag{13b}$$

where γ is a new constant. Substituting these expressions into Eq. (8) yields

$$x = p_k \ln\left(1 + \frac{\Delta\theta_k}{\theta_k}\right), \tag{14}$$

which upon rearranging terms from Eq. (7) gives

$$\Delta\theta_k = \theta_k\left[\exp\left(\frac{d'}{a p_k^b}\right)^{1/b} - 1\right]. \tag{15}$$

In what follows the values of a, b, and d' from the studies referenced in Table I and the exponents p_i from Eqs. (3)–(6) (and listed in Table II) are used to compute null regions for each acoustic parameter in the physical parameter space according to Eq. (15). For material the physical parameter space included three variables: bar density ρ, Young's modulus E, and the coefficient of viscoelasticity η. For geometry there were two variables: bar radius r and bar length L. The effect of two other aspects of bar geometry, shape and hollowness, are principally the same as for bar radius and so are treated along the way. Table III summarizes all variables included in the analysis and their definitions, while Table IV gives the specific values of the physical parameters used. The values of ρ and E in Table IV correspond nominally to wood (oak), glass, aluminum, and plastic (plexiglass) as given in published tables (Cremer, 2005, pp. 191 and 196). So too, the values of η can be considered only nominal as there is much reported variability in their measurement (Cremer, 2005). Finally, the values of r and L were chosen for these materials to produce frequencies and decay moduli that extended over most of the audible range above 400 Hz, where, as earlier noted, the predominant source of damping in the Chaigne and Doutaut (1997) model is viscoelastic.

TABLE III. Definitions of variable quantities and their units.

Symbol	Definition (unit)
f	Frequency (Hz)
A	Amplitude
$\tau_n^{(a)}$	Decay modulus, fluid damping model (s)
$\tau_n^{(v)}$	Decay modulus, viscoelastic damping model (s)
E	Young's modulus (dyn/cm^2)
ρ	Mass density (g/cm^3)
L	Bar length (cm)
r	Bar radius (cm)
κ	Radius of gyration (cm)
η	Coefficient of viscoelasticity (ns)
c	Air damping factor
ϕ	Generic symbol for acoustic parameter
θ	Generic symbol for physical parameter
a	Intercept of the psychometric function
b	Slope of the psychometric function
d'	Index of discriminability
p_i	Exponents relating acoustic to physical parameters

III. ANALYTIC RESULTS

A. Null regions for material properties

Example NRs in the $\rho \times E$ parameter space are given by the color-shaded regions in Fig. 2; red, green, and blue, respectively, for frequency, amplitude, and decay. The decay is for the fluid damping model; viscoelastic damping is considered later. Increasing values of $d' = 1-3$ are represented by shade from dark to light for both increments and decrements in the acoustic parameters. The NRs are for the first partial ($n=1$) of a small glass (Gl) and wood (Wd) bar, left and right panels, respectively (note the cross-over of the NRs at the points for glass and wood). The NRs give the predicted discriminable changes in the first partial corresponding to a change in the material of the bar from glass or from wood. For example, in going from glass to aluminum, the change in frequency is estimated to be highly discriminable ($d' \gg 3$), whereas the change in decay is not ($d' < 3$). The properties of the other bars selected for investigation are listed in Table IV. These properties were selected so that the first partial, at least, would fall within the range for which there were psychophysical data relevant to the discrimination of decay. Over this range the different geometries had little noticeable effect on the NRs. The largest effect overall was on fluid damping in going from the small to the thick wood bar; the corresponding NR ($d' = 3$) for the thick wood bar is given by the blue dashed line in the right panel of Fig. 2.

The NRs of Fig. 2, as noted, are for the first partial. To compute the NRs for the complex as a whole one would next need to determine the sensitivities to changes in the remaining audible partials and combine these in some fashion to estimate sensitivity to changes in the complex as a whole (cf. Moore *et al.*, 1984; Gockel *et al.*, 2007). Though clearly doable, our analyses indicate that the outcome would hardly differ from that for the first partial alone. Figure 3 demonstrates the reason why. It shows by shade the NRs ($d' = 3$) for the first three partials of the glass and wood bar (the darkest shade representing the lowest frequency partial, and the

TABLE IV. Values of physical and acoustic parameters corresponding to the reference bars represented by the NRs in Fig. 1. Amplitude of the reference is fixed at 40 dB SL. The estimates of the Weber fractions, $\Delta\theta/\theta$, for the physical parameters in columns 9–13 is based on the NR for frequency for $d'=1$.

Bar	ρ (g/cm^3)	E (dyn/cm$^2\times10^{11}$)	η (ns)	$L\times r$ (cm$\times$cm)	f_1 (Hz)	$\tau_1^{(a)}$ (ms)	$\tau_1^{(v)}$ (ms)	$\Delta E/E$ $\times10^{-3}$	$\Delta\rho/\rho$ $\times10^{-3}$	$\Delta\eta/\eta$ $\times10^{-3}$	$\Delta L/L$ $\times10^{-5}$	$\Delta r/r$ $\times10^{-3}$
Lg Wd	0.7	1.00	100	20×6	1587	25	201	5.7	5.6	5.6	1.9	2.8
Sm Wd	0.7	1.00	100	10×3	3174	12	50	7.6	7.6	7.6	0.6	3.8
Tk Wd	0.7	1.00	100	15×6	2821	4	64	7.1	7.0	7.1	0.9	3.5
Lg Gl	2.5	6.00	30	20×9	3085	5	177	7.5	7.4	7.4	1.2	3.7
Sm Gl	2.5	6.00	30	10×3	4114	20	100	9.4	9.3	9.4	0.6	4.7
Tk Gl	2.5	6.00	30	15×6	3657	7	126	8.5	8.4	8.4	0.9	4.2
Sm Al	2.7	7.20	10	9×3	5354	10	177	12.4	12.2	12.4	0.5	6.2
Tk Al	2.7	7.20	10	15×6	3855	7	34	8.9	8.8	8.8	0.9	4.4
Lg Pl	1.1	0.60	200	20×9	1471	21	117	5.6	5.6	5.6	1.9	2.8
Sm Pl	1.1	0.60	200	10×3	1961	81	66	6.0	6.0	6.0	0.8	2.8
Tk Pl	1.1	0.60	200	15×6	1743	29	83	5.8	5.8	5.8	1.3	2.9

lightest shade the highest). The geometries of the bars in this case ($r=1.5$ and $L=20$ cm) were selected so that all three partials would fall within the audible range.[4] What is evident from the figure is that the NRs are either unchanged (as for amplitude), or they become much larger with increasing order of the partial (for frequency and decay). The increase in the NRs for frequency and decay is a general result and is to be expected since the respective Weber fractions are larger for the higher frequencies and shorter decays associated with the higher order partials [cf. Eqs. (9), (11a), and (11b)]. The significant point to be made here is that changes in the few higher order partials that are audible will contribute little if at all to overall sensitivity if changes in the first partial are already below threshold for detection. For simplicity of presentation we focus in what follows on the NRs for the first partial alone as a limiting case. A subsequent empirical test is provided in Sec. IV.

Returning to Fig. 2, several gross features of the NRs are worth noting. First, the NRs for frequency are by far the smallest. This is due to the fact that small changes in bar density and elasticity produce relatively large changes in frequency [cf. Eq. (3)] while, at the same time, the Weber fraction for the discrimination of frequency is exceedingly small, $\Delta f/f<0.005$. Table IV shows what is predicted to be the smallest detectable change in either material density or elasticity ($d'=1$) based on the NRs for frequency. The smallest detectable change in each case is of the same order of magnitude as the Weber fraction for the discrimination of frequency. The implication of the small frequency NRs is that changes in frequency could provide a robust cue for material when differences in material are small. This might go some way in explaining the results of the aforementioned studies where listeners have been observed to base their judgments of material on simple differences in frequency (Lutfi and Oh, 1997; Giordano and McAdams, 2006). We will return to this point in Sec. V.

A second feature to note is that the NRs for amplitude are roughly orthogonal to the NRs for frequency and decay (although not shown, this is also true for decay due to viscoelastic damping). As such, changes in amplitude, in special circumstances, could provide an independent source of information about material. A case in point is given in the left panel of Fig. 2 where the change from glass to aluminum produces a greater increase in d' for amplitude than it does for $\tau^{(a)}$. This specific case was, in fact, investigated by Lutfi and Oh (1997). They used the fluid damping model to synthesize impact sounds of glass, aluminum, and quartz bars, quartz being very similar to aluminum in mass density and elasticity. Listeners were able to distinguish among these materials, but three of four listeners participating in the discrimination of glass from aluminum gave a significant,

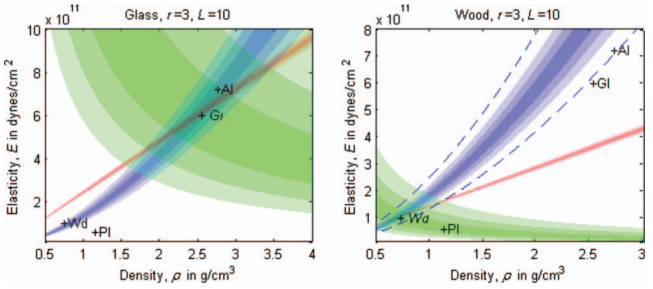


FIG. 2. NRs for changes in frequency (red), amplitude (green), and decay based on fluid damping (blue) are plotted as a function of bar density, ρ, and elasticity, E, for glass (Gl) and wood (Wd) bars in the left and right panels, respectively. The locations of plastic (Pl) and aluminum (Al) bars in this space are also shown (see Table IV). Shading of the NRs from dark to light represents increasing values of d' from 1 to 3. The NRs are for the small (Sm) bars in Table IV. The NRs are essentially unchanged for the other bar geometries in Table IV with the exception of the thick wood bar, for which the NR for decay ($d'=3$) is given by the dashed line.

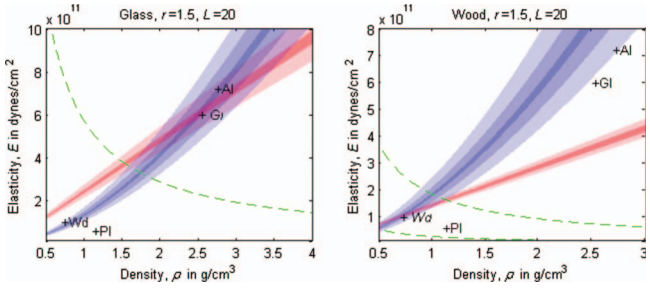


FIG. 3. NRs ($d'=3$) for the first three partials of the glass and wood bar, represented by shade with the darkest shade representing the lowest frequency partial and the lightest shade representing the highest frequency partial. The geometries of the bars ($r=1.5$ and $L=20$ cm) were selected so that all three partials would fall within the audible range.

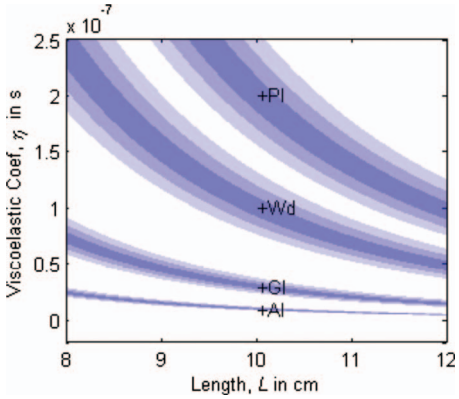


FIG. 4. NRs (blue) for changes in decay (viscoelastic damping) are plotted as a function of the viscoelastic damping factor η and bar length L for plastic (Pl), wood (Wd), glass (Gl), and aluminum (Al) bars (see Table IV). Shading from dark to light represents increasing values of d' from 1 to 3 as before.

though not predominant, weight to amplitude. All three listeners participating in the discrimination of glass from quartz gave a significant weight to amplitude with one giving a predominant weight to amplitude. Unfortunately, in many other studies, the role of amplitude cannot be gauged because the amplitudes are either not reported or the levels of signals are equalized in some manner.

Finally we note that for glass, in the left panel of Fig. 2, both aluminum and wood fall in or near the NR for $\tau^{(a)}$. This means that if mass density and elasticity were the only factors affecting decay, as is true for $\tau^{(a)}$, then a listener might often detect a change in frequency corresponding to the difference in these materials without detecting the corresponding change in decay. A similar situation might be thought to exist for $\tau^{(v)}$; however, $\tau^{(v)}$ also depends on the material parameter η whose effect is not evident from Fig. 2. Figure 4 shows the corresponding NRs for $\tau^{(v)}$ in the parameter space given by η and L. The figure reveals significant separation among the NRs plotted in this space, indicating clearly discriminable differences in $\tau^{(v)}$ for these materials. The Weber fractions for η based on the discriminable changes in frequency are listed in Table IV and are of the same order of magnitude as those for the discrimination of frequency. Viscoelastic damping as given by η is one of a number of internal damping factors that has figured prominently in the analysis of invariant acoustic information for material (Wildes and Richards, 1988). We return to this point in Sec. V.

B. Null regions for geometric properties

Figure 5 shows the NRs for changes in the parameter space given by $L \times r$, color- and shade-coded as before. In the left panel the bar is relatively large (Lg), having dimensions $L=20$ and $r=6$; in the right panel a thick (Tk) bar is represented with dimensions $L=15$ and $r=6$. The locations of the small (Sm) bars of Fig. 2 are also indicated. The NRs are for the glass bar; however, the differences between these NRs and those for the other materials in Table IV were barely noticeable.

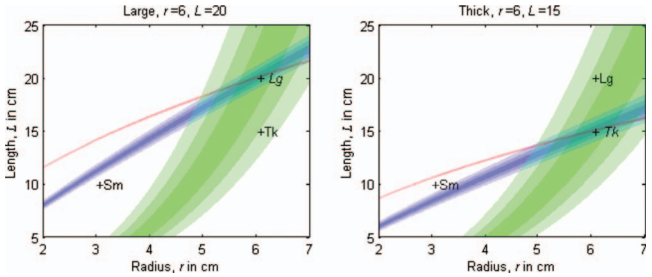


FIG. 5. NRs for changes in frequency (red), amplitude (green), and decay based on fluid damping (blue) are plotted as a function of bar length L and radius r for large (Lg) and thick (Tk) glass bars in the left and right panels, respectively. The location of the small (Sm) bars in this space is also shown (see Table IV). Shading of the NRs from dark to light represents increasing values of d' from 1 to 3.

Certain features of the NRs in Fig. 5 are similar to those for material. As was the case for material, the NRs for frequency are smallest. This again is to be expected as small changes in geometry, or material, tend to produce large changes in frequency where the Weber fraction is smallest. The corresponding Weber fractions for length and radius are given in Table IV. Here, the Weber fractions for length are clearly the smallest, being two orders of magnitude smaller than those for any of the other four physical parameters. This is expected from Eq. (3), where length has the largest-valued exponent and so produces proportionally the largest changes in frequency. A second common feature with material is that the NRs for amplitude, though not orthogonal to, cross the other NRs at considerable angles, particularly in the local region of the bar reference. The relation is such that, while a change in bar length is likely to produce a more discriminable change in frequency or decay than in amplitude, a change in bar radius with length may produce a more discriminable change in amplitude. This reinforces the point made earlier that amplitude in special cases could serve as a cue for bar properties.

In addition to bar length and radius, Fig. 5 provides estimates of sensitivity to changes in two other features of bar geometry: hollowness and shape. These properties affect the cross-sectional area of the bar [S in Eq. (1)] and so their acoustic effect is principally the same as that for bar radius. The specific effect is given by the radius of gyration κ, which is a measure of how the cross-sectional area of the bar is distributed about its center axis. For the cylindrical and rectangular bar $\kappa=r/2$ and $\kappa=r/\sqrt{12}$, respectively, where r is bar radius. For the hollow cylindrical bar $\kappa=\frac{1}{2}\sqrt{r^2+s^2}$, where s is the inner radius of hollowness (Morse and Ingard, 1968). Sensitivity to specific changes in each of these cases can be determined from the NRs in Fig. 5 by scaling the abscissa according to these expressions, but within the acoustic range investigated the general conclusions will be the same as for bar radius.

IV. EMPIRICAL RESULTS

Momentarily we consider how the general features of the NRs described above might account for some apparent discrepancies in past studies. First, however, we offer some preliminary data testing specific predictions of the NRs for

the different effects of fluid and viscoelastic damping on the identification of bar material. The experiment demonstrates an intended use of Eq. (15) as a vehicle for identifying conditions of interest. It also tests the earlier premise that sensitivity to changes in the first partial can be taken to reflect sensitivity to changes in the signal as a whole.

In a standard two-interval, forced-choice procedure with feedback listeners were presented the impact sounds (partials $n=1-3$ present) of the small glass and aluminum bars ($r=3$ and $L=10$ cm), those represented in Figs. 2 and 4. The listeners were asked to judge on each trial which of the two sounds was produced by the aluminum bar. In different conditions the impact sounds were synthesized according to either the fluid or the viscoelastic damping model. As mentioned earlier, both models yield a specific relation between frequency and decay as invariant acoustic information for material (more is said about this in Sec. V). However, the NRs of Figs. 2 and 4 predict that only for viscoelastic damping will the relation be above threshold for detection. For viscoelastic damping aluminum falls well outside the NRs for both frequency (Fig. 2) and decay (Fig. 4), but for fluid damping aluminum falls within the same NR for decay (Fig. 2). The prediction is that listeners will base their judgments of material on the combination of frequency and decay when the sounds are synthesized according to the viscoelastic damping model, but will base their judgments predominantly on frequency when the sounds are synthesized according to the fluid damping model.

A standard perturbation analysis procedure was used to test these predictions (see Berg, 1990 and Lutfi, 2001). On each presentation a small normal deviate was added to the length of the bar ($\sigma_L=0.2$ and 1.0 cm, respectively, for fluid and viscoelastic damping), and in the case of viscoelastic damping, to η ($\sigma_\eta=10^{-8}$ s). The latter perturbation served to decouple the perfect correlation between f and $\tau^{(v)}$, which was necessary for the analysis. It was not so great as to prevent the listener from performing perfectly if they based their judgments on $f\tau^{(v)}$. Correlations between each listener's judgments and the resulting perturbations in f, τ, and $f\tau$ were computed by multivariate regression using the glmfit routine of MATLAB with a logistic link function. The correlations, based on the average of 400 trials, were then converted to relative decision weights by normalizing with respect to the sum of their absolute values (cf. Berg, 1990). The listeners were one male and two female University of Wisconsin students (ages 28–29). All had previous extensive experience in the task and all had normal hearing by standard audiometric tests (ANSI S3.6-1996, 1996).

The results of the experiment are given in Fig. 6 where the obtained decision weights (symbols representing different listeners) are compared to predictions in each condition (horizontal lines). Also shown (legend) is the predicted and obtained proportion of correct responses. Though the obtained weights were not exclusively for one or the other cue, the results are in general agreement with the predictions. All listeners show a predominant weight on f for the fluid damping model and a predominant weight on $f\tau$ for the viscoelastic damping model. Moreover, all listeners as predicted perform more poorly with the fluid damping model. The results

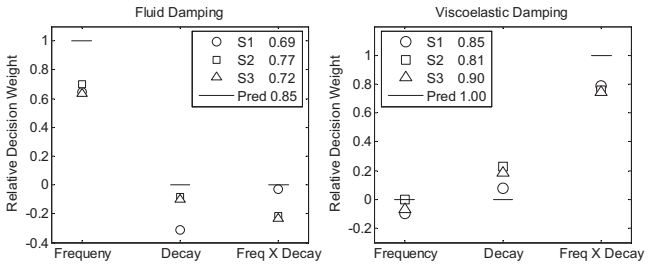


FIG. 6. Obtained decision weights of listeners (symbols) are given for the two-interval, forced-choice distinction between the impact sounds of the small glass and aluminum bars. Left and right panels show, respectively, the results when the sounds were synthesized according to the fluid and viscoelastic damping models. Predictions for the data in each case were derived from the NRs of Figs. 2 and 4 and are given by the short horizontal lines in each panel.

support the use of the first partial as an index of sensitivity for changes in the complex as a whole, although it is always possible to use Eq. (15) to estimate the additional influence of the higher order partials. They also demonstrate how Eq. (15) might be used as practical tool in selecting conditions where one wishes to examine the role of limited auditory sensitivity or where one wishes to select conditions for which limited auditory sensitivity can reasonably be ruled out as a factor.

V. DISCUSSION

We return now to consideration of the possible influence of auditory sensitivity on the availability of invariant acoustic information for bar attributes in past studies. Wildes and Richards (1988) provided the seminal work on the invariant acoustic information for material. Their analysis focuses on the damping due to internal friction created within the object as it vibrates. Internal friction may have several sources, viscoelastic processes being one, but the degree of damping resulting collectively from these sources is commonly measured by the material's coefficient of internal friction, tan φ. The coefficient gives the ratio of elastic moduli associated with the stress and strain acting on the object. Wildes and Richards (1988) recognized that, as a unique property of materials, tan φ could convey invariant acoustic information for material through its effect on the decay of partials $\tau^{(e)}$. Citing literature on anelastic solids (Zener, 1948), they proposed a model in which $\tau^{(e)}$ at any frequency f is linearly related to tan φ as follows:

$$\tan\varphi = \frac{1}{\pi f \tau^{(e)}}. \tag{16}$$

Thus, assuming that internal friction is the predominant source of damping, a listener who has access to the frequency f and decay $\tau^{(e)}$ of any one partial could, in theory, recover a measure of internal friction that would allow for a material classification. That measure would be a frequency-independent decay $f\tau^{(e)}$, corresponding to the number of cycles required for the partial to decay to $1/e$ of its original value.

Other cues for material, since implicated in studies, can be shown to be closely related to $f\tau^{(e)}$. In the Chaigne and

Doutaut (1997) model, for example, on which the present analysis is based, internal friction is determined by the viscoelastic damping factor η—a unique material property. From Eq. (5) the corresponding acoustic information for material is given by $f\tau^{(v)}$ for any partial (Note that unlike tan φ, η has units of seconds; hence, f is not squared in this relation). Also, because both Eqs. (5) and (16) dictate an inverse relation between frequency and decay, the spectral centroid of the impact sound will shift to lower frequencies over time. McAdams *et al.* (2004) identified this cue as a correlate of η in listeners' dissimilarity ratings of impact sounds. Given the close relation among these cues a strong correlation of listener judgments with any one would likely yield a strong correlation with the others. It probably matters little therefore that most studies, in fact, have restricted their analyses to $f\tau^{(e)}$. Collectively, these studies show that listeners are, indeed, able to judge the material of bars and plates from their emitted sound, regardless of whether the sound is live, recorded, or synthesized based on a physical model (Gaver, 1988; Lutfi and Oh, 1997; Kunkler-Peck and Turvey, 2000; Carello *et al.*, 1998; Tucker and Brown, 2003; Giordano and McAdams, 2006). However, the results are mixed as to whether listeners' judgments are based on $f\tau_e$. Using synthesized sounds of struck clamped bars and a forced-choice classification of material, Lutfi and Oh (1997) found the frequency f of partials to be a better predictor than $f\tau_e$ of the listeners' trial-by-trial judgments. Hermes (1998) reported a similar outcome for both free- and forced-choice classifications of synthesized impact sounds. Klatzky *et al.* (2000), on the other hand, used synthesized sounds of struck, clamped bars, very similar to those of Lutfi and Oh (1997), and reported greater reliance of judgments on $f\tau_e$ than on f for both material classification and similarity judgments of material. Their results appear to be corroborated by Avanzini and Rocchesso (2001) using synthesized impact sounds, and by Tucker and Brown (2003) using recordings of sounds from loosely suspended plates, although Tucker and Brown (2003) only analyzed for $f\tau_e$. Giordano and McAdams (2006) obtained forced-choice judgments of material using live sounds generated by loosely suspended plates. They found judgments to be more strongly correlated with f when materials were confused (as in the case of steel vs glass and wood vs plastic), but more strongly correlated with $f\tau_e$ when they were not (as in the case when comparisons were made across these material pairs). They suggested based on this outcome that f may mediate judgments for materials having similar physical properties, whereas $f\tau_e$ may mediate judgments for materials having very different physical properties.

Many of the apparent discrepancies among these studies may be understood by considering the predicted effect of auditory sensitivity based on Eq. (15). A key prediction is that differences in decay associated with different materials are less likely to be detected if the predominant source of damping is external to the object (e.g., fluid) as opposed to internal (e.g., viscoelastic). Klatzky *et al.* (2000) and Avanzini and Rocchesso (2001) synthesized their sounds using precisely the internal damping model proposed by Wildes and Richards (1988) [Eq. (16) above]. These studies involved classification of wood and glass, among other materials, which according to the NRs of Fig. 4 would be clearly identifiable based on $f\tau^{(v)}$. Lutfi and Oh (1997), on the other hand, used the fluid damping model given by Eq. (6). Their study involved the classification of glass and aluminum, among other materials, which according to the NRs of Fig. 2 would produce a clearly discriminable change in f, but a much less discriminable change in $\tau^{(a)}$. Hermes (1998), whose results were similar to those of Lutfi and Oh (1997), also used a means of synthesizing sounds that did not include the influence of internal damping.

Though Eq. (15) was developed specifically for bars, the predicted difference in sensitivity to the source of damping may also account for the different results obtained in the studies using real sounds emitted by plates. Plates, by virtue of their geometry, are subject to significant damping by air, generally more so than bars. Moreover, in the studies using plates, the plates have had large surface areas, thus increasing the expected influence of fluid damping. Giordano and McAdams (2006), for example, reported the values of f_1 for nearly all their plates to fall below 400 Hz, well within the region where fluid damping is expected to have a significant impact for bars according to the model of Chaigne and Doutaut (1997). As fluid damping begins to play a more prominent role, the expectation from Fig. 2 is that confusions would first occur between the generic categories of glass and metal (aluminum) since both fall within the same NR for decay. Finally, the NRs of Fig. 2 suggest that as the difference in materials to be classified becomes large amplitude would become a salient cue. Giordano and McAdams (2006), in fact, observed that, related to amplitude, "peak loudness" was as highly correlated with listener judgments in such cases.[5]

The results of studies investigating auditory perception of size and shape are somewhat more difficult to interpret than those for material. To begin, these studies find that size and shape are not nearly as well judged from sound as material (Gaver, 1988, 1993a, 1993b; Tucker and Brown, 2003; Kunkler-Peck and Turvey, 2000; Carello *et al.*, 1998; Ottaviani and Rocchesso, 2004; Rocchesso, 2001; Lakatos *et al.*, 1997). This may have to do with the fact that, unlike material, information about size and shape can be conveyed through the spatial dimension of sound, which is lacking in studies where the sounds were recordings (cf. Lutfi, 2008). Somewhat better, though still rather poor, performance has been observed in studies using live sounds (Kunkler-Peck and Turvey, 2000; Carello *et al.*, 1998). Also there is considerably more variation across these studies in the manner in which the sounding objects were excited to vibrate. In the Gaver (1988) study, for example, the objects were bars struck with a mallet as they rested on a carpeted surface; in the Carello *et al.* (1998) study they were rods that were dropped on a hard floor and allowed to bounce. Notwithstanding these complications, two general statements can be made from the present analysis regarding the identification of bar size and shape. First, because the acoustic effect of these properties according to the model is determined strictly by the bar's length L and radius of gyration κ, there are no additional parameters that might be expected to yield better sensitivity for size and shape than is indicated by the NRs of

Fig. 5. Notably, this was not the case for material, where differences in η yielded clearly discriminable changes in $\tau^{(v)}$ beyond those of the other two material parameters, E and ρ. This distinction, in fact, might go some way in explaining the generally better performance observed for the identification of material in past studies. Second, the NRs for frequency in Fig. 5 would support the informal conjecture of past authors that pitch is a salient cue for bar size, or at least bar length (Gaver 1988, 1993a, 1993b). As earlier noted, the estimated Weber fraction for a change in bar length based on the change in frequency is by far the smallest of all physical parameters investigated.

Compared to size and shape, one aspect of bar geometry that does not have a significant spatial acoustic component is hollowness. Lutfi (2001) analyzed the invariant acoustic information for the hollowness of a cylindrical bar using the fluid damping model given by Eq. (6). Consistent with the foregoing analysis for material, the solution was given by a specific relation between the frequency and decay for any one partial. In the same study listeners were asked to judge hollowness in a two-alternative, forced-choice task using synthesized sounds of hollow and solid bars of different materials and sizes. A regression analysis of each listener's trial-by-trial responses showed for half of the listeners a strong correlation with the relation between frequency and decay, but for the other half of the listeners a stronger correlation with f alone. Simulations of an internal noise model applied to these conditions suggested that the reliance on f reflected the robust nature of this cue with respect to sensory noise. The hypothesis can be checked by applying Eq. (15). Recall that the acoustic effect of hollowness for the cylindrical bar is given by $\kappa=\frac{1}{2}\sqrt{r^2+s^2}$, where s is the inner radius of hollowness. For the solid aluminum bar used in the study ($L=10$ cm and $r=2$ cm), Eq. (15) yields a value of $\Delta\kappa=0.1$ cm for $d'=3$ based on $\Delta\tau^{(a)}$, and a value of $\Delta\kappa=0.03$ based on Δf. The actual value of $\Delta\kappa$ in the study was 0.1 cm ($s=1$ cm) well outside the NR for f, but right on the border of the NR for $\Delta\tau^{(a)}$. A similar outcome is obtained for the wood bar in that study. The predictions of Eq. (15) thus generally support the outcome of the programed simulations by Lutfi (2001).

We conclude by noting some shortcomings of the analysis. First, because the analysis is based on data measuring sensitivity to changes in single tones, it does not take into account possible interactions that can occur among partials (e.g., mutual masking, combination tones, beats, and the like). This is not likely to be a problem for the widely spaced transverse modes of the homogenous bar [cf. Eq. (3)], and even for plates there are likely to be resolved partials that adequately convey information about source properties. Still, there are degenerate cases (e.g., square plates) where few if any partials are likely to be separately resolved. In these cases, Eq. (15) would not apply. Second, the derivation of Eq. (15) assumes that sensitivity to one acoustic parameter is not affected by the value of the others. This, in fact, is largely true over much of the parametric range investigated (Jesteadt *et al.*, 1977; Wier *et al.*, 1977; Abel, 1972; Freyman and Nelson, 1986). The one notable exception is for the discrimination of frequency where threshold Δf can be an order of magnitude greater for durations below 20 ms than above (Freyman and Nelson, 1986). In the present study the durations were never less than 20 ms, but in applications where they are a correction to Eq. (15) would have to be imposed. Finally, the psychophysical data on which the analysis is based are somewhat weak for the discrimination of the longer decay constants ($\tau>100$), where our estimates have been taken from data on the discrimination of duration.[6] Future studies may help to correct this shortcoming. Notwithstanding these limitations, Eq. (15) should provide, at least, a preliminary workable framework for estimating the bound limited sensitivity places on the auditory identification of rudimentary bar attributes.

ACKNOWLEDGMENTS

The authors wish to thank Dr. B. C. J. Moore and two anonymous reviewers for helpful comments on a previous version of this manuscript. This research was supported by NIDCD Grant No. 5R01DC006875-05.

[1]One can, of course, imagine conditions where the manner of impact could provide information about bar material or geometry, as, for example, where the surface area of contact is affected by bar shape. Generally, however, we expect the effects of bar material and geometry on the acoustics to be the primary source of information about these properties.

[2]This simply arises from the fact that for $0<\Delta\phi/\phi\ll 1$, $\ln(1+\Delta\phi/\phi)\approx\Delta\phi/\phi$.

[3]Some previous studies have reported different results. However, Buus and Florentine (1991) pointed to the small range of intensity increments used in these studies as the likely reason for the discrepancies.

[4]Note that for all three partials to fall within the audible range the decay of the first partial had to exceed that for which there are available psychophysical data on the discrimination of decay. The outcome was the same, however, when the decay of the first two partials fell within the range of the psychophysical data.

[5]In Giordano and McAdams, 2006 the peak levels of signals ranged from 54 to 72 dB sound pressure level (SPL). Loudness was computed by passing signals through a bank of auditory filters, compressing the output of each filter and then summing the individual loudness of the partials according to the model of Zwicker and Fastl (1999).

[6]An additional complicating factor here is that durations of the sounds can be so long that, in practice, they would need to be truncated. This, in fact, has been the approach taken in a number of past studies.

Abel, S. M. (**1972**). "Duration discrimination of noise and tone bursts," J. Acoust. Soc. Am. **51**, 1219–1223.

ANSI S3.6-1996 (**1996**). American National Standards Specification for Audiometers, American National Standards Institute, New York.

Avanzini, F., and Rocchesso, D. (**2001**). "Controlling material properties in physical models of sound objects," Proceedings of the International Computer Music Conference, La Habana, Cuba, pp. 17–22.

Berg, B. G. (**1990**). "Observer efficiency and weights in a multiple observation task," J. Acoust. Soc. Am. **88**, 149–158.

Buus, S., and Florentine, M. (**1991**). "Psychometric functions for level discrimination," J. Acoust. Soc. Am. **90**, 1371–1380.

Carello, C., Anderson, K. A., and Kunkler-Peck, A. J. (**1998**). "Perception of object length by sound," Psychol. Sci. **9**, 211–214.

Chaigne, A., and Doutaut, V. (**1997**). "Numerical simulations of xylophones. I. Time-domain modeling of the vibrating bars," J. Acoust. Soc. Am. **101**, 539–557.

Cremer, L. (**2005**). *Structure-Borne Sound: Structural Vibrations and Sound Radiation at Audio Frequencies* (Springer, New York).

Dai, . (**1995**). "On measuring psychometric functions: A comparison of the constant-stimulus and adaptive up–down methods," J. Acoust. Soc. Am. **98**, 3135–3139.

(**2004**). *Ecological Psychoacoustics*, edited by J. G. Neuhoff (Academic, New York).

Egan, J. P., Lindner, W. A., and McFadden, D. (**1969**). "Masking-level dif-

ferences and the form of the psychometric function," Percept. Psychophys. **6**, 209–215.
Freed, D. J. (**1990**). "Auditory correlates of perceived mallet hardness for a set of recorded percussive sound events," J. Acoust. Soc. Am. **87**, 311–322.
Freyman, R. L., and Nelson, D. A. (**1986**). "Frequency discrimination as a function of tonal duration and excitation-pattern slopes in normal and hearing-impaired listeners," J. Acoust. Soc. Am. **79**, 1034–1044.
Gaver, W. W. (**1988**). "Everyday listening and auditory icons," Ph.D. thesis, University of California, San Diego, CA.
Gaver, W. W. (**1993a**). "What in the world do we hear? An ecological approach to auditory event perception," Ecological Psychol. **5**, 1–29.
Gaver, W. W. (**1993b**). "How do we hear in the world? Explorations in ecological acoustics," Ecological Psychol. **5**, 285–313.
Gibson, J. J. (**1966**). *The Senses Considered as Perceptual Systems* (Houghton-Mifflin, Boston, MA).
Giordano, B. L., and McAdams, S. (**2006**). "Material identification of real impact sounds: Effects of size variation in steel, glass, wood and plexiglass plates," J. Acoust. Soc. Am. **119**, 1171–1181.
Gockel, H., Moore, B. C. J., Carlyon, R. P., and Plack, C. J. (**2007**). "Effect of duration on the frequency discrimination of individual partials in a complex tone and on the discrimination of fundamental frequency," J. Acoust. Soc. Am. **121**, 373–382.
Green, D. M., and Swets, J. A. (**1966**). *Signal Detection Theory and Psychophysics* (Wiley, New York).
Hermes, D. J. (**1998**). "Auditory material perception," IPO Annual Progress Report **33**, 95–102.
Jesteadt, W., and Bilger, R. C. (**1974**). "Intensity and frequency discrimination in one- and two-interval paradigms," J. Acoust. Soc. Am. **55**, 1266–1276.
Jesteadt, W., and Sims, L. (**1975**). "Decision processes in frequency discrimination," J. Acoust. Soc. Am. **57**, 1161–1168.
Jesteadt, W., Wier, C. C., and Green, D. M. (**1977**). "Intensity discrimination as a function of frequency and sensation level," J. Acoust. Soc. Am. **61**, 169–177.
Kac, M. (**1966**). "Can one hear the shape of a drum?," Mathematics Monthly **13**, 21–23.
Klatzky, R. L., Pai, D. K., and Krotkov, E. P. (**2000**). "Perception of material from contact sounds," Presence: Teleoperators and Virtual Environment **9**, 399–410.
Kunkler-Peck, A. J., and Turvey, M. T. (**2000**). "Hearing shape," J. Exp. Psychol. **26**, 279–294.
Lakatos, S., McAdams, S., and Causse, R. (**1997**). "The representation of auditory source characteristics: Simple geometric form," Percept. Psychophys. **59**, 1180–1190.
Lutfi, R. A. (**2000**). "Source uncertainty, decision weights, and internal noise as factors in auditory identification of a simple resonant source," Assoc. Res. Otolaryngol. Abstr. **23**, 171.
Lutfi, R. A. (**2001**). "Auditory detection of hollowness," J. Acoust. Soc. Am. **110**, 1010–1019.
Lutfi, R. A. (**2008**)."Sound source identification," in *Springer Handbook of Auditory Research: Auditory Perception of Sound Sources*, edited by W. A. Yost and A. N. Popper (Springer-Verlag, New York).
Lutfi, R. A., and Liu, C.-J. (**2007**). "Individual differences in source identification from synthesized impact sounds," J. Acoust. Soc. Am. **122**, 1017–1028.
Lutfi, R. A., and Oh, E. (**1997**). "Auditory discrimination of material changes in a struck-clamped bar," J. Acoust. Soc. Am. **102**, 3647–3656.
Lutfi, R. A., Oh, E., Storm, E., and Alexander, J. M. (**2005**). "Classification and identification of recorded and synthesized impact sounds by practiced listeners, musicians and nonmusicians," J. Acoust. Soc. Am. **118**, 393–404.
Lutfi, R. A., Stoelinga, C. N. J., and Liu, C.-J. (**2008**). "Level dominance in source identification from synthesized impact sounds," J. Acoust. Soc. Am. **124**, 3784–3792.
McAdams, S., Chaigne, A., and Roussarie, V. (**2004**). "The psychomechanics of simulated sound sources: Material properties of impacted bars," J. Acoust. Soc. Am. **115**, 1306–1320.
Micheals, C. F., and Carello, C. (**1981**). *Direct Perception* (Prentice-Hall, Englewood Cliffs, NJ).
Moore, B. C. J., Glasberg, B. R., and Shailer, M. J. (**1984**). "Frequency and intensity difference limens for harmonics within complex tones," J. Acoust. Soc. Am. **75**, 550–561.
Morse, P. M., and Ingard, K. U. (**1968**). *Theoretical Acoustics* (Princeton University Press, Princeton, NJ), pp. 175–191.
Ottaviani, L., and Rocchesso, D. (**2004**). "Auditory perception of 3D size: Experiments with synthetic resonators," ACM Trans. Appl. Percept. **1**, 118–129.
Rocchesso, D. (**2001**). "Acoustic cues for 3-D shape information," Proceedings of the 2001 International Conference on Auditory Display, Finland.
Schlauch, R. S., Reis, D. T., and DiGiovani, J. J. (**2001**). "Duration discrimination and subjective duration for ramped and damped sounds," J. Acoust. Soc. Am. **109**, 2880–2887.
Stoelinga, C. N. J., and Lutfi, R. A. (**2008**). "Spectral density discrimination," Assoc. Res. Otolaryngol. Abstr. **31**, 310.
Tucker, S., and Brown, G. J. (**2003**). "Modelling the auditory perception of size, shape and material: Applications to the classification of transient sonar sounds," presented at the 114th Audio Engineering Society Convention, Amsterdam, The Netherlands, 22–25 March.
Ullman, S. (**1980**). "Against direct perception," Behav. Brain Sci. **3**, 373–415.
Van Heuven, V. J., and Van Den Broecke, M. P. R. (**1979**). "Auditory discrimination of rise and decay times in tone and noise bursts," J. Acoust. Soc. Am. **66**, 1308–1315.
Wier, C. C., Jesteadt, W., and Green, D. M. (**1977**). "Frequency discrimination as a function of frequency and sensation level," J. Acoust. Soc. Am. **61**, 178–184.
Wildes, R., and Richards, W. (**1988**). "Recovering material properties from sound," in *Natural Computation*, edited by W. Richards (MIT, Cambridge, MA), pp. 356–363.
Zener, C. (**1948**). *Elasticity and Anelasticity of Metals* (University of Chicago Press, Chicago, IL).
Zwicker, E., and Fastl, H. (**1999**). *Psychoacoustics: Facts and Models*, 2nd ed. (Springer-Verlag, New York).

The role of suppression in psychophysical tone-on-tone masking[a)]

Joyce Rodríguez,[b)] Stephen T. Neely, Harisadhan Patra, Judy Kopun, Walt Jesteadt, Hongyang Tan, and Michael P. Gorga
Boys Town National Research Hospital, 555 North 30th Street, Omaha, Nebraska 68131

(Received 25 June 2009; revised 25 September 2009; accepted 29 September 2009)

This study tested the hypothesis that suppression contributes to the difference between simultaneous masking (SM) and forward masking (FM). To obtain an alternative estimate of suppression, distortion-product otoacoustic emissions (DPOAEs) were measured in the presence of a suppressor tone. Psychophysical-masking and DPOAE-suppression measurements were made in 22 normal-hearing subjects for a 4000-Hz signal/f_2 and two masker/suppressor frequencies: 2141 and 4281 Hz. Differences between SM and FM at the same masker level were used to provide a psychophysical estimate of suppression. The increase in L_2 to maintain a constant output (L_d) provided a DPOAE estimate of suppression for a range of suppressor levels. The similarity of the psychophysical and DPOAE estimates for the two masker/suppressor frequencies suggests that the difference in amount of masking between SM and FM is at least partially due to suppression.
 [DOI: 10.1121/1.3257224]

PACS number(s): 43.66.Dc, 43.64.Jb, 43.64.Kc [BLM] Pages: 361–369

I. INTRODUCTION

Suppression, defined as the reduction in the response to one stimulus by the simultaneous presentation of another stimulus, has been demonstrated at several different levels in the auditory system. Psychophysical masking occurs when the presence of one sound causes an elevation in the threshold of another sound. Although controversial, it has been hypothesized that suppression contributes to differences between simultaneous masking (SM) and forward masking (FM). The purpose of this study was to investigate the role of suppression in psychophysical tone-on-tone masking.

Suppression was first described by Wever *et al.* (1940) based on measurements of the cochlear microphonic, and later demonstrated in physiological studies of auditory-nerve fibers (ANFs) (e.g., Galambos and Davis, 1944; Sachs and Kiang, 1968; Arthur *et al.*, 1971; Abbas and Sachs, 1976; Javel *et al.*, 1983). Sachs and Kiang (1968) recorded a reduction in firing rate to one tone, usually at a fiber's characteristic frequency (CF), by the addition of a second tone of appropriate frequency and intensity. Suppression has been observed in the responses of outer hair cells (Sellick and Russell, 1979) and in compound action potentials (Dallos *et al.*, 1974; Harris, 1979). Evidence of suppression has also been observed in the vibration patterns of the basilar membrane (BM) (e.g., Rhode, 1977; Ruggero *et al.*, 1992). Both mechanical and neural studies have observed that suppression grows more rapidly for suppressors lower in frequency than CF, compared to suppressors close to CF (e.g., Delgutte, 1990b; Pang and Guinan, 1997; Ruggero *et al.*, 1992).

Evidence of suppression also has been observed in otoacoustic emission (OAE) data (e.g., Brown and Kemp, 1984; Harris *et al.*, 1992; Kummer *et al.*, 1998; Abdala, 1998; Gorga *et al.*, 2003, 2008). Interestingly, psychophysical and OAE data appear to share the same dependence on the relation between signal and suppressor frequency that is evident in both mechanical and neural responses from lower animals.

Excitation, suppression, and adaptation are physiological mechanisms thought to contribute to psychophysical masking (Delgutte, 1990a, 1990b). Studies of single ANFs reveal that excitatory or line-busy masking occurs when the overall discharge rate in the presence of a signal and a masker is not higher than the discharge rate observed when only the masker is present. In contrast, suppression does not produce an increase in discharge rate, but shifts the rate-level function for the signal toward higher intensities, resulting in threshold elevation. This effect is especially evident for off-frequency suppressors where the suppressor is about an octave below the signal. Both line-busy masking and suppression require that the signal and masker be presented simultaneously. Masking due to adaptation occurs when the masker is presented prior to the presentation of the signal. Under these conditions, the response of an ANF to a subsequently presented signal will be reduced compared to the discharge rate the signal would typically elicit when presented alone (e.g., Smith, 1977, 1979; Harris and Dallos, 1979). Adaptation has been suggested as the primary mechanism underlying FM, although temporal integration of masker and signal has been suggested as an alternative (e.g., Delgutte, 1990a, 1990b; Pang and Guinan, 1997; Oxenham, 2001). Delgutte (1990a) suggested that the main difference between SM and FM was the absence of suppression in the FM condition. Depending on stimulus conditions, all of these mechanisms may contribute to the masking of a signal.

[a)]Portions of this work were presented at the 2009 Midwinter Meeting of the Association for Research in Otolaryngology.
[b)]Author to whom correspondence should be addressed. Electronic mail: rodríguezj@boystown.org

The contribution of suppression to masking has been described in psychophysical studies (e.g., Shannon, 1976; Weber and Green, 1978; Duifhuis, 1980; Moore *et al.*, 1984; Bacon *et al.*, 1999), but it is difficult to separate suppression effects from other effects based on psychophysical data alone. In the psychophysical-masking task, an increase in masker level (ML) requires a similar increase in signal level when masker and probe are close in frequency because response growths of the masker and signal are similar. Masking is strongest as the masker frequency (f_m) approximates the signal frequency (f_p). When f_p and f_m are close to one another, it is assumed that their excitation patterns overlap and are being processed by the same compressive nonlinearity; thus, the response to probe and masker grows at about the same rate. For maskers lower in frequency than the signal, masker level must be increased in order for masking to be observed, but once masking threshold is exceeded, masking grows by as much as 2–2.5 dB/dB for low-frequency maskers (e.g., Wegel and Lane, 1924; Egan and Hake, 1950; Stelmachowicz *et al.*, 1987; Plack and Oxenham, 1998). This result is often referred to as upward spread of masking. The contribution of suppression to upward spread of masking is unclear (Oxenham and Plack, 1998; Yasin and Plack, 2005).

Moore *et al.* (1984) suggested that when psychophysical tuning curves (PTCs) are measured using procedures to minimize off-frequency listening, the main factor contributing to the SM-FM difference was suppression. Several earlier psychophysical studies accounted for these differences in terms of suppression as well (e.g., Houtgast, 1972; Shannon, 1976; Vogten, 1978; Wightman *et al.*, 1977), but later studies suggested that suppression was not the only factor involved (e.g., Weber, 1983; Jesteadt and Norton, 1985; Neff, 1986). The extent to which suppression contributes to SM-FM differences remains unresolved.

Distortion-product otoacoustic emissions (DPOAEs) may be observed when two tones (f_2 and f_1, $f_2>f_1$) are presented at the same time. The response (at frequency $f_d=2f_1-f_2$) is not present in the original two-tone stimulus. The response is typically measured at the $2f_1-f_2$ frequency because it is the largest distortion product in humans. It is generally assumed that DPOAEs arise from two places along the BM, a distortion source near the f_2 place and a coherent-reflection source at the DPOAE-frequency place (e.g., Zweig and Shera, 1995; Talmadge *et al.*, 1998; Shera and Guinan, 1999). The distortion component results from the nonlinear interaction between f_1 and f_2; this interaction creates energy at the DPOAE frequency that then travels both apically and basally within the cochlea. The apically traveling energy reaches the $2f_1-f_2$ place on the BM and is then reflected back basally. Brown and Kemp (1984) demonstrated that the introduction of a third, suppressor tone (f_3), in addition to the two-tone probe, could result in the suppression of the DPOAE. By keeping the level of the two-tone probe constant and varying f_3 in frequency and level, growth of DPOAE suppression can be measured. Suppression-growth functions share many similarities with psychoacoustical and physiological measures of cochlear response, and have been studied in normal and hearing-impaired ears (e.g., Abdala, 1998; Martin *et al.*, 1998; Abdala and Fitzgerald, 2003; Gorga *et al.*, 2003). In studies of DPOAE suppression, the f_2, f_1 primary pair is regarded as the "probe," because it is assumed that their interaction near the f_2 place results in the initial generation of the DPOAE, and f_3 is viewed as the equivalent of the "masker" typically used in psychoacoustical studies (e.g., Abdala and Chaterjee, 2003; Gorga *et al.*, 2002, 2008).

Growth of DPOAE suppression follows the same pattern as the growth of masked threshold as a function of masker level, in that the slope is steepest for low-frequency suppressors and shallow for high-frequency suppressors relative to f_2. A slope of nearly 1 is observed for suppressor frequencies near f_2 (e.g., Brown and Kemp, 1984; Harris *et al.*, 1992; Abdala, 1998; Abdala and Chaterjee, 2003; Gorga *et al.*, 2003, 2008). As stated earlier, these general trends are at least qualitatively similar to trends observed in both ANF and BM responses.

It is generally thought that suppression is the mechanism that accounts for changes in DPOAE level as a consequence of the presentation of a third tone. In contrast, SM combines both suppression and excitatory effects (such as line-busy masking). It may be possible, therefore, to use DPOAE-suppression measurements to gain insights into the causes of differences between SM and FM. This study investigated the role of suppression in psychophysical masking by comparing behavioral estimates of suppression (defined as the difference between SM and FM) and DPOAE estimates of suppression in the same group of subjects.

II. METHODS

A. Subjects

Twenty-two subjects participated in this study. They were selected on the basis of hearing sensitivity and production of DPOAEs for a wide range of levels, including low-level stimuli. Subjects ranged in age from 16 to 47 years, with a mean age of 20 years. Each subject had thresholds ≤25 dB hearing loss (HL) (re ANSI, 1996) for standard octave and inter-octave audiometric frequencies from 250 to 8000 Hz. Behavioral thresholds for the purposes of meeting inclusion criteria were measured using routine clinical procedures. Subjects were also required to have thresholds of 20 dB sound pressure level (SPL) or better at the stimulus frequencies of 2141, 4000, and 4281 Hz. A normal 226-Hz tympanogram was also required on each day on which DPOAE measurements were made. Only one ear of each subject was selected for study, and was chosen as the ear with the lowest thresholds at 2141, 4000, and 4281 Hz, and most favorable tympanometric results. If there were no differences between ears of a given subject, the test ear was chosen randomly.

B. Stimuli and apparatus

For the psychophysical experiments, the probe signal (f_p) was set to 4000 Hz. The masker frequencies (f_m) were 2141 and 4281 Hz. The term "off-frequency" will be used to describe conditions in which the masker/suppressor frequency=2141 Hz and "on-frequency" to refer to conditions in which the masker/suppressor frequency=4281 Hz.

During the psychophysical measurements, probe level was held constant at levels ranging from 20 to 45 dB SPL (5-dB steps) for FM and from 20 to 70 dB SPL (10-dB steps) for SM; masker level (L_m) was adaptively varied, with a maximum level of 95 dB SPL. The signal was gated with 5-ms rise and fall Blackman windows, and no steady-state portion. The maskers were 200 ms in duration including rise and fall times of 5 ms (Blackman windows). For the FM condition, the masker-signal delay was 0 ms (measured from the final point of the masker to the initial point of the probe) to minimize recovery from adaptation or maximize masker persistence. For the SM condition, the signal was presented 15 ms before the end of the masker. All stimuli for the psychophysical portion of the study were generated digitally via MATLAB at a sampling rate of 44 100 Hz with 24-bit resolution and output by a soundcard (CardDeluxe, Digital Audio Labs, Minneapolis, MN). The headphone output of the soundcard was fed to a remote passive attenuator in a sound-treated room, and then to a Sennheiser HD 250 Linear II headphone.

DPOAEs were elicited with f_2=4000 Hz (i.e., the same as the signal frequency in the masking measurements) and f_1=3279 Hz. The f_2, f_1 primary pair is viewed as a probe for the DPOAE measurement in much the same way that the signal toneburst is viewed as a probe for the masking measurements. The suppressor frequencies (f_3) were the same as the masker frequencies (2141 and 4281 Hz) in the psychophysical measurements. The higher frequency (4281 Hz) was selected to serve as the on-frequency masker/suppressor for SM, FM, and DPOAE studies because it produced the greatest amount of suppression in a previous study of DPOAE suppression when f_2=4000 Hz (Gorga *et al.*, 2008). Additionally, the growth of suppression is nearly linear when f_2=4000 Hz and f_3=4281 Hz (Gorga *et al.*, 2008), indicating that both frequencies are being processed similarly. Although this observation provides additional support for the use of 4281 Hz as the on-frequency suppressor, it should be noted that it is not on-frequency as defined in the psychophysics literature; as a consequence, the on-frequency slopes obtained for the psychophysical portion of the experiment might differ from previously reported findings. The lower masker/suppressor frequency is an octave below the higher masker/suppressor frequency and was selected to be low enough to be outside the frequency range where compressive response growth is observed, while still being near enough to the signal/f_2 frequency to influence its response. It should be noted that there is ongoing debate regarding compression estimates obtained with a low-frequency masker an octave below f_2 as a linear reference (Lopez-Poveda and Alves-Pinto, 2008; Wojtczak and Oxenham, 2009), but that issue is beyond the scope of this study. Although the low-frequency masker/suppressor may not result in a completely linear response at the signal/f_2 frequency, the influence of this problem is mitigated by using the same low-frequency masker/suppressor for SM, FM, and DPOAE conditions.

For all DPOAE measurements, response waveforms with duration of approximately 250 ms were averaged in two alternating buffers. These buffers were summed and, after a Fourier transformation, the frequency component in the $2f_1-f_2$ bin was used to estimate DPOAE level. These two buffers were subtracted and the squared-magnitudes in the $2f_1-f_2$ bin and the five bins above and below $2f_1-f_2$ were averaged to provide an estimate of noise level.

DPOAE stimuli were produced at a sampling rate of 32 000 Hz by a 24-bit soundcard (CardDeluxe, Digital Audio Labs) that drove a probe-microphone system (Etymotic Research, ER-10C). The "receiver equalization" of the ER-10C was bypassed to allow for the production of high stimulus levels. DPOAE data were collected with custom-designed software (EMAV, Neely and Liu, 1994). Both channels of the soundcard and probe-microphone system were used during DPOAE measurements, with f_2 presented on one channel and f_1 presented on the other. When a suppressor was included, it was presented on the same channel as f_2. For DPOAE measurements, in-the-ear forward-pressure level (FPL) calibration (Scheperle *et al.*, 2008) was used to determine stimulus levels. FPL calibration avoids the influence of standing waves and has been shown to result in less variability in DPOAE measurements (Scheperle *et al.*, 2008).

C. Procedure

For the psychophysical measurements, trials were presented in blocks of 50. Each trial consisted of a 500-ms warning interval, two 300-ms observation intervals separated by 300 ms, and a 300-ms feedback interval following the response of the subject. There was an interval of 500 ms before the beginning of the next trial. Subjects were given visual markers for the warning and observation intervals and correct-interval feedback in a message window on a keypad that they used to indicate their responses. In one interval, the masker and signal were presented together, while in the other interval, the masker was presented alone. A two-down, one-up adaptive tracking procedure was used to estimate the 71%-correct point on the psychometric function (Levitt, 1971). The masker level (L_m) was initially varied with a step size of 4 dB, which was reduced to 2 dB after the first four reversals. The threshold estimate was taken as the mean L_m at the turn points after the first four reversals of each 50-trial block. Trial blocks with standard deviations exceeding 5 dB were not accepted.

For DPOAE measurements, the level of the higher-frequency primary (L_2) was set to one of eight levels (25–60 dB FPL in 5-dB steps). The level of f_1 (L_1) was set to $0.4L_2+39$ dB (Kummer *et al.*, 1998). DPOAE-suppression measurements were obtained by presenting a third (suppressor) tone (f_3) at one of two frequencies (f_3=2141 and 4281 Hz), whose level (L_3) was set to each of 22 levels (−20 to 85 dB FPL, 5-dB steps). A control condition with no suppressor was included before and after all suppressor levels at each f_3. This sequence of suppressor conditions was presented at both suppressor frequencies and all L_2 levels for each subject. Measurements continued until the noise floor averaged down to −25 dB SPL, or until 64 s of artifact-free averaging had taken place, whichever occurred first.

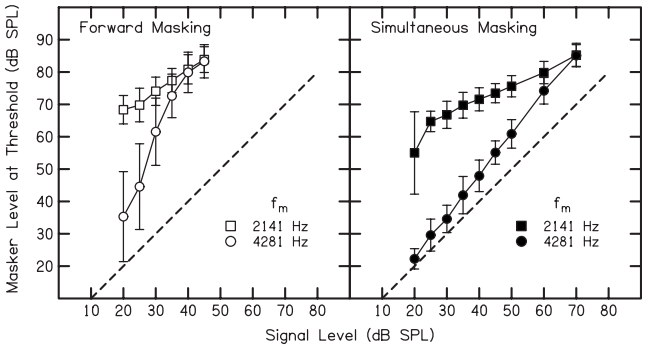


FIG. 1. Mean masker level and standard deviations based on data from all 22 subjects. Each panel shows masker level at threshold as a function of signal level for on- and off-frequency maskers, with results for FM and SM shown in left and right panels, respectively. Off-frequency (2141 Hz) masker levels are represented by open and filled squares and on-frequency (4281 Hz) masker levels are represented by open and filled circles.

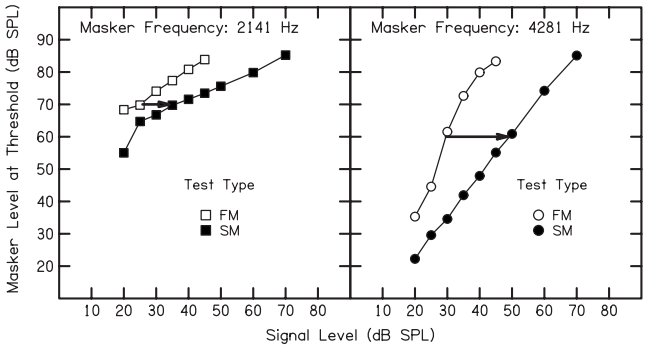


FIG. 2. Mean masker level as a function of signal level based on data from all 22 subjects, with results from on- and off-frequency maskers shown in left and right panels, respectively. Within each panel, open and filled symbols represent FM and SM, respectively. The arrows illustrate the DAM between FM and SM for fixed masker levels.

III. RESULTS

A. Masker-level functions

The mean masker levels (L_m) at threshold as a function of signal level are shown in Fig. 1, with data from FM and SM provided in left and right panels, respectively. These will be referred to as ML functions to distinguish them from growth of masking (GOM) functions, where signal level is plotted as a function of masker level. The parameter within each panel is f_m. Standard deviations for the FM thresholds were 3–5 dB for the off-frequency and 5–6 dB for the on-frequency masker, except at the lowest signal levels (20–30 dB SPL) where the standard deviations for the on-frequency condition ranged from 10 to 13 dB. For the SM conditions, standard deviations were 3–4 dB for both off- and on-frequency maskers, except for the off-frequency masker SM condition at a signal level of 20 dB SPL, where the standard deviation was 12 dB. Although there was variability among subjects, the within-subject standard deviation (based on three repeated measurements) did not exceed 3 dB for any condition.

The ML data in Fig. 1 were fitted with linear least-squares functions to allow for comparisons with previous research (Oxenham and Plack, 1997, 1998; Plack and Oxenham, 1998; Yasin and Plack, 2005). As expected, the FM ML functions were characterized by higher masker thresholds for both on- and off-frequency maskers, with the on-frequency forward masker showing the steepest function. The slopes for the off-frequency and on-frequency FM conditions were 0.6 and 2.0 dB/dB, respectively. The on-frequency ML function for the SM condition grows in a nearly linear fashion with a slope of 1.2 dB/dB; in contrast, the off-frequency ML function for the SM condition exhibits a shift in threshold of about 30 dB for the lowest signal level (20 dB SPL), compared to the same condition for the on-frequency masker, and grows at a slower rate thereafter (0.5 dB/dB). The masking data in Fig. 1 are similar to the data of Oxenham and Plack (1997, Fig. 2). However, our estimates of compression (2:1 for SM and 3:1 for FM) based on taking the ratio of on- and off-frequency GOM slopes (not shown) are not as high as their compression estimate (6:1). The difference in compression estimates might be related to the use of an on-frequency masker that was higher than the signal probe or to off-frequency listening. The use of a higher on-frequency masker (e.g., 4281 Hz) might have changed the slope of the on-frequency functions, resulting in a reduced compression estimate based on the on- and off-frequency ratio comparison. Similarly, given that no additional noise masker was used to restrict listening, the influence of off-frequency listening on the slopes of the functions cannot be ruled out.

B. Estimates of suppression based on comparison of SM and FM

In Fig. 2, the data in Fig. 1 are recast to compare ML functions under FM and SM conditions at each f_m. Because the absolute threshold for the signal is the same in SM and FM conditions, the differences in signal level shown in Fig. 2 can be viewed as differences in amount of masking (DAM). The reason for plotting FM and SM together is to visualize differences in amount of masking between FM and SM at equivalent masker levels. These differences were used as estimates of suppression. This definition is equivalent to the definition of suppression used with rate-level functions in ANF studies (e.g., Javel *et al.*, 1983) and described by Delgutte (1990a) in the context of masking. For f_m=4281 Hz and L_m=60 dB SPL (see arrow in Fig. 2, right panel), the signal levels at threshold were 29 and 48 dB SPL for FM and SM, respectively. The dB difference between these signal levels is 19 dB. A second example, f_m=2141 Hz and L_m=70 dB SPL (see arrow in the left panel of Fig. 2), estimates the amount of suppression (SM-FM difference) as 9.6 dB.

DAM between SM and FM is plotted in Fig. 3. The filled circle and square in Fig. 3 correspond to the 4281 and 2141 Hz conditions illustrated for one L_m by the arrows in Fig. 2. DAM ranged from 5 to 26 dB for masker levels of 30–80 dB SPL for the 4281 Hz masker, and from 8 to 21 dB for masker levels of 60–80 dB SPL when f_m=2141 Hz.

C. DPOAE-suppression data

Measured DPOAE levels (both with and without a suppressor tone present) are shown in Fig. 4 as input/output (I/O) functions. DPOAE level (L_d) is plotted as a function of

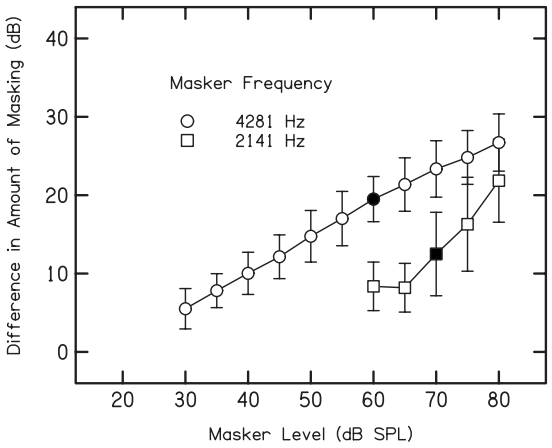


FIG. 3. The amount of suppression, estimated as the DAM for SM and FM as a function of masker level; the parameter is masker frequency. The open circles and squares represent the results for on-frequency (4281 Hz) and off-frequency maskers (2141 Hz), respectively. The filled circle and square represent the SM-FM difference illustrated by the horizontal arrows in Fig. 2.

L_2, with L_3 as the parameter. Control conditions (without a suppressor) are shown as filled symbols in Fig. 4. The I/O functions shift toward the right in the presence of suppressors. We define amount of suppression relative to L_2 by fixing L_d and determining how much L_2 must increase to maintain the same output (L_d) in the presence of suppressors. This measure of suppression is not independent of DPOAE level. Because the present study was exploratory in nature, it was necessary to empirically determine the L_d criterion output with and without suppressors that could be used to determine the L_2 levels at which this constant output (L_d) was observed. Although it was apparent from visual inspection of DPOAE I/O functions that L_d=−3 dB SPL was the output criterion that best agreed with the DAM results, additional analyses were performed to support this choice. The mean data were analyzed in 0.25 dB steps. The amount of suppression, calculated as the shift in L_2 as a function of L_3 (i.e., suppressor

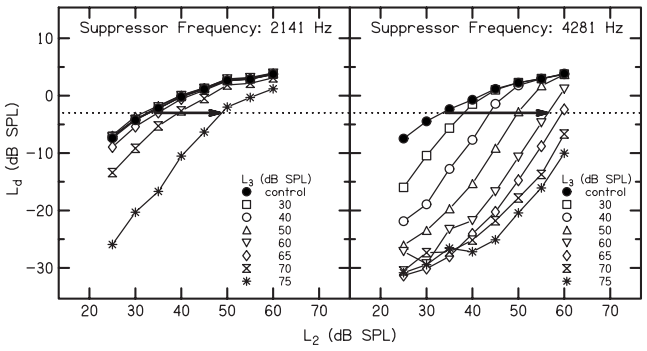


FIG. 4. Mean DPOAE level (L_d) as a function of L_2 (f_2=4000 Hz), with suppressor level (L_3) as the parameter. The filled symbols within each panel represent the DPOAE levels for control conditions, in which no suppressor was presented. The left and right panels show data for off-frequency (2141 Hz) and on-frequency (4281 Hz) suppressors, respectively. The horizontal arrows in each panel are drawn at an L_d of −3 dB SPL. They illustrate the extent to which L_2 had to be increased in order to maintain a DPOAE level of −3 dB SPL as L_3 increases from the control condition. The standard deviations (4–9 dB) are consistent with previously reported DPOAE variability.

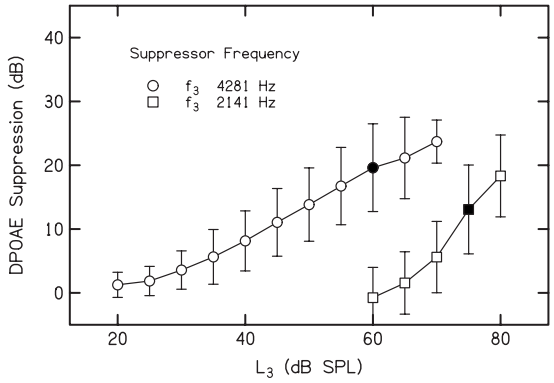


FIG. 5. Amount of DPOAE suppression in dB as a function of suppressor level. The open circles and squares represent the conditions in which f_3 =4281 Hz and f_3=2141 Hz, respectively. The filled circle and square correspond to the L_2 difference described by the arrows in Fig. 4.

level), varied by only about 2 dB for L_d between 0 and −5 dB SPL. Thus, the initial selection of L_d=−3 dB SPL provided the least deviation from the DAM data, while spanning the greatest number of DPOAE I/O functions. While the selection of −3 dB SPL for our definition of DPOAE suppression is somewhat arbitrary, it was sufficient for the objectives of this study to find agreement between DAM and *any* definition of DPOAE suppression.

Estimates of suppression for the DPOAE functions are plotted in Fig. 5 and ranged from 1 to 23 dB for L_3 =20–70 dB SPL for the on-frequency suppressor. The off-frequency suppressor produced suppression estimates that ranged from 0 to 18 dB for L_3=60–80 dB SPL. The filled circle and square in Fig. 5 correspond to the 4281 and 2141 Hz conditions indicated by the arrows in Fig. 4.

D. Comparison of psychophysical and physiological estimates

Figure 6 compares DAM for the on- and off-frequency maskers to the DPOAE-suppression estimates for the same frequencies. The differences between the mean DAM and DPOAE suppression were less than 3 dB for all conditions in which comparisons could be made for the on-frequency masker. For the off-frequency masker, estimates of DAM and DPOAE suppression exhibited differences of 3–9 dB for

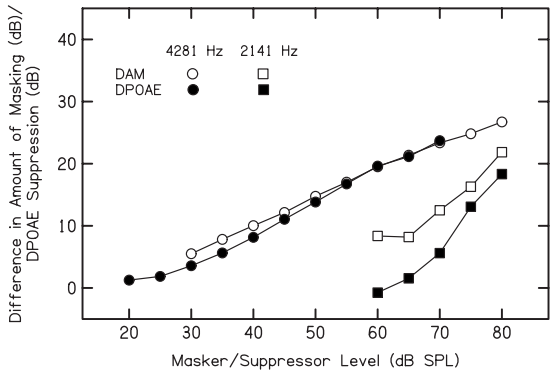


FIG. 6. Mean DAM and amount of DPOAE suppression as a function of masker/suppressor level for all subjects. Open circles and squares represent DAM for 4281 and 2141 Hz, respectively. Filled circles and squares represent DPOAE suppression for 4281 and 2141 Hz, respectively.

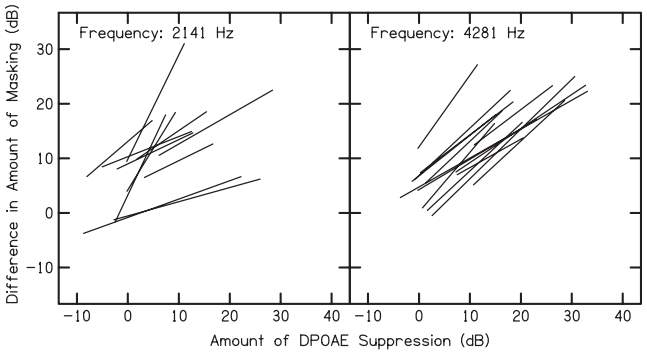


FIG. 7. Relation between DAM and DPOAE suppression represented by best-line fits to the data from individual subjects. Best-line fits for the off-frequency (2141 Hz) and on-frequency (4281 Hz) masker/suppressors are depicted in the left and right panels, respectively.

masker/suppressor levels up to 80 dB SPL. Even so, the general shapes of the DAM and DPOAE-suppression functions were similar, suggesting that suppression is at least partially responsible for the difference between SM and FM.

Figure 7 plots DAM as a function of the amount of DPOAE suppression for individual subjects, where points derived at three or more masker levels are approximated by the best-fitting straight lines shown in the figure. To evaluate this relation, pairs of values for DAM and DPOAE suppression were analyzed for all masker/suppressor levels at which they were available for individual subjects. For example, if at 60 dB SPL, the DAM was 20 dB and the DPOAE suppression was 22 dB, those two values constituted a pair (i.e., the coordinates for a data point) and were included in the analysis. If DAM and DPOAE suppression were measured at masker/suppressor levels of 50, 60, and 70 dB SPL, those three pairs were used to obtain a best-fit line. A minimum of three pairs per subject was required in order to fit a line. The relation between DAM and amount of DPOAE suppression varied across subjects for both masker/suppressor frequencies. The trends across subjects were at least qualitatively similar for the on-frequency case; greater variability was evident, however, for the off-frequency masker/suppressor condition.

Table I contains the individual results for the DAM/DPOAE-suppression correlations. As stated above, results are included only for subjects for whom three or more DAM/DPOAE-suppression pairs were available. As can be seen in Table I, three points were not available from all subjects and (not surprisingly) the number of available points for fitting a line was about twice as large for the on-frequency case, compared to the number of available points for the off-frequency case. Still, the slopes for both masker frequencies have positive correlation, and the correlation for DAM/DPOAE-suppression functions was significant at the 0.05 level, suggesting that a relation exists between these two different measures of suppression. The wide range of regression intercepts demonstrates that the DAM/DPOAE-suppression relation is variable across individuals.

The individual results shown in Fig. 7 were averaged across subjects to produce the mean data shown in Fig. 8. The mean data indicate that DAM grows at about two-thirds the rate of DPOAE suppression. The correlation between DAM and DPOAE suppression was measured separately for the on- and off-frequency conditions. Because many of the correlations were high, the individual correlations were converted to Fisher z values prior to averaging, then converted back to r values. At both frequencies, the mean correlation was positive and significant ($2141: r=0.96$, $p<0.05$; $4281: r=0.98$, $p<0.05$). These results indicate that the relationship between DAM and DPOAE suppression is independent of masker/suppressor frequency, at least for the two frequencies used in the present study.

IV. DISCUSSION

In general, the present masking results are in agreement with previously reported findings. Bacon *et al.* (1999) examined GOM using a SM task for signal frequencies between

TABLE I. Number of paired comparisons (N), slope, intercept, and correlation coefficients of the DAM/DPOAE-suppression function for individual subjects for each masker/suppressor.

	2141 Hz				4281 Hz			
Subject	Slope	Intercept	r	N	Slope	Intercept	r	N
01	0.36	10.3	0.87	3	0.74	6.7	0.99[a]	6
02	1.97	3.2	0.85[a]	7	0.55	4.2	0.91[a]	6
03	0.80	13.0	0.96	3	1.09	0.1	0.98[a]	6
04	0.63	8.7	0.88	3	0.52	3.1	0.99[a]	4
05					0.70	4.7	0.99[a]	5
06	0.26	−0.5	0.90[a]	8	0.53	4.7	0.94[a]	8
07	0.53	7.2	1.00[a]	3	0.71	7.1	0.97[a]	8
08	0.45	5.0	0.99[a]	3	0.62	2.9	0.99[a]	7
09					0.87	4.3	0.96[a]	6
10	1.91	9.8	0.88	4	1.31	12.1	0.94[a]	7
11	1.52	4.2	0.89[a]	6	0.91	−2.8	0.97[a]	8
12	0.33	−0.8	0.89[a]	5	0.86	−0.9	0.98[a]	8
13	0.44	8.9	0.65	3	0.87	6.8	0.94[a]	7
14					0.86	−4.1	0.99[a]	6

[a]$p<0.05$.

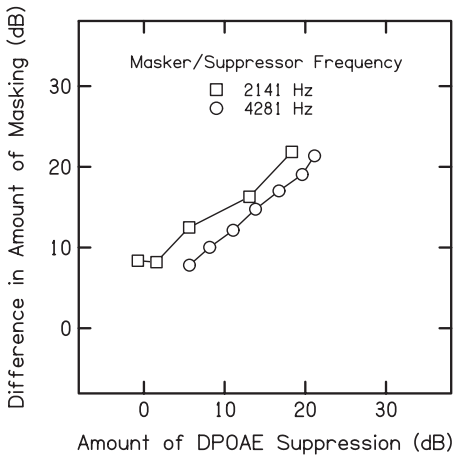


FIG. 8. Relation between DAM and DPOAE estimates of suppression. Open squares and circles represent mean values of paired DAM/suppression comparisons across subjects for 2141 and 4281 Hz, respectively.

400 and 5000 Hz in the presence of maskers about a half octave below (f_m/f_p=0.7) and ranging in level from 40 to 95 dB SPL; stimulus duration was 100 ms for f_m and 5 ms for f_p. For f_p=4300 Hz and f_m=3000 Hz, they found an average slope of 1.9 (dB/dB) for levels similar to ours (60–90 dB SPL). Bacon *et al.* (1999) concluded that for signal frequencies at or above 750 Hz, the slope of the growth of the SM function changed from a value greater than 1 (average ≈1.9) for 60–80 dB SPL signals to close to 1 at levels ≥80 dB SPL. Other studies of GOM using either a FM or FM-SM paradigm (Oxenham and Plack, 1997; Plack and Oxenham, 1998; Yasin and Plack, 2005), with f_p=4000 Hz, f_m=2400 and 4000 Hz, and stimulus parameters similar to ours, have reported that, on average, the on-frequency FM slopes were 2.4 dB/dB, a value greater than the on-frequency SM slopes when plotted as functions of signal level. Our on-frequency ML slope for the FM condition was about 2 dB/dB, which is greater than the ML slope for the SM condition (e.g., 1.2 dB/dB).

Previous research (e.g., Moore *et al.*, 1984, Delgutte, 1990a, 1990b) has suggested that the difference between SM and FM might be due to suppression; however, this hypothesis had not been directly tested against an alternative measure of suppression. The goal of this experiment was to test this hypothesis by comparing a behavioral estimate of suppression (DAM) to measures of DPOAE suppression in the same group of subjects. A primary concern of this study was to develop a set of experimental conditions that would allow comparisons between psychophysical and DPOAE data with the hope of gaining insights into the role of suppression in SM. In our previous DPOAE-suppression studies, we have kept primary levels constant while varying suppressor level. We designed the masking measurements for this study to parallel the DPOAE measurements by keeping signal level constant and varying masker level. Unfortunately, we found it difficult to compare results at constant signal levels because DPOAE generation and psychophysical masking are both only indirectly linked to the cochlear response to the probe stimulus. Equating masker levels with DPOAE suppressor levels made more sense on the assumption that the role of masker and suppressor are more similar between the two measurements. In other words, it made more sense to quantify suppression in terms of *input* levels because the *output* measures were not directly comparable. In retrospect, it would have been better to obtain masking measurements at fixed masker levels to avoid the need to interpolate between masker levels. The decision to hold suppressor level constant had less impact on the DPOAE measurement because we collected these data with suppressor levels varied in 5-dB steps.

It is worth noting that f_1, in addition to interacting with f_2 to generate distortion, also suppresses the response to f_2 (Geisler, 1998) at the "optimum" levels for DPOAE generation; the combination of f_1 with f_2 causes enough suppression that when the level of f_2 is at the subject's audiometric threshold for f_2 in quiet, the subject is unable to detect the presence of f_2 in the stimulus. Thus, L_2 must be above the quiet threshold level for f_2 to be heard by the subject. Based on previous, unpublished measurements in our laboratory and our interpretation of the data described in this paper, we suspect that the threshold of audibility for f_2 (4000 Hz) (in the presence of f_1 at levels used to elicit a DPOAE) occurs at L_2 levels that produce L_d between −5 and 0 dB SPL. We could have chosen to define suppression at any L_d between −5 and 0 dB SPL since the same trend was observed across this range. However, the amount of suppression when L_d= −3 dB SPL was most similar to the SM-FM masking difference.

The data for SM in Fig. 2 show that, in agreement with previous studies, the slope of the off-frequency ML function is less than unity, while the slope of the on-frequency ML function is close to unity (Oxenham and Plack, 1997). Near-linear growth of SM for an on-frequency masker is expected, even though the response of the basilar membrane at a specific place is nonlinear for best-frequency tones (when cochlear function is normal), because both the masker and the signal are being processed through the same nonlinearity (Oxenham and Plack, 1997; Plack and Oxenham, 1998). Similarly, a near-linear growth in suppression (defined as the shift in L_2 necessary to maintain an L_d of −3 dB SPL as a function of L_3) was observed when $f_3 \approx f_2$ (Fig. 3), a pattern that has been observed previously (e.g., Abdala, 1998; Martin *et al.*, 1998; Gorga *et al.*, 2008). The psychophysical off-frequency masker produced a more rapid growth of masking, which also is consistent with findings previously reported in the literature (e.g., Oxenham and Plack, 1997, 1998; Plack and Oxenham, 1998; Stelmachowicz *et al.*, 1987).

In agreement with physiological data (Pang and Guinan, 1997), our off-frequency data, when f_p is plotted as a function of L_m, reveal that suppression threshold is about 55–65 dB SPL and the SM slope is 2 dB/dB. Pang and Guinan (1997), in a study of GOM in ANFs of the cat using a SM condition, found that the threshold of masking for an off-frequency masker was typically higher than 60 dB SPL and the slope of the GOM function was 2 dB/dB or higher. Our 2 dB/dB SM slope (L_p as a function of L_m) is also close to the 2.4 dB/dB slope suggested by the Allen and Sen (1998) model for the slope of an off-frequency masker.

The similarity of the psychophysical and DPOAE estimates shown in Fig. 6, to a first approximation, suggests that DAM is, at least in part, due to suppression. The correlations between these two measures of suppression shown in Table I support this view. It is likely that the absolute differences in amount of suppression between these two measures would change, depending on stimulus conditions. However, for the conditions chosen for the present study (chosen to minimize recovery from adaptation during FM), the mean difference in suppression estimates between the two paradigms is less than 3 dB at any level for the on-frequency condition, with slightly larger differences (3–9 dB) for the off-frequency case.

The individual best-line fits in Fig. 7 and their correlation coefficients in Table I provide more details concerning the relation between DAM and DPOAE suppression. This figure reveals an orderly relationship for the on-frequency condition, whereas the off-frequency case shows more scatter. As seen in Table I, the individual correlations of the DAM/DPOAE-suppression function for both masker/suppressors are ≥ 0.8, except for one condition. These data suggest that for a given amount of suppression, DAM may be expected to be two-thirds the suppression value for parameters similar to those used in this study. At present, we do not have an explanation for the rate of growth of DAM versus DPOAE suppression, although one possible explanation might relate to the criterion output level ($L_d = -3$ dB SPL) selected for data analysis of the DPOAE I/O functions. It is also possible that if we had used a psychophysical paradigm in which L_p was varied while L_m was held constant, we might have observed results more in alignment with the suppression function; at the very least, such a paradigm would have avoided the data transformations that were needed in the present study to compare the DPOAE and behavioral-masking data.

Our estimates of suppression (3–26 dB) are consistent with other estimates reported in the psychophysical literature. Specifically, other psychophysical studies estimating suppression as a change in signal threshold during a pulsation-threshold task (Duifhuis, 1980; Shannon, 1986) or GOM measurements (Oxenham and Plack, 1998; Yasin and Plack, 2005) have reported suppression values of 6–33 dB. Duifhuis (1980), in a study of psychophysical two-tone suppression, observed suppression estimates of about 16 dB for $f_p = 1000$ Hz and $f_m = 600$ Hz for L_m between 60 and 90 dB SPL using a pulsation-threshold paradigm. Shannon (1986) also used a pulsation-threshold task, and estimated about 20 dB of suppression for $f_p = 1000$ Hz and $f_m = 400$ Hz. Other psychophysical studies have estimated suppression as a function of signal level by subtracting interpolated masker levels for SM from the FM condition for a 2400 Hz off-frequency masker when $f_p = 4000$ Hz (Oxenham and Plack, 1998; Yasin and Plack, 2005). In one case, estimates of suppression across subjects ranged from 15 to 32 dB for signal levels from 40 to 60 dB SPL (Oxenham and Plack, 1998), and, in the other case, it ranged from 6 to 17 dB for signal levels of 10–80 dB SPL (Yasin and Plack, 2005). These ranges are similar to the range observed in the present study (8–21 dB) for the off-frequency case. Using a measure of neural synchrony with $f_p = 1000$ Hz and $f_m = 400$ Hz, Javel (1981) reported 28 dB of suppression for the shift in neural-response functions in the presence of an off-frequency masker.

Taken together, previous data, as summarized above, suggest that the amount of suppression exhibits variability that is dependent on the relationship between signal and masker frequency, level, and test paradigm. Still, the similarity of values across studies suggests that suppression does not appear to exceed 35 dB for levels between 30 and 90 dB SPL. Thus, the values estimated for DAM in the present study are in agreement with the previously reported data for psychophysical and physiological suppression estimates.

V. CONCLUSIONS

The goal of this experiment was to examine the extent to which differences between SM and FM can be attributed to suppression that occurs during SM, but not during FM. The results for most of the signal range under study provide evidence to support the view that the difference between SM and FM (DAM) is at least partly due to suppression.

ACKNOWLEDGMENTS

This work was supported by the NIH (Grant Nos. NIDCD R01 DC002251, R01 DC006350, T32 DC000013, and P30 004662). We thank Sarah Michael and Tom Creutz for their help with instrumentation, Sandy Estee for her assistance in subject recruitment, and the subjects who participated in the study.

Abbas, P. J., and Sachs, M. B. (**1976**). "Two-tone suppression in auditory-nerve fibers: Extension of a stimulus-response relationship," J. Acoust. Soc. Am. **59**, 112–122.

Abdala, C. (**1998**). "A developmental study of distortion product otoacoustic emission ($2f_1-f_2$) suppression in humans," Hear. Res. **121**, 125–138.

Abdala, C., and Chaterjee, M. (**2003**). "Maturation of cochlear nonlinearity as measured by DPOAE suppression growth in humans," J. Acoust. Soc. Am. **114**, 932–943.

Abdala, C., and Fitzgerald, T. (**2003**). "Ipsilateral distortion product otoacoustic emission ($2f_1-f_2$) suppression in children with sensorineural hearing loss," J. Acoust. Soc. Am. **114**, 919–931.

Allen, J. B., and Sen, D. (**1998**). "A unified theory of two-tone suppression and the upward spread of masking," in Proceedings of the 16th International Congress in Acoustics and 135th Meeting of the Acoustical Society of America.

American National Standards Institute. (**1996**). "Specifications for audiometers," ANSI S3.6-1996, New York.

Arthur, R. M., Pfeiffer, R. R., and Suga, N. (**1971**). "Properties of two-tone inhibition in primary auditory neurons," J. Physiol. **212**, 593–609.

Bacon, S. P., Boden, L. N., Lee, J., and Repovsch, J. (**1999**). "Growth of simultaneous masking for fm<fs: Effects of overall frequency and level," J. Acoust. Soc. Am. **106**, 341–350.

Brown, A. M., and Kemp, D. T. (**1984**). "Suppressibility of the $2f_1-f_2$ stimulated acoustic emissions in gerbil and man," Hear. Res. **13**, 29–37.

Dallos, P., Cheatham, M. A., and Ferraro, J. (**1974**). "Cochlear mechanisms, nonlinearities, and cochlear potentials," J. Acoust. Soc. Am. **55**, 597–605.

Delgutte, B. (**1990a**). "Physiological mechanisms of psychophysical masking: Observations from auditory-fibers," J. Acoust. Soc. Am. **87**, 791–809.

Delgutte, B. (**1990b**). "Two-tone suppression in auditory-nerve fibers: Dependence on suppressor frequency and level," Hear. Res. **49**, 225–246.

Duifhuis, H. (**1980**). "Level effects in psychophysical two-tone suppression," J. Acoust. Soc. Am. **67**, 914–927.

Egan, J. P., and Hake, H. W. (**1950**). "On the masking pattern of a simple auditory stimulus," J. Acoust. Soc. Am. **22**, 622–630.

Galambos, R., and Davis, H. (**1944**). "Inhibition of activity in single auditory nerve fibers by acoustic stimulation," J. Neurophysiol. **7**, 287–303.

Geisler, C. D. (**1998**). *From Sound to Synapse: Physiology of the Mammalian Ear* (Oxford University Press, New York).
Gorga, M. P., Neely, S. T., Dierking, D. M., Dorn, P. A., Hoover, B. M., and Fitzpatrick, D. F. (**2003**). "Distortion product otoacoustic emission suppression tuning curves in normal-hearing and hearing-impaired human ears," J. Acoust. Soc. Am. **114**, 263–278.
Gorga, M. P., Neely, S. T., Dierking, D. M., Kopun, J., Jolkowski, K., Groenenboom, K., Tan, H., and Stiegemann, B. (**2008**). "Low-frequency and high-frequency distortion product otoacoustic emission suppression in humans," J. Acoust. Soc. Am. **123**, 2172–2190.
Gorga, M. P., Neely, S. T., Dorn, P. A., Dierking, D. M., and Cyr, F. (**2002**). "Evidence of upward spread of suppression in DPOAE measurements," J. Acoust. Soc. Am. **112**, 2910–2920.
Harris, D. M. (**1979**). "Action potential suppression, tuning curves and thresholds: comparison with single nerve fiber," Hear. Res. **1**, 133–154.
Harris, D. M., and Dallos, P. (**1979**). "Forward masking of auditory nerve fibers responses," J. Neurophysiol. **42**, 1083–1107.
Harris, F. P., Probst, R., and Xu, L. (**1992**). "Suppression of the $2f_1-f_2$ otoacoustic emission in humans," Hear. Res. **64**, 133–141.
Houtgast, T. (**1972**). "Psychophysical evidence for lateral inhibition in hearing," J. Acoust. Soc. Am. **51**, 1885–1894.
Javel, E. (**1981**). "Suppression of auditory nerve responses I: Temporal analysis, intensity effects and suppression contours," J. Acoust. Soc. Am. **69**, 1735–1745.
Javel, E., McGee, J., Walsh, E., and Gorga, M. P. (**1983**). "Suppression of auditory nerve responses. II. Suppression threshold and growth, iso-suppression contours," J. Acoust. Soc. Am. **74**, 801–813.
Jesteadt, W., and Norton, S. J. (**1985**). "The role of suppression in psychophysical measures of frequency selectivity," J. Acoust. Soc. Am. **78**, 365–374.
Kummer, P., Janssen, T., and Arnold, W. (**1998**). "The level and growth behavior of the $2f_1-f_2$ distortion product otoacoustic emission and its relationship to auditory sensitivity in normal and cochlear hearing loss," J. Acoust. Soc. Am. **103**, 3431–3444.
Levitt, H. (**1971**). "Transformed up-down methods in psychoacoustics," J. Acoust. Soc. Am. **49**, 467–477.
Lopez-Poveda, E. A., and Alves-Pinto, A. (**2008**). "A variant temporal-masking-curve method for inferring peripheral auditory compression," J. Acoust. Soc. Am. **123**, 1544–1554.
Martin, G. K., Jassir, D., Stagner, B. B., Whitehead, M. L., and Lonsbury-Martin, B. L. (**1998**). "Locus of generation for the $2f_1-f_2$ vs. $2f_2-f_1$ distortion-product otoacoustic emissions in normal-hearing humans revealed by suppression tuning, onset latencies, and amplitude correlations," J. Acoust. Soc. Am. **103**, 1957–1971.
Moore, B. C. J., Glasberg, B. R., and Roberts, B. (**1984**). "Refining the measurement of psychophysical tuning curves," J. Acoust. Soc. Am. **76**, 1057–1066.
Neely, S. T., and Liu, Z. (**1994**). "EMAV: Otoacoustic emission averager," Technical Memo No. 17, Boys Town National Research Hospital, Omaha, NE.
Neff, D. L. (**1986**). "Confusion effects with sinusoidal and narrow-band noise forward maskers," J. Acoust. Soc. Am. **79**, 1519–1529.
Oxenham, A. J. (**2001**). "Forward masking: Adaptation or integration?," J. Acoust. Soc. Am. **109**, 732–741.
Oxenham, A. J., and Plack, C. J. (**1997**). "A behavioral measure of basilar-membrane nonlinearity in listeners with normal and impaired hearing," J. Acoust. Soc. Am. **101**, 3666–3675.
Oxenham, A. J., and Plack, C. J. (**1998**). "Suppression and the upward spread of masking," J. Acoust. Soc. Am. **104**, 3500–3510.
Pang, X. D., and Guinan, J. J. (**1997**). "Growth rate of simultaneous masking in cat auditory-nerve fibers: Relationship to the growth of basilar-membrane motion and the origin of two-tone suppression," J. Acoust. Soc. Am. **102**, 3564–3575.
Plack, C. J., and Oxenham, A. J. (**1998**). "Basilar membrane nonlinearity and the growth of forward masking," J. Acoust. Soc. Am. **103**, 1598–1608.
Rhode, W. S. (**1977**). "Some observations of two-tone interactions measured using the Mössbauer technique," in *Psychophysics and Physiology of Hearing*, edited by E. F. Evans and J. P. Wilson (Academic, London), pp. 27–41.
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.
Scheperle, R. A., Neely, S. T., Kopun, J. G., and Gorga, M. P. (**2008**). "Influence of *in situ*, sound-level calibration on distortion-product otoacoustic emission variability," J. Acoust. Soc. Am. **124**, 288–300.
Sellick, P. M., and Russell, I. J. (**1979**). "Two-tone suppression in cochlear hair cells," Hear. Res. **1**, 227–236.
Shannon, R. V. (**1976**). "Two-tone unmasking and suppression in a forward masking situation," J. Acoust. Soc. Am. **59**, 1460–1470.
Shannon, R. V. (**1986**). "Psychophysical suppression of selective portions of pulsation threshold patterns," Hear. Res. **21**, 257–260.
Shera, C., and Guinan, J. J. (**1999**). "Evoked otoacoustic emissions arise by two fundamentally different mechanisms: A taxonomy for mammalian OAEs," J. Acoust. Soc. Am. **105**, 782–798.
Smith, R. L. (**1977**). "Short-term adaptation in single auditory nerve fibers: Some post-stimulatory effects," J. Neurophysiol. **40**, 1098–1112.
Smith, R. L. (**1979**). "Adaptation, saturation and physiological masking in single auditory nerve fibers," J. Acoust. Soc. Am. **96**, 795–800.
Stelmachowicz, P. G., Lewis, D. E., Larson, L. L., and Jesteadt, W. (**1987**). "Growth of masking as a measure of response growth in hearing-impaired listeners," J. Acoust. Soc. Am. **81**, 1881–1887.
Talmadge, C. L., Long, G. R., and Piskorski, P. (**1998**). "Modeling otoacoustic emission and hearing threshold fine structures," J. Acoust. Soc. Am. **104**, 1517–1543.
Vogten, L. L. M. (**1978**). "Low-level pure-tone masking: A comparison of tuning curves obtained with simultaneous and forward masking," J. Acoust. Soc. Am. **63**, 1520–1527.
Weber, D. L. (**1983**). "Do off-frequency simultaneous maskers suppress the signal?," J. Acoust. Soc. Am. **73**, 887–893.
Weber, D. L., and Green, D. M. (**1978**). "Temporal factors and suppression in backward and forward masking," J. Acoust. Soc. Am. **66**, 396–399.
Wegel, R. L., and Lane, C. E. (**1924**). "The auditory masking of one pure tone by another and its probable relation to the dynamics of the inner ear," Phys. Rev. **23**, 266–285.
Wever, E. G., Bray, C. W., and Lawrence, M. (**1940**). "The interference of tones in the cochlea," J. Acoust. Soc. Am. **12**, 268–280.
Wightman, F. L., McGee, T., and Kramer, M. (**1977**). "Factors influencing frequency selectivity in normal and hearing impaired listeners," in *Psychophysics and Physiology of Hearing*, edited by E. F. Evans and J. P. Wilson (Academic, London), pp. 295–308.
Wojtczak, M., and Oxenham, A. J. (**2009**). "Pitfalls in behavioral estimates of basilar-membrane compression in humans," J. Acoust. Soc. Am. **125**, 270–281.
Yasin1, I., and Plack, C. J. (**2005**). "The role of suppression in the upward spread of masking," J. Assoc. Res. Otolaryngol. **6**, 368–377.
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.

Computation of the head-related transfer function via the fast multipole accelerated boundary element method and its spherical harmonic representation

Nail A. Gumerov,[a] Adam E. O'Donovan, Ramani Duraiswami, and Dmitry N. Zotkin
Perceptual Interfaces and Reality Laboratory, Institute for Advanced Computer Studies, University of Maryland, College Park, Maryland 20742

(Received 18 June 2009; revised 6 October 2009; accepted 7 October 2009)

The head-related transfer function (HRTF) is computed using the fast multipole accelerated boundary element method. For efficiency, the HRTF is computed using the reciprocity principle by placing a source at the ear and computing its field. Analysis is presented to modify the boundary value problem accordingly. To compute the HRTF corresponding to different ranges via a single computation, a compact and accurate representation of the HRTF, termed the *spherical spectrum*, is developed. Computations are reduced to a two stage process, the computation of the spherical spectrum and a subsequent evaluation of the HRTF. This representation allows easy interpolation and range extrapolation of HRTFs. HRTF computations are performed for the range of audible frequencies up to 20 kHz for several models including a sphere, human head models [the Neumann KU-100 ("Fritz") and the Knowles KEMAR ("Kemar") manikins], and head-and-torso model (the Kemar manikin). Comparisons between the different cases are provided. Comparisons with the computational data of other authors and available experimental data are conducted and show satisfactory agreement for the frequencies for which reliable experimental data are available. Results show that, given a good mesh, it is feasible to compute the HRTF over the full audible range on a regular personal computer. © *2010 Acoustical Society of America.* [DOI: 10.1121/1.3257598]

PACS number(s): 43.66.Pn, 43.20.Fn [NX] Pages: 370–386

I. INTRODUCTION

Scattering of sound off the bodies of human and animal listeners, especially their heads and external ears, provides them with cues that enable localization of the sound source.[1,2] The scattering process can be modeled as a linear filter applied to the sound emanating from the source. The "head related transfer function (HRTF)" captures the scattering behavior of the ear, head, and body of the listener. If the head is centered at a point P and the sound source located at elevation θ, azimuth φ, and distance r in a head-centered spherical coordinate system, then the HRTF $H(r,\theta,\varphi,f)$ is the ratio of the Fourier transform of the signal at the ear $F_e(f)$ to the Fourier transform of the signal that would have been received at the point P in free-field $F_P(f)$, where f is the signal frequency. Roughly speaking, the size of the head is important above 1 kHz, the general characteristics of the torso are important below 3 kHz, and the detailed structure of the head and pinnae becomes significant above 3 kHz, with the details of the pinnae itself becoming important at frequencies over 7 kHz.

Often the HRTF is measured. Both direct and reciprocal HRTF measurements have been reported. In the former method a source (a speaker) is moved to various locations and the response measured at a microphone placed at the entrance to the blocked ear canal (see, e.g., Refs. 3 and 4). In the latter, an outwardly directed source (a microspeaker) is placed at the entrance to the blocked ear canal, and the HRTF determined using Helmholtz' reciprocity principle by simultaneously measuring the received signal at microphones placed at locations of interest.[5]

Since the HRTF results from a scattering process, many computational acousticians have attempted to compute it by solving the wave equation subject to boundary conditions on the surface of the head.[6–10] Another area of interest has been the numerical simulation of the HRTF of various animals, e.g., the dolphin[11] and the bat.[12] The numerical HRTF is usually obtained by solving the Fourier transform of the wave equation (the Helmholtz equation) at wavenumbers corresponding to frequencies of interest, though direct simulation of the wave equation in the time domain has also been attempted.[13] When the solution is obtained using the direct experimental configuration, the simulations must be repeated for each source position.

Numerical simulation is attractive since it offers the possibility of extracting HRTFs without subjecting the user to measurement (beyond those needed to create the discretization of their body). Further, numerical simulation, if it were easy and accurate, offers the promise of allowing one to relate features in the HRTF with the anatomical structure of the head, pinna, and body. By manipulating the mesh and observing the resulting computed HRTF it might be possible to explore the sensitivity of the HRTF to particular features. However, the match between experiment and simulation over the entire auditory range has not yet been presented, and such sensitivity analysis still remains to be done. Moreover,

[a]Author to whom correspondence should be addressed. Electronic mail: gumerov@umiacs.umd.edu

TABLE I. A list of recent work on numerical HRTF computation describing the numerical method used, whether a direct or reciprocal formulation was employed, the maximum frequency employed, the size of the mesh, the time taken, and whether the computation results are compared with experiments. Data in italics refer to projections rather than actual computations. Reference 9 uses a different numerical method. None of the papers referred to include the torso in the computations.

Ref.	Method	Dir/Rec	f_{max} (kHz)	Mesh	Time	Expt.
7	BEM	Recip.	5.4	22 000	28 h/freq	Y
6	BEM	Recip.	10.3	30 570	9.15 h/freq	Y
13	FDTD-PML	Direct	8	$[130]^3$	0.66 h/direction	N
9	BEM	***	20	28 000	480 h	N
14	BEM	Direct	15	21 693	*52.5 h*	Y
This paper	FMBEM	Recip.	20	445 276	70 h	Y

computations are relatively time consuming (taking days of computer time), and simulations cannot yet be used for parametric study.

As can be seen from Table I, progress on simulations has been gradual, and the ability to completely synthesize HRTFs from meshes and simulations is not yet available. This paper presents a first such computation. Progress has been hindered by several factors. First, simulations are relatively time consuming. Since the wavelengths that must be resolved are relatively small, and approximately six to ten elements are needed per wavelength, large meshes result. When the boundary element method (BEM) is solved directly, the solution time scales as the cube of the discretization size leading to long simulation times, making numerical simulation unattractive in comparison with experiments. Further, the amount of memory needed to achieve a simulation at higher frequencies means that these simulations can only be done on supercomputers, and this has meant that most studies do not perform the simulations for high frequencies. Second, the comparison between experimentally measured HRTFs and numerically computed ones has not been demonstrated satisfactorily at all frequencies of interest. The reasons for this could lie in inaccuracies in the simulation procedure or in the quality of the mesh discretization.

For the computations to be useful, the HRTF information must be extracted and presented in a manner similar to the experiments. Further, the results of the computations and experiments must be compared to see that they agree, and any discrepancies explained. Table I summarizes the results of several papers. Due to the inability of previous authors to simulate the entire frequency range, such comparisons were attempted only for restricted ranges, and the results showed less than satisfactory agreement between experimental and the computed data.[14]

In this paper we use the preconditioned fast multipole accelerated boundary element method,[15] a new formulation of the HRTF problem, and good meshes of the head and torso to compute the HRTF. We then compare the computations to experiments. In Sec. II, the theoretical background for HRTF computations is discussed. In particular, we derive the equations necessary to use the reciprocal set-up for computations. In Sec. III, we introduce the spherical representation of the HRTF corresponding to insonification from plane-waves from various directions. It is shown how one may then derive the HRTF corresponding to a source at a finite range from this representation. In Sec. IV we briefly describe the numerical methods used in this paper, namely, the preconditioned fast multiple method (FMM) BEM, and its use to compute the spherical harmonic spectrum. Section V presents numerical results for computations with an analytical solution (a sphere), the Neumann KU-100 "Fritz" head, and the Knowles Electronics Manikin for Acoustic Research (KEMAR) head and head-and-torso models. Section VI concludes the paper.

II. PROBLEM FORMULATION

A. Direct computations

To compute the HRTF for a given frequency $f=\omega/(2\pi)$ the three dimensional (3D) Helmholtz equation must be solved to determine the scattered field for different incident fields. Let the origin of the reference frame be placed at some point inside the head, which is referred to as the "head center" and let $\mathbf{r}$ denote the radius-vector of any spatial point in this frame. Along with the Cartesian coordinates (x,y,z) the spherical coordinates (r,θ,φ) defined as

$$\mathbf{r}=(x,y,z)=r(\sin\theta\cos\varphi,\sin\theta\sin\varphi,\cos\theta),$$

$$0\leqslant\varphi<2\pi,\quad 0\leqslant\theta\leqslant\pi,\quad r=|\mathbf{r}| \tag{1}$$

are also used. Let S denote the surface of the scatterer and V the infinite external domain.

When the incident field is generated by a source, the sphere S' or radius R (range) centered at the origin, which includes surface S and the source location, must also be considered. This sphere is characterized by the vector $\mathbf{r}'$, or alternately by the radius R and a unit vector $\mathbf{s}$,

$$\mathbf{r}'=R\mathbf{s},\quad \mathbf{s}=(\sin\theta'\cos\varphi',\sin\theta'\sin\varphi',\cos\theta'),\quad |\mathbf{s}|=1. \tag{2}$$

Thus, the incident complex pressure field generated by a unit intensity monopole source placed on S' is

$$\Phi^{\text{in}}(\mathbf{r};\mathbf{s})=\Phi^{\text{in}}_R(\mathbf{r};\mathbf{r}')=G(|\mathbf{r}-\mathbf{r}'|)=G(|\mathbf{r}-R\mathbf{s}|), \tag{3}$$

$$G(r)=\frac{e^{ikr}}{4\pi r},\quad k=\frac{\omega}{c}. \tag{4}$$

Here G is the free-field Green's function for the Helmholtz equation, k is the wavenumber, and c is the speed of sound in air.

The other case considered is when the field is a plane-wave incident from direction $\mathbf{s}$:

$$\Phi^{\text{in}}(\mathbf{r};\mathbf{s})=e^{-ik\mathbf{s}\cdot\mathbf{r}},\quad |\mathbf{s}|=1. \tag{5}$$

This can be considered as the limit of the field generated by the monopole at large R since

$$\Phi^{\text{in}}(\mathbf{r};\mathbf{s})=e^{-ik\mathbf{s}\cdot\mathbf{r}}=\lim_{R\to\infty}[4\pi Re^{-ikR}G(\mathbf{r}-R\mathbf{s})]$$
$$=\lim_{R\to\infty}[4\pi Re^{-ikR}\Phi^{\text{in}}_R(\mathbf{r};R\mathbf{s})]. \tag{6}$$

The total acoustic field can be represented as

$$\Phi = \Phi^{\text{in}} + \Phi^{\text{scat}}, \tag{7}$$

where the scattered field Φ^{scat} is a regular function in V which can be found by solution of the Helmholtz equation

$$\nabla^2\Phi^{\text{scat}} + k^2\Phi^{\text{scat}} = 0, \quad \mathbf{r} \in V, \tag{8}$$

with the Sommerfeld radiation condition

$$\lim_{r\to\infty}\left[r\left(\frac{\partial\Phi^{\text{scat}}}{\partial r} - ik\Phi^{\text{scat}} \right)\right] = 0, \tag{9}$$

and boundary conditions on S, which for simplicity and as is customary in HRTF calculations, is assumed to be a sound-hard surface (the impedance boundary conditions could also be imposed):

$$\left.\frac{\partial\Phi^{\text{scat}}}{\partial n}\right|_S = -\left.\frac{\partial\Phi^{\text{in}}}{\partial n}\right|_S, \quad \left(\left.\frac{\partial\Phi}{\partial n}\right|_S = 0 \right). \tag{10}$$

This problem is closed and its solution is unique. However, this formulation is wasteful of computations since to compute the HRTF we need to compute Φ only at one point on the surface S, corresponding to the receiver location, whereas this procedure returns the solution everywhere on S. A more efficient formulation is presented later.

Let the location of the receiver be denoted as $\mathbf{r}_*$. We must determine the function

$$H(\mathbf{s}) = \begin{cases} \Phi(\mathbf{r}_*; R\mathbf{s}) \\ \Phi(\mathbf{r}_*; \mathbf{s}) \end{cases}, \quad \mathbf{r}_* \in S, \tag{11}$$

where the upper and lower formulas correspond to the monopole and plane-wave incident fields.

When the complex valued function $H(\mathbf{s})$ above is normalized by the value of the incident field at the head center, which by the choice of the reference frame is the origin, we obtain the conventional HRTF. Note that the unitary plane-wave satisfies $e^{-ik\mathbf{s}\cdot\mathbf{r}} = 1$ at $\mathbf{r} = \mathbf{0}$, so $H(\mathbf{s})$ is already normalized and is the complex HRTF.

For the incident field generated by a source we have

$$\begin{aligned} \text{HRTF}(\mathbf{s};R) &= \frac{H(\mathbf{s};R)}{G(|R\mathbf{s}|)} = \frac{H(\mathbf{s};R)}{G(R)} = 4\pi R e^{-ikR} H(\mathbf{s};R) \\ &= 4\pi R e^{-ikR}\Phi(\mathbf{r}_*; R\mathbf{s}). \end{aligned} \tag{12}$$

Due to the linearity of the scattering problem and Eq. (6),

$$\text{HRTF}(\mathbf{s};\infty) = \text{HRTF}^{(\text{pw})}(\mathbf{s};\infty) = H^{(\text{pw})}(\mathbf{s}). \tag{13}$$

The superscript (pw) is used to show that the quantity is related to plane-wave incidence.

B. Modified boundary value problem for the reciprocal case

Even though direct solution can provide the HRTF, it is obvious that HRTF computations will be excessively long, as a separate solution is needed for each direction. Further, the only part of the solution retained is the acoustic pressure at one receiver point, with the rest discarded! To avoid this, the reciprocity principle[16] can be used (see Refs. 6 and 7). This principle states that the positions of the source and receiver in a given domain can be exchanged, while not affecting the value of the potential at the receiver location. Hence, in our case the source is placed on the domain boundary. However, the singularity of the source can cause numerical and conceptual problems unless it is explicitly treated, as here. In this section we will derive a technique that treats these problems by separating the leading order singular behavior of the scattered field by using an analytical expression for it (the image source), and then obtaining a better-conditioned boundary value problem for the remaining part of the field.

To derive appropriate forms of the boundary value problem for this situation, we place the source at the point

$$\mathbf{r}_{*\epsilon} = \mathbf{r}_* + \epsilon\mathbf{n}, \tag{14}$$

where $\mathbf{n}$ is the normal to the boundary directed inside V and ϵ is assumed small enough so that due to the continuity of the pressure, the values obtained for $\Phi(\mathbf{r}_*;\mathbf{s})$ and $\Phi(\mathbf{r}_{*\epsilon};\mathbf{s})$ are close. We will then consider the limiting case $\epsilon\to 0$ which will provide the result for the source placed exactly on the boundary *independent* of ϵ.

1. Incident field generated by a source

Let $\phi(\mathbf{r};\mathbf{r}_{*\epsilon})$ be a solution of the following problem:

$$\begin{aligned} &\phi = \phi^{\text{in}} + \phi^{\text{scat}}, \quad \phi^{\text{in}} = G(|\mathbf{r} - \mathbf{r}_{*\epsilon}|), \\ &\nabla^2\phi^{\text{scat}} + k^2\phi^{\text{scat}} = 0, \quad \mathbf{r} \in V, \\ &\lim_{r\to\infty}\left[r\left(\frac{\partial\phi^{\text{scat}}}{\partial r} - ik\phi^{\text{scat}} \right)\right] = 0, \\ &\left.\frac{\partial\phi^{\text{scat}}}{\partial n}\right|_S = -\left.\frac{\partial\phi^{\text{in}}}{\partial n}\right|_S, \quad \left(\left.\frac{\partial\phi}{\partial n}\right|_S = 0 \right). \end{aligned} \tag{15}$$

Consider the case when the incident field in the direct problem formulation is generated by a source located at $\mathbf{r}'$, the solution to which we denote as $\Phi(\mathbf{r};\mathbf{r}')$. Due to the reciprocity principle, we have

$$\phi(\mathbf{r}';\mathbf{r}_{*\epsilon}) = \Phi(\mathbf{r}_{*\epsilon};\mathbf{r}'). \tag{16}$$

Therefore, $H(\mathbf{s})$ (11) can be computed as

$$H(\mathbf{s}) = \lim_{\epsilon\to 0}\Phi(\mathbf{r}_{*\epsilon};\mathbf{r}') = \lim_{\epsilon\to 0}\phi(\mathbf{r}';\mathbf{r}_{*\epsilon}) = \lim_{\epsilon\to 0}\phi(R\mathbf{s};\mathbf{r}_{*\epsilon}). \tag{17}$$

Decomposing the total field into incident and scattered parts (15) this can be rewritten as

$$H(\mathbf{s}) = G(|\mathbf{r}_* + R\mathbf{s}|) + \lim_{\epsilon\to 0}\phi^{\text{scat}}(R\mathbf{s};\mathbf{r}_{*\epsilon}). \tag{18}$$

Since $\mathbf{r}' \notin S$, and this point is far from the boundary, the function $\phi^{\text{scat}}(\mathbf{r}';\mathbf{r}_{*\epsilon})$ is regular and well behaved. The limit in Eq. (18) exists and is finite. However, in numerical solution of the problem the boundary value $\phi^{\text{scat}}(\mathbf{r}_*;\mathbf{r}_{*\epsilon}) \sim \epsilon^{-1}$, which causes substantial numerical errors for small ϵ. (In

fact, previous authors have attempted to deal with the locally high gradients by modifying the mesh to account for this, see, e.g., Ref. 6.)

A straightforward use of Green's identity for determining $\phi^{\text{scat}}(\mathbf{r}';\mathbf{r}_{*\epsilon})$ is also problematic. To treat the singular behavior of the scattered field, we will use the fact that for a sound hard flat boundary the scattered field for a source located at $\mathbf{r}_{*\epsilon}$ can be found by placing the source of the same intensity (*image*) at $\mathbf{r}^{-}_{*\epsilon}=\mathbf{r}_{*}-\epsilon\mathbf{n}$. So when we have a smooth curved boundary we will represent ϕ^{scat} as a sum of the leading image term and an additional regular radiating function ψ, which satisfies the Helmholtz equation in V,

$$\phi^{\text{scat}} = G(|\mathbf{r}-\mathbf{r}^{-}_{*\epsilon}|) + \psi, \tag{19}$$

and the following boundary condition:

$$\left.\frac{\partial\psi}{\partial n}\right|_S = \left[-\frac{\partial}{\partial n}G(|\mathbf{r}-\mathbf{r}^{-}_{*\epsilon}|) - \frac{\partial\phi^{\text{in}}}{\partial n}\right]\Bigg|_S = -\left.\frac{\partial}{\partial n}[G(|\mathbf{r}-\mathbf{r}^{-}_{*\epsilon}|) + G(|\mathbf{r}-\mathbf{r}_{*\epsilon}|)]\right|_S. \tag{20}$$

This formula allows us to take the limit $\epsilon\to 0$, which shows that

$$\left.\frac{\partial\psi}{\partial n}\right|_S = \begin{cases} 0, & \mathbf{r}=\mathbf{r}_* \\ -2\dfrac{\partial}{\partial n}G(|\mathbf{r}-\mathbf{r}_*|), & \mathbf{r}\neq\mathbf{r}_*. \end{cases} \tag{21}$$

The required function $H(\mathbf{s})$ can be found from Eqs. (18) and (19) as

$$H(\mathbf{s}) = 2G(|\mathbf{r}_*-R\mathbf{s}|) + \psi(R\mathbf{s};\mathbf{r}_*). \tag{22}$$

2. Use of Green's identity

Since ψ is a regular function in V, Green's identity can be applied to determine ψ at any field point including $\mathbf{r}'$ via its boundary values, which can be computed using the BEM:

$$-\psi(\mathbf{r}') = L[q] - M[\psi], \quad q = \left.\frac{\partial\psi}{\partial n}\right|_S, \tag{23}$$

where L and M denote the boundary operators

$$L[q] = \int_S q(\mathbf{r})G(|\mathbf{r}-\mathbf{r}'|)dS(\mathbf{r}),$$

$$M[\psi] = \int_S \psi(\mathbf{r})\frac{\partial G(|\mathbf{r}-\mathbf{r}'|)}{\partial n(\mathbf{r})}dS(\mathbf{r}), \tag{24}$$

where the normal orientation is consistent with Eq. (14).

Use of Green's identity allows us to link the function $H(\mathbf{s})$ (22) with the boundary values of ψ:

$$H(\mathbf{s}) = 2G(|\mathbf{r}_*-R\mathbf{s}|) + \int_S\left[\psi(\mathbf{r};\mathbf{r}_*)\frac{\partial G(|\mathbf{r}-R\mathbf{s}|)}{\partial n(\mathbf{r})} - q(\mathbf{r};\mathbf{r}_*)G(|\mathbf{r}-R\mathbf{s}|)\right]dS(\mathbf{r}). \tag{25}$$

3. Plane-wave incident field

For plane-wave incidence (5) $H(\mathbf{s})$ can be found via the solution of the source on the boundary as well. Consider the solution of the problem where instead of a unit intensity source we have a source of intensity $4\pi Re^{-ikR}$. Due to linearity $H(\mathbf{s})$ can be computed according to Eq. (25) as

$$H(\mathbf{s}) = 4\pi Re^{-ikR}\left[2G(|\mathbf{r}_*-R\mathbf{s}|) + \int_S\left[\psi(\mathbf{r};\mathbf{r}_*)\frac{\partial G(|\mathbf{r}-R\mathbf{s}|)}{\partial n(\mathbf{r})} - q(\mathbf{r};\mathbf{r}_*)G(|\mathbf{r}-R\mathbf{s}|)\right]dS(\mathbf{r})\right]. \tag{26}$$

Consider now the limit $R\to\infty$. Equation (6) shows that in this case the incident field for the direct problem will be $e^{ik\mathbf{s}\cdot\mathbf{r}}$. The same formula applied to Eq. (26) results in

$$H^{(\text{pw})}(\mathbf{s}) = 2e^{-ik\mathbf{s}\cdot\mathbf{r}_*} - \int_S e^{-ik\mathbf{s}\cdot\mathbf{r}}[ik\mathbf{n}(\mathbf{r})\cdot\mathbf{s}\psi(\mathbf{r};\mathbf{r}_*) + q(\mathbf{r};\mathbf{r}_*)]dS(\mathbf{r}), \tag{27}$$

since

$$\lim_{R\to\infty}\left[4\pi Re^{-ikR}\frac{\partial G(|\mathbf{r}-R\mathbf{s}|)}{\partial n}\right] = \mathbf{n}\cdot\nabla_{\mathbf{r}}\lim_{R\to\infty}[4\pi Re^{-ikR}G(|\mathbf{r}-R\mathbf{s}|)] = \mathbf{n}\cdot\nabla_{\mathbf{r}}e^{ik\mathbf{s}\cdot\mathbf{r}} = -ik\mathbf{n}\cdot\mathbf{s}e^{ik\mathbf{s}\cdot\mathbf{r}}. \tag{28}$$

III. SPHERICAL REPRESENTATIONS

The HRTF corresponding to plane-wave incidence is a function of direction. In this section we will represent this function, $H(\mathbf{s})$, as a regular function defined on the unit sphere. Since the spherical harmonics form a basis on this surface, it is natural to represent $H(\mathbf{s})$ as

$$H(\mathbf{s}) = \sum_{n=0}^{\infty}\sum_{m=-n}^{n} H_n^m Y_n^m(\mathbf{s}), \tag{29}$$

where the orthonormal spherical harmonics for $\mathbf{s}=(\sin\theta\cos\varphi, \sin\theta\sin\varphi, \cos\theta)$ are defined as

$$Y_n^m(\mathbf{s}) = (-1)^m\sqrt{\frac{2n+1}{4\pi}\frac{(n-|m|)!}{(n+|m|)!}}P_n^{|m|}(\cos\theta)e^{im\varphi}, \quad n=0,1,2,\ldots, \quad m=-n,\ldots,n, \tag{30}$$

and $P_n^{|m|}(\mu)$ are the associated Legendre functions consistent with that in Ref. 17.

In this basis the coefficients of expansion (29) can be found as

$$H_n^m = \int_{S_u} H(\mathbf{s})Y_n^{-m}(\mathbf{s})dS(\mathbf{s}), \tag{31}$$

(where S_u is the unit sphere surface), using orthonormality

$$\int_{S_u} Y_{n'}^{m'}(\mathbf{s}) Y_n^{-m}(\mathbf{s}) dS(\mathbf{s}) = \delta_{mm'}\delta_{nn'}, \tag{32}$$

where δ is the Kronecker symbol.

A. Plane-wave incident field

Substituting Eq. (27) in Eq. (31)

$$H_n^{(\mathrm{pw})m} = \int_{S_u} \left[2e^{-ik\mathbf{s}\cdot\mathbf{r}_*} - \int_S e^{ik\mathbf{s}\cdot\mathbf{r}} [ik\mathbf{n}(\mathbf{r})\cdot\mathbf{s}\psi(\mathbf{r};\mathbf{r}_*) + q(\mathbf{r};\mathbf{r}_*)] dS(\mathbf{r}) \right] Y_n^{-m}(\mathbf{s}) dS(\mathbf{s}). \tag{33}$$

The Gegenbauer formula for plane-wave expansion and the addition theorem for spherical harmonics (e.g., see Ref. 18) provide

$$e^{-ik\mathbf{s}\cdot\mathbf{r}} = 4\pi \sum_{n'=0}^{\infty} \sum_{m'=-n'}^{n'} i^{-n'} Y_{n'}^{m'}(\mathbf{s}) R_{n'}^{-m'}(\mathbf{r}),$$

$$R_{n'}^{m'}(\mathbf{r}) = j_{n'}(kr) Y_{n'}^{m'}\left(\frac{\mathbf{r}}{r}\right),$$

$$-ik\mathbf{n}\cdot\mathbf{s}e^{-ik\mathbf{s}\cdot\mathbf{r}} = \mathbf{n}\cdot\nabla_{\mathbf{r}} e^{-ik\mathbf{s}\cdot\mathbf{r}} = 4\pi \sum_{n'=0}^{\infty} \sum_{m'=-n'}^{n'} i^{-n'} Y_{n'}^{m'}(\mathbf{s}) [\mathbf{n}\cdot\nabla R_{n'}^{-m'}(\mathbf{r})], \tag{34}$$

where $j_n(kr)$ are the spherical Bessel functions of the first kind and $R_n^m(\mathbf{r})$ are the regular spherical wave functions. Substituting these in Eq. (33) and using orthonormality (32),

$$H_n^{(\mathrm{pw})m} = 4\pi i^{-n} \left\{ 2R_n^{-m}(\mathbf{r}_*) + \int_S [\psi(\mathbf{r};\mathbf{r}_*)\mathbf{n}(\mathbf{r})\cdot\nabla R_n^{-m}(\mathbf{r}) - q(\mathbf{r};\mathbf{r}_*) R_n^{-m}(\mathbf{r})] dS(\mathbf{r}) \right\}. \tag{35}$$

B. Incident field generated by a source

A regular incident field can be represented as a sum of plane-waves, and so the coefficients of the expansion of $H(s)$ over the spherical harmonics basis, H_n^m, can be expressed via the corresponding coefficients for plane-wave incidence $H_n^{(\mathrm{pw})m}$. For example, the incident field generated by a source (3) inside the sphere S' can be represented outside it as (e.g., Ref. 18)

$$\Phi^{\mathrm{in}}(\mathbf{r};\mathbf{s}) = G(|\mathbf{r} - R\mathbf{s}|) = ik \sum_{n=0}^{\infty} \sum_{m=-n}^{n} S_n^m(R\mathbf{s}) R_n^{-m}(\mathbf{r}),$$

$$S_n^m(\mathbf{r}) = h_n(kr) Y_n^m\left(\frac{\mathbf{r}}{r}\right), \tag{36}$$

where $h_n(kr)$ are the spherical Hankel functions of the first kind and $S_n^m(\mathbf{r})$ are the regular spherical wave functions. Compared with Eq. (34), we see that for the incident field of a monopole source

$$H_n^m = \frac{ik}{4\pi} i^n h_n(kR) H_n^{(\mathrm{pw})m}. \tag{37}$$

This expression provides an easy way of relating HRTFs measured at finite range with those corresponding to plane-wave incidence and with those corresponding to other ranges as well.

IV. COMPUTATIONAL METHODS

The FMM accelerated BEM formulation for the Neumann problem for the Helmholtz equation in 3D (Ref. 15) is at the heart of the computational technique. Computation is done in two parts: solution and evaluation. The first step is solution of the problem for the scattered field ψ subject to the boundary conditions (21). The second step is determination of the expansion coefficients $H_n^{(\mathrm{pw})m}$ of the function $H^{(\mathrm{pw})}(\mathbf{s})$, which is done using Eq. (35). Note that the coefficients $H_n^{(\mathrm{pw})m}$ (whose number depends on the frequency) provide the theoretically most compact representation of $H^{(\mathrm{pw})}(\mathbf{s})$ within a given bandwidth. Further, the use of the spherical representation provides a fast and accurate interpolation of $H^{(\mathrm{pw})}(\mathbf{s})$ within a given bandwidth for any other direction.

The evaluation part can then be performed relatively quickly for an arbitrary range R. The case $R\to\infty$ is, of course, just $H_n^{(\mathrm{pw})m}$ itself. Computation of the coefficients H_n^m for fields generated by monopole sources is fast, as it can be done using Eq. (37). After that the function $H(\mathbf{s})$ can be determined for any given direction using Eq. (29), and the HRTF or any other related functions can be computed.

A. BEM accelerated by the FMM

The boundary element method used in the present study is described in detail in Ref. 15, where a review of related work can also be found. This software uses the Burton–Miller[19] formulation. The boundary integral equation solved is

$$-\tfrac{1}{2}[\psi(\mathbf{r}') + \lambda q(\mathbf{r}')] = (L + \lambda L')[q] - (M + \lambda M')[\phi], \quad \mathbf{r}' \in S, \tag{38}$$

where $\mathbf{r}'$ is a point on the boundary, λ is a complex valued regularization parameter, the operators L and M are provided by Eq. (24), while the other operators are

$$L'[q] = \int_S q(\mathbf{r}) \frac{\partial G(|\mathbf{r} - \mathbf{r}'|)}{\partial n(\mathbf{r})} dS(\mathbf{r}),$$

$$M'[\psi] = \frac{\partial}{\partial n(\mathbf{r})} \int_S \psi(\mathbf{r}) \frac{\partial G(|\mathbf{r} - \mathbf{r}'|)}{\partial n(\mathbf{r})} dS(\mathbf{r}). \tag{39}$$

The boundary is discretized by a mesh with N_{vert} vertices and N_{el} flat triangular elements. The discretized equation is collocated either at the vertices or at the centers of elements using a constant panel approximation. The flexible preconditioned GMRES (Ref. 26) was used to solve the resulting linear system.

The FMM used in the BEM solver has the following scaling for CPU time: for problems with large kD, where D is the maximum size of the computational domain and k the wavenumber, this algorithm has theoretical time complexity $O((kD)^3)$ for $kD \gg 10^3$ and $O((kD)^{2+\alpha})$ with a relatively small α for $kD \lesssim 10^3$. These scalings were validated experimentally on benchmark problems for $0.0001 < kD < 500$ in Ref. 15. At high frequencies the algorithm has memory complexity $O((kD)^2 \log(kD))$. At low kD the time complexity depends on kD weakly and is determined rather by the size of the system N [complexity $O(N \log N)$] not by kD. It also requires $O(N \log N)$ memory. This is in contrast to the scaling of conventional BEM solvers which have a time complexity of $O((kD)^6)$ and a memory complexity of $O((kD)^4)$. Because of this, the HRTF calculation presented here can be done for the entire audible frequency range on a mesh that includes the torso on a conventional desktop workstation.

B. Spherical harmonic analysis

1. Truncation error and error bound

The number of spherical harmonics needed to represent the function $H(\mathbf{s})$ for any given frequency or wavenumber k can be deduced from the Gegenbauer plane-wave decomposition (34), where the infinite sum over n is truncated to a maximum value $n=p-1$. This provides a p^2 term approximation via an expansion over the spherical functions R_n^m. Using the addition theorem for spherical harmonics, we can rewrite Eq. (34) as

$$e^{-ik\mathbf{s}\cdot\mathbf{r}} = 4\pi \sum_{n=0}^{\infty} i^{-n} j_n(kr) \sum_{m=-n}^{n} Y_n^m(\mathbf{s}) Y_n^{-m}\left(\frac{\mathbf{r}}{r}\right)$$
$$= \sum_{n=0}^{\infty} (2n+1) i^{-n} j_n(kr) P_n\left(\mathbf{s}\cdot\frac{\mathbf{r}}{r}\right)$$
$$= \sum_{n=0}^{p-1} (2n+1) i^{-n} j_n(kr) P_n\left(\mathbf{s}\cdot\frac{\mathbf{r}}{r}\right) + \epsilon_p(kr), \quad (40)$$

where $\epsilon_p(kr)$ is the truncation error and P_n are the Legendre polynomials, $|P_n| \leq 1$. This error for $n \geq p > kr$, for which $j_n(kr)$ are strictly positive functions, can be bounded as

$$|\epsilon_p(kr)| = \left| \sum_{n=p}^{\infty} (2n+1) i^{-n} j_n(kr) P_n\left(\mathbf{s}\cdot\frac{\mathbf{r}}{r}\right) \right|$$
$$\leq \sum_{n=p}^{\infty} (2n+1) j_n(kr) < \sum_{n=p}^{\infty} (2n+1) j_n(ka), \quad (41)$$

where a is the maximum possible value of r. Since for computation of H_n^m parameter r is the distance from the scatterer center to a point located on the scatterer surface [see Eq. (33)], a is the minimal possible value of range R for which spherical decomposition can be used, so we set $a=R_{\min}$.

The error bounds shown in the literature for the sum (41) are related to decompositions of plane-wave over the wave functions (e.g., Ref. 20). A more strict bound valid for any ka for this sum is found in Ref. 21 and is

$$|\epsilon_p| \leq \sum_{n=p}^{\infty} \frac{(ka)^n}{(2n-1)!!} < \left(\frac{ka}{2}\right) \sum_{n=p-1}^{\infty} \frac{(ka)^n}{2^n n!}$$
$$= \frac{e^{ka/2}}{(p-1)!}\left(\frac{ka}{2}\right)^p, \quad p \geq 4. \quad (42)$$

This substantially overestimates the p required to obtain a given accuracy $|\epsilon_p|$ for large ka for which the asymptotic bound[21]

$$|\epsilon_p| \lesssim (2p+1) j_p(ka) \lesssim \frac{p^{2/3}}{(ka)^{1/2}} \exp\left(-\frac{1}{3}(2\eta_p^{(a)})^{3/2}\right),$$

$$\eta_p^{(a)} = \frac{p-ka+1/2}{(ka)^{1/3}}, \quad p > ka \gg 1 \quad (43)$$

is more appropriate. This formula results in the following estimate of p for prescribed error $\epsilon = |\epsilon_p|$:[21]

$$p \gtrsim ka + \frac{1}{2}\left(3 \ln\frac{1}{\epsilon} + \frac{1}{2}\ln(ka)\right)^{2/3} (ka)^{1/3}, \quad ka \gg 1. \quad (44)$$

We also note a more simplistic approximation of $p(ka)$ for any ka used in the literature (e.g., Ref. 22)

$$p = ka + 4(ka)^{1/3} + 3. \quad (45)$$

The error of this approximation grows with ka [as it misses a slow growth $\sim\ln(ka)$ of the coefficient near $(ka)^{1/3}$, which actually can be estimated using Eq. (43) (which is a more accurate estimate than Eq. (44))] with p from Eq. (45).

We also note that for computation of the HRTF for finite range we have more restrictive conditions on p which increase it due to the requirement of the accuracy of source field expansion (36). A detailed error bound analysis for this case is provided in Ref. 18, where also the following approximation for p was obtained and checked for a broad range of frequencies:

$$p \approx \left\{ \left[\frac{1}{\ln\delta} \ln \frac{\delta^{3/2}}{\epsilon(\delta-1)^{3/2}} + 1 \right]^4 + \left[ka + \frac{1}{2}\left(3\ln\frac{1}{\epsilon} + \frac{1}{2}\ln(ka)\right)^{2/3} (ka)^{1/3} \right]^4 \right\}^{1/4}, \quad \delta = \frac{R}{a}. \quad (46)$$

Note that this formula is modified for the HRTF, which is a normalized function (12) (in Ref. 18 the equation is derived based on the absolute error). It shows that for $\delta=1$, i.e., when the source is located on the boundary of the expansion domain, $p=\infty$, while p is finite for $\delta>1$. For ranges $R>2a$ the order of magnitude of the truncation error will be the same as that for the expansion of the plane-wave and more simplistic estimates can be used.

2. Computation of expansion coefficients

As soon as the required bandwidth for a given error bound, frequency, and domain size for the spherical harmonic representation is known, the p^2 coefficients $H_n^{(\mathrm{pw})m}$ can be computed using the direct formula (35). Note that computation of functions $R_n^m(\mathbf{r})$ and their directional derivatives $\mathbf{n}(\mathbf{r})\cdot\nabla R_n^m(\mathbf{r})$ is done via a standard recursive procedure used

in the FMM for the Helmholtz equation, which is inexpensive with a computational complexity of $O(Np^2)$. Also the integrals should be computed in a way consistent with the procedure of computation of the BEM integrals (e.g., see treatment of singular integrals[15]). While computation of integrals (35) is usually much faster than solution of the linear system, for large problems these computations can be accelerated via the FMM using the fact that function $H(\mathbf{s})$ is a sum of dipoles and monopoles [see Eq. (25)] of given intensities, and coefficients of expansion of this function over the spherical basis can be retrieved immediately from the expansions appeared in the FMM (see Ref. 15).

V. NUMERICAL TESTS

We first conducted some tests on simple geometrical models for which the HRTF can be obtained either analytically or semi-analytically without using the BEM, reciprocity principle, and decompositions/synthesis over spherical basis functions and compare it with the present method which uses all these. These tests were used to validate the method and figure out the settings appropriate for computations with more complex geometries. After that we performed several tests on a complex head geometry in the presence and absence of the torso and compared this with the results from simpler geometries. Also, we tested the BEM solutions against analytical results for arbitrary meshes, such as by computing the field generated by a source placed inside the body. For the meshes used the relative errors in the solution for the worst cases were of order 1% which are consistent with usual BEM errors. These tests provided us with the confidence that the method can be used for modeling more complex situations, and we then compared the results of the computations with the HRTF measured in experiments.

A. Tests for head mesh

1. Mesh generation

Computation of scattering off complex shapes, such as human head, requires quality surface meshes. A quality triangular mesh satisfies the requirements that the mesh is fine enough to satisfy the requirement of at least five elements per wavelength, that all triangles be as close to equilateral as possible, that the number of triangles associated with each vertex of the mesh be as close to 6 as possible, and finally that the rate of change in triangle areas across different regions of the mesh be gradual. The use of the version of the FMM accelerated BEM described in Ref. 15 requires high quality meshes, which ideally should have a uniform distribution of the mesh vertices over the surface.

On the other hand, the data coming from laser scans of heads and meshed using standard triangulation techniques, such as those based on the Delaunay triangulation, are acceptable for graphics, but have a poor quality for the BEM. Several such meshes were available to the researchers. To make them appropriate for the BEM we developed a remeshing technique which produced meshes of the desired quality. Note that in the literature several techniques for generation of quality meshes are available [e.g., in the Computational Geometry Algorithms Library (CGAL) project]. A detailed description of our technique is beyond the scope of the present paper, and we only mention that the technique is based on the 3D radial basis function (RBF) interpolation combined with a relaxation technique, which allows the mapping of meshes between topologically equivalent objects. Details of the technique can be found in our technical report.[23] We note that for high frequency computations up to 20 kHz the head mesh contained of order 10^5 elements, and the maximum value of the computational parameter kD was 181. The RBF technique for such a large data set was enabled by the use of the FMM for the biharmonic equation and $O(N \log N)$ preconditioning technique, as described in Ref. 24.

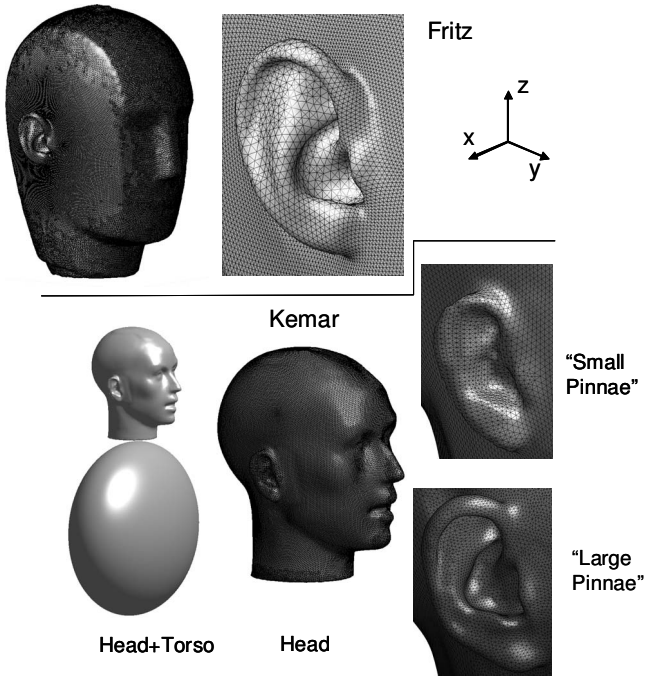


FIG. 1. The meshes used for computation of the HRTF using the FMM accelerated BEM for the Fritz and Kemar models. The Fritz head mesh contains 152 666 triangular panels, while Kemar head+torso meshes contain 426 846 and 445 276 panels for "small pinnae" and "large pinnae" cases, respectively.

The resulting mesh was checked for quality using several geometrical parameters (minimal and maximum size of the triangles, aspect ratios, etc.) and for several elements an additional subdivision procedure was applied. In this way it was guaranteed that the surface is discretized with not less than six elements per wavelength for the highest frequency computed. Figure 1 shows the computational mesh for the Neumann KU 100 Dummy Head model (referred to as Fritz). This mesh contains 152 666 triangles. We also did computations with two Kemar head meshes which contain 99 166 triangles ("small pinnae" mesh) and 117 596 triangles ("large pinnae" mesh), respectively (see Fig. 1). Note that the small pinnae Kemar mesh was generated from a laser scan, while the large pinnae mesh was obtained in the following way. A mesh of good quality was generated from the CT-scan of the pinnae provided by Duda and Algazi.[25] The pinna areas of the original Kemar head were cut off, and the pinnae meshes were positioned at an angle corresponding to measurements

TABLE II. Performance for the Fritz and Kemar (head with small pinnae) models.

Frequency (kHz)	0.172	0.517	1.034	2.067	4.996	9.991	14.987	20.155
				BEM solution (s)				
Fritz	131	125	159	334	558	919	1254	1483
Kemar	62	64	86	160	341	417	520	628
				Computation of $H_n^{(\mathrm{pw})m}$ (s)				
Fritz	1.2	1.8	2.9	5.6	13	35	64	103
Kemar	0.9	1.3	1.8	3.3	8	21	36	58
				Memory (Gbyte)				
Fritz	2.07	2.10	2.14	2.20	2.30	2.37	2.44	2.60
Kemar	3.12	3.16	3.20	3.25	3.30	3.32	3.34	3.43

in the UC Davis Center for Image Processing and Integrated Computing (CIPIC) database.[3] After that a mesh "stitching" and relaxation technique was used to merge the meshes. Below, if it is not mentioned otherwise specifically, the term "Kemar head mesh" means the original model with the small pinnae.

2. Performance

In the present study we used the BEM with the panel-center collocation method, since it is more appropriate to the image source method. The algorithm was executed on a four core PC (Intel Core 2 Extreme QX6700 2.66 GHz, 8 Gbyte RAM) using OpenMP parallelization. Table II shows the wall clock time (includes input/output time for each procedure) and memory required for several cases. The total execution time for the run, which includes BEM solution, and computation of the spherical spectrum over 117 frequencies equispaced from 172 to 20 155 Hz took about 32 h for Fritz and 14 h for the Kemar head.

3. Spherical spectra

We analyzed the spherical spectra for the computed cases and found that the truncation occurs in the region of spectral (exponential) convergence, and the values of the truncated modes are several orders of magnitude smaller than the values of the modes stored for the HRTF representation. Computed spectra were compared with analytical results for a sphere and showed a good agreement if the size of the effective sphere is selected as the minimal radius of the head (the shortest distance from the head surface to the center). Particularly, the limiting value of the mean mode $H_0^{(\mathrm{pw})0}(0) = (4\pi)^{1/2}$ was achieved with an accuracy of four digits; good quantitative match was found for $ka < 1$ (frequencies below 440 Hz) and for frequency of oscillations of the mean mode $H_0^{(\mathrm{pw})0}$ as a function of ka.

4. HRTF

Figure 2 shows a comparison between the plane-wave HRTF for the Fritz model and for a sphere of radius $a_{\min}$. The bright side of the imaging sphere corresponds to azimuth angles $-90° \leqslant \varphi \leqslant 90°$ [see Eq. (1) and the reference frame in Fig. 1], while the dark side corresponds to azimuths $90° \leqslant \varphi \leqslant 270°$. It is seen that at low frequencies, such as at 1 kHz the HRTF for the sphere and for the Fritz head is qualitatively similar. It is also seen that at these frequencies "the bright spot" on the dark side of the imaging sphere is well defined for both the head shapes. At higher frequencies, such as 10 kHz the HRTFs for the Fritz model and sphere are very different. On the ipsilateral side (the side of the ear) the maximum amplitude of the HRTF for the sphere is at the ear location. At the same time for the Fritz model a notch is observed at this location, where the HRTF amplitude can be 20 dB or so less than its maximum value. The pictures for the contralateral side are also different. If for the sphere we can see the bright spot surrounded by a circular interference pattern, the Fritz model demonstrates rather irregular filamentary structure of the notches, where the amplitude of the HRTF can be tens of decibels lower than the average (in fact, the bright spot and circular interference pattern are present also in the Fritz model, but it is substantially masked by the notch structure).

The frequency dependence of the plane-wave HRTF along with its dependence on the spherical angles is presented in Fig. 3. Again we can see the difference between the HRTF for the sphere and for the Fritz and Kemar head models. The dynamic range of the HRTF for Fritz and Kemar is much larger than that for the sphere. The amplitude at the notches can be tens of decibels lower than the average values. It is also seen that while the HRTF for the sphere is clearly symmetric for elevation angles $\theta = 90° \pm \alpha$, the HRTF for the head models is different for the planes defined by elevation angles $\theta = 45°$ and $\theta = 135°$ which cut the head above and below the elevation angle at the plane of ear location, $\theta = 90°$. The figure also shows a difference between the HRTFs for Fritz and Kemar head models and the effect of the pinnae shape since the difference between the two Kemar head meshes is only due to the different pinnae.

5. Comparison with the computations of other authors

Fortunately, there are computational results for the Fritz head model computed by other authors. Two groups computed the HRTF for frequency range from 100 Hz to 15 kHz using conventional BEM on substantially non-uniform meshes containing 21 693 and 12 054 panels, respectively, and compared their results.[14] The results for the "gain" in decibels for two incident angles were scanned from the cited paper and renormalized to obtain true HRTFs. As the HRTF is plotted using a logarithmic scale (decibels), any renormal-

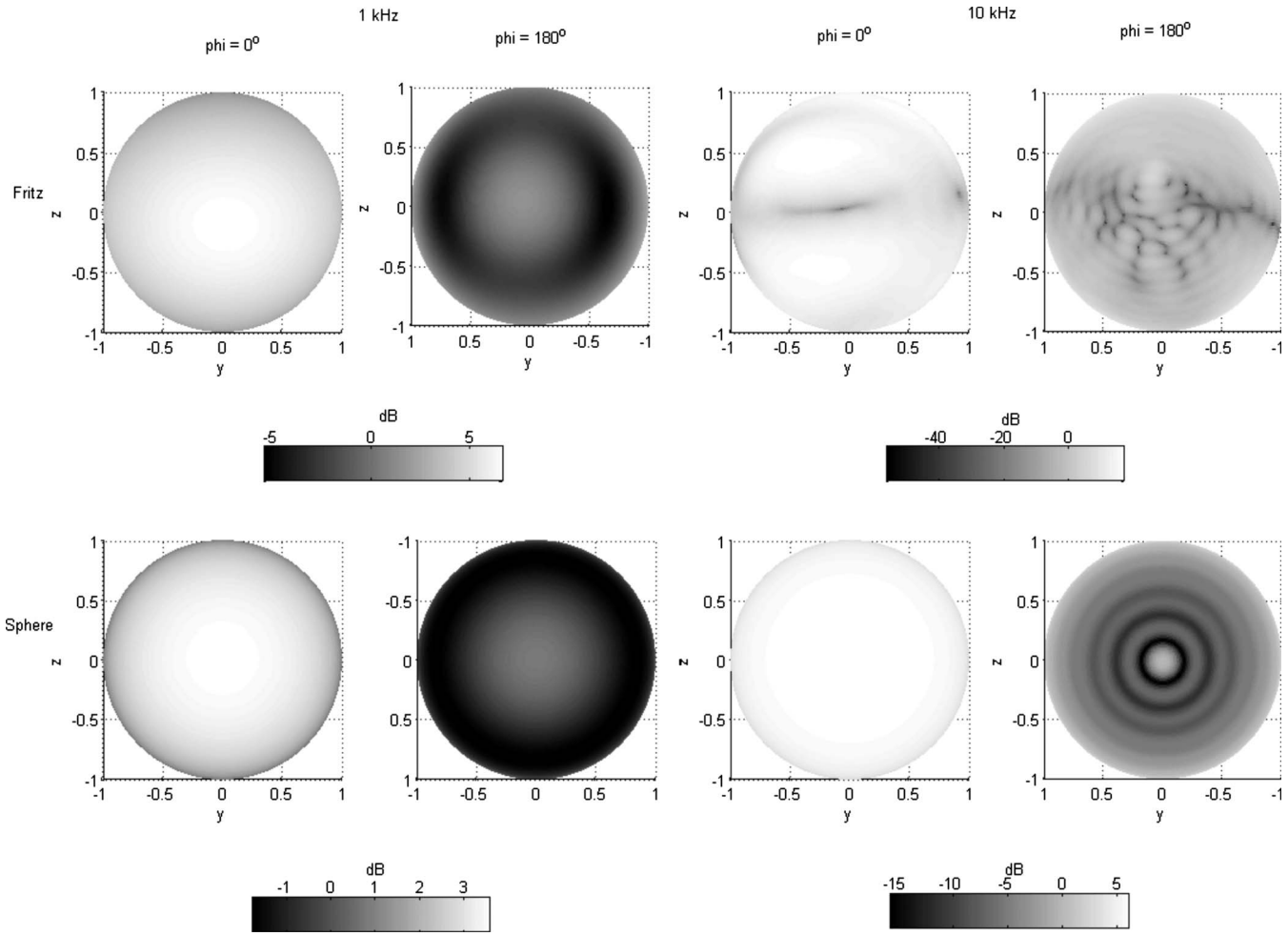


FIG. 2. The plane-wave HRTF ($20\log_{10}|H^{(\mathrm{pw})}|$) computed for the Fritz model (the top row) and analytical for a sphere of radius $a=a_{\min}=0.0705$ m (the bottom row) at frequencies 1.034 and 9.991 kHz. The HRTF is plotted as a function of spherical angles θ and φ [see Eq. (1)] on a unit sphere. The center of the right ear is located at the $\theta=90°$ and $\varphi=0°$. The views corresponding to $\varphi=0°$ are from the side of the right ear, while $\varphi=180°$ are from the side of the left ear. Axes (x,y,z) are oriented as shown in Fig. 1.

ization results in a shift of graphs along the y-axis. Since at low frequencies we have $|H^{(\mathrm{pw})}|\to 1$, or 0 dB, we shifted the cited results to this level at the lowest frequency (100 Hz).

Figure 4 shows comparison between the present computations and that of Ref. 14. It is seen that the computations reproduce the patterns in the HRTF. Quantitatively a good agreement is observed for the range of frequencies below 5 kHz, while for larger frequencies some mismatch takes place. There can be several reasons for this mismatch (differences in meshes and methods and accuracy, slight misalignment of the model [e.g., our measurement point was located actually at angle $\theta\approx 89°$ instead of ideal $\theta=90°$]). The comparison shows that the HRTF in all cases is computed well enough, and the difference in computations is smaller than usual difference between experiments and computations.

6. Comparison with experiments

Two sets of experimental data for the Fritz model were available to the authors via their own measurements, using both a direct and a reciprocal set-up. These measurements were conducted as a part of the "Club Fritz" activity.[27] Comparisons with both direct and reciprocal experiments provided approximately the same quality of agreement, so only data on the direct measurements which were conducted with a denser sampling are reported below for the sake of space.

In these experiments, the impulse response function was measured at both ears using external microphones and then its spectrum was analyzed to determine the HRTF for 823 directions. The signal generation/processing technique used windowing, so the experimental data for frequencies above 14 kHz are not reliable and frequencies above 16 kHz are completely suppressed. Also, as is the case with most HRTF measurements (see, e.g., Ref. 28 for a discussion) the experimental spectra were not reliable for low frequencies below 700 Hz. So for comparisons with computations only experimental HRTF for the range of frequencies from 700 Hz to 14 kHz was used.

We compared experiments and computations for all measurement locations, and several typical comparisons are shown in Fig. 5, where all non-normalized experimental data were shifted by 12 dB to provide consistent magnitudes with computations. A match in the patterns produced by the experiments and computations is clearly seen, while a quantitative discrepancy is also observed. Such a discrepancy may arise due to several factors, including modeling issues and

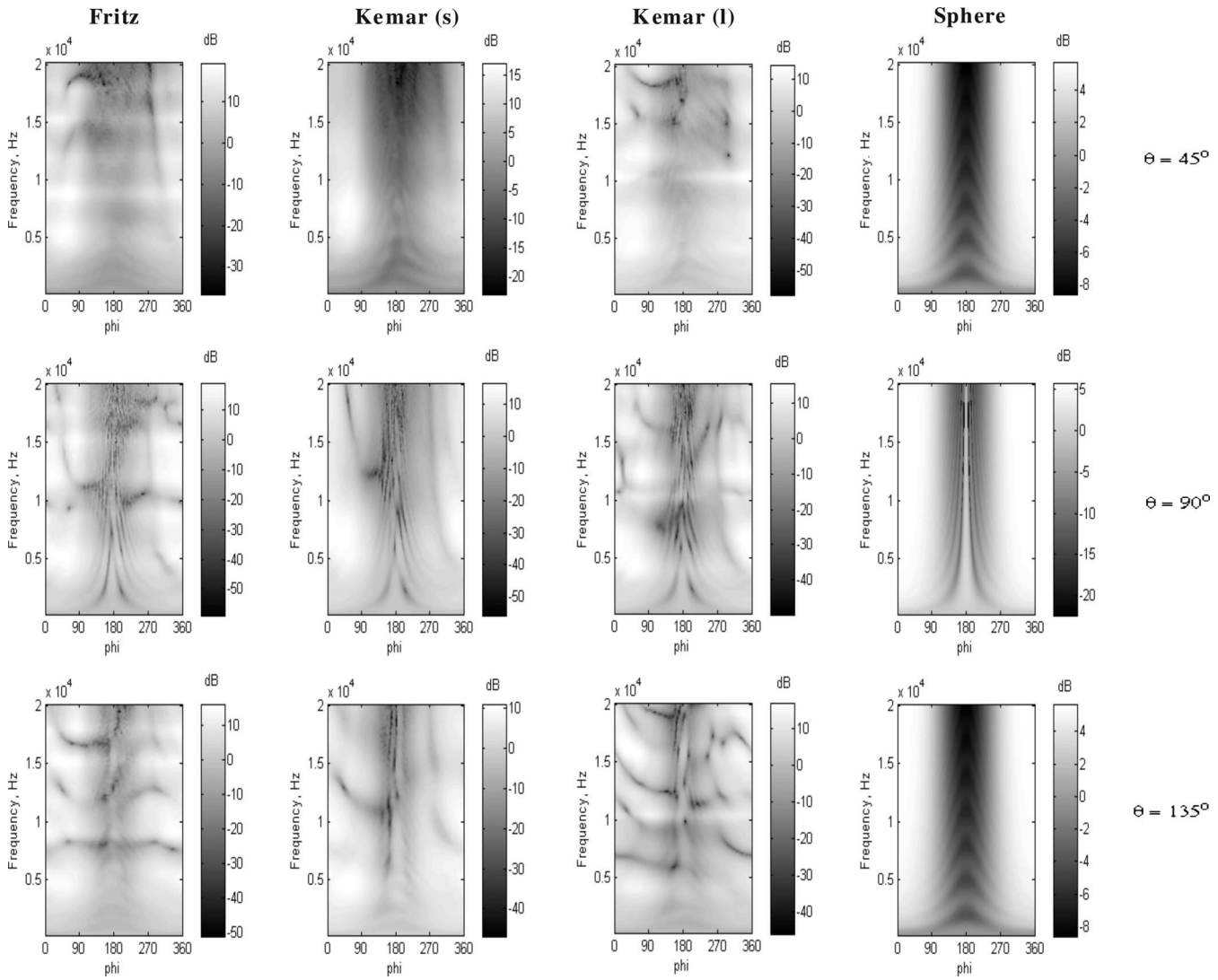


FIG. 3. The plane-wave HRTF ($20\log_{10}|H^{(\mathrm{pw})}|$) for the right ear ($\theta$=90°, φ=0°) computed for Fritz (the left column) and Kemar heads (the left center column for the "small pinnae" model and the right center column for the "large pinnae" model) and analytical for a sphere of radius $a=a_{\min}$=0.0705 m (the right column) at fixed spherical angles θ for the range of audible frequencies and spherical angles φ [see Eq. (1) and Fig. 1 for definitions].

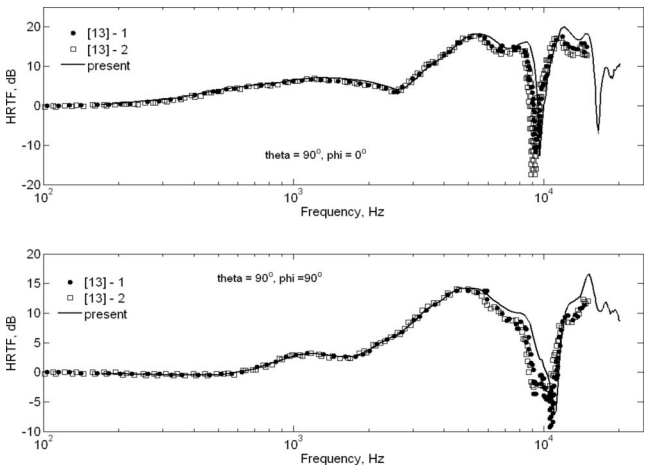


FIG. 4. Comparison of the plane-wave HRTFs ($20\log_{10}|H^{(\mathrm{pw})}|$) for two incident angles computed for the Fritz model by two groups and reported in Ref. 14 and using the present method. Original normalized data of Greff and Katz are renormalized to match the HRTF asymptotics at zero frequency ($|H^{(\mathrm{pw})}|\to 1$).

experimental errors. The latter, for example, include imprecision in recording of the tilt in the model position, which may result in a few degrees of inconsistency in the directions for which the HRTF was measured and computed. To evaluate such a type of the error, we plotted in Fig. 5 by the dashed lines measurements for the left ear for which directions were modified to get the "right ear data from the left ear measurements." This was easy to do since the 823 directions were symmetric with respect to the azimuth φ=90°, and the same set could be used for evaluation of the response of both ears (pairs $\varphi=90°\pm\alpha$ with the same θ were swapped). In the case of perfect symmetry and ideal measurements the actual data for the right ear and derived from the data for the left ear should coincide for this symmetric model. However, Fig. 5 shows discrepancy of these data, which can be used to judge the experimental errors and model misalignment. It is seen then that in some cases inconsistency of the experimental data is of the same order of magnitude as the inconsistency between the experiment and computations, while in some other cases the experimental error is smaller than between the experiments and computations (e.g., 1–2 kHz

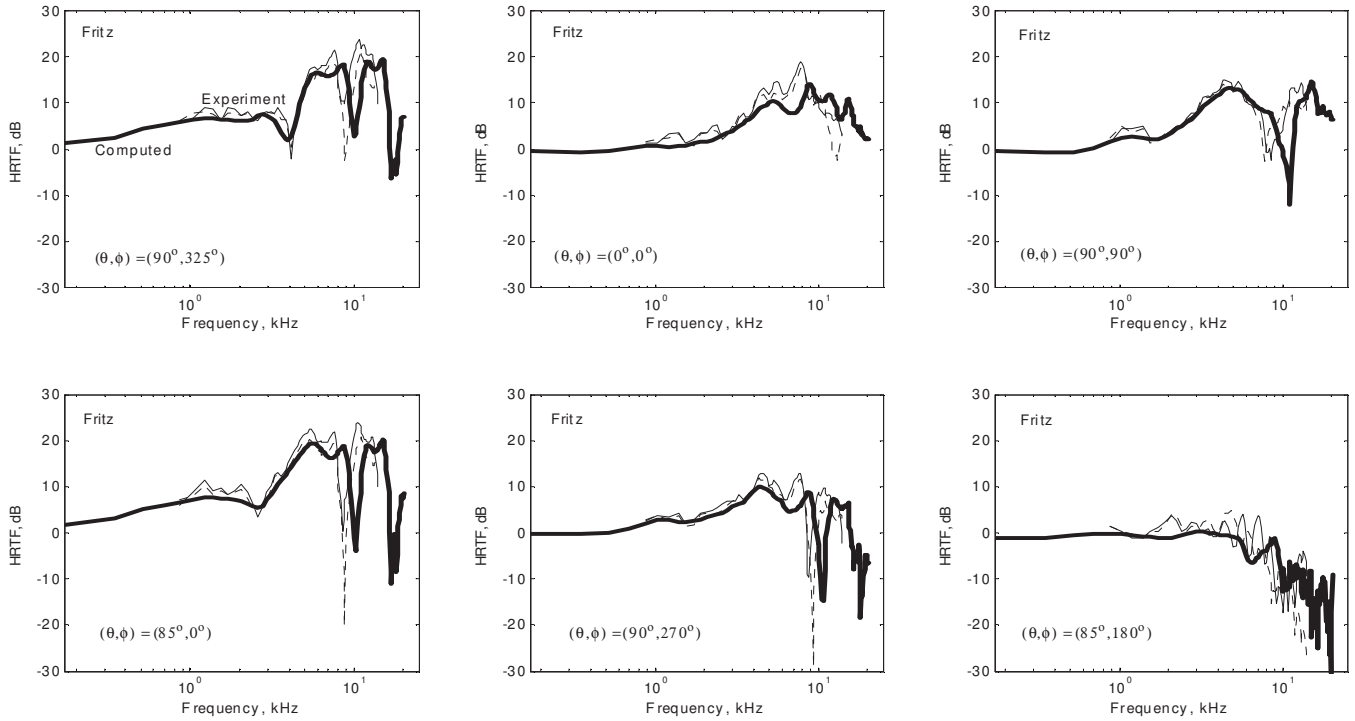


FIG. 5. Comparison of the computed and experimental HRTFs for the Fritz model for several angles. Computed data for the right ear are shown by the thick lines. The thin solid lines show experimental data for the right ear, while the dotted lines show data for the right ear derived from the experimental data for the left ear assuming symmetry of the model.

shift in the HRTF notches), which may be explained, e.g., by the differences in the computational and actual shapes. Figure 6 shows the data measured and computed at a constant θ=90° elevation (coronal plane through the ears) for the frequency range of validity of the experimental data. Experimental data for θ=90° measured at 58 azimuth angles φ were used for computations. This figure shows that an overall match exists between the computations and experiments, while quantitative discrepancy (in terms of notch locations) is clear for frequencies above 7–8 kHz.

B. Head-and-torso model computations

While computations for the head alone provide important information on the HRTF, the effect of torso on the HRTF is substantial. Modeling of this effect using a simple geometric approximation was performed in Refs. 28 and 29, while the shape influence of the head was considered in Ref. 30. In the present study we used a hybrid model consisting of detailed head shape of Kemar and torso modeled as an ellipsoid disjoint from the head. The parameters of the ellipsoid fitting the actual torso can be found in Ref. 28.

1. Meshes

The finest small pinnae mesh used for computations had 426 846 panels (445 276 for the finest "large pinnae mesh"). Both head-and-torso meshes satisfy the quality requirements for the BEM/FMM. Figure 1 illustrates the model and meshes used.

These meshes are suitable for computations of the HRTF in the entire audible range and provide not less than six elements per the shortest wavelength computed (frequency of 20.155 kHz). The dimensionless size of the computational domain for this frequency was kD=410, which is in the range of the BEM/FMM capabilities for personal computers (PCs). To reduce the computational cost we used two more meshes ("mid frequency" and "low frequency" meshes), which had 181 086 and 119 646 panels, respectively (for the small pinnae). These meshes had the same head meshing as the high frequency mesh, while we used coarser discretizations for the torso. The meshes were applied for frequency ranges 7.2–14.2 and below 7.2 kHz, respectively. In these ranges the head mesh certainly satisfied six elements/per wavelength criteria, while the torso discretizations were a bit rougher, but anyway not less than 5.5 elements/per wavelength.

2. Performance

Adding the torso substantially increases the computation time, as the size of the computational domain increases. The total time for the BEM solution was slightly below 70 h and required about 8 Gbyte RAM at the peak for the highest frequency.

3. Effect of torso

As a result of larger model size, plus the fact that the spectral representations can be applied to the region outside a large sphere concentric with the head, which includes the full torso, the number of harmonics for the head+torso model was substantially larger than that for the head alone. We found that for all frequencies several low frequency spatial modes of models with and without the torso match almost perfectly. The difference arises at high frequency spatial

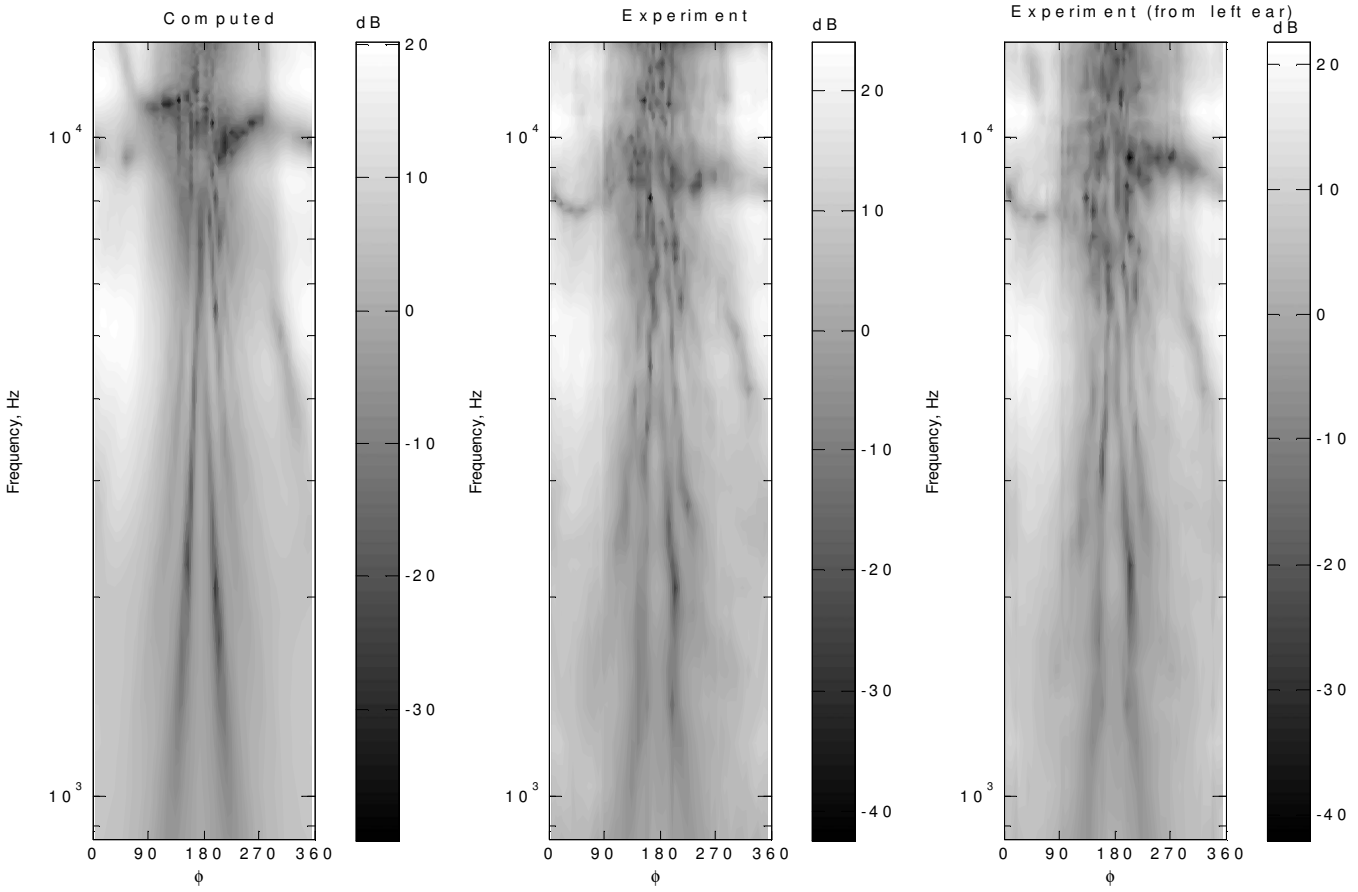


FIG. 6. Computed and measured HRTFs (in decibels) for the Fritz model (right ear) at angle $\theta=90°$ for frequency range 800 Hz–14 kHz. The middle plot is obtained from the right ear measurements, while the right plot is obtained from the left ear measurements by applying symmetry.

modes starting from some value dependent on the frequency of the acoustic field. The higher the acoustic frequency the higher is this starting value. So, e.g., at 500 Hz spectral differences are observable for modes with $n \geqslant 3$, while for 5 kHz the difference is visible for $n \gtrsim 10$ and for 15 kHz for $n \gtrsim 25$. The high frequency spatial modes in the presence of torso have a considerable magnitude in a wider range.

Figure 7 shows the plane-wave HRTF. It is seen that the presence of torso results in interference patterns due to reflections from the torso, which spatial frequency is higher at higher acoustic frequencies. According to the spectral analysis at high acoustic frequencies the torso does not affect low frequency spatial modes, while high frequency spatial component shows up as a wave reflected from the torso, which is superposed with the HRTF for the head alone. At low frequencies when the wavelength is comparable with the head size the torso affects low frequency spatial components, which create effects of totally different spatial dependence of the HRTF.

Figure 8 represents the HRTF along the centerlines when the azimuth varies at constant elevation ($\theta=90°$) and the elevation varies at constant azimuth ($\varphi=0°$, $\delta=90°-\theta$ and $\varphi=180°$, $\delta=90°+\theta$; angle δ is the same as in Ref. 28). The effect of torso is seen on both pictures, while for the latter case especially clear characteristic pattern for angles $180° < \delta < 270°$ is visible. This pattern is similar to that measured for the snowman model and the Kemar model without ears reported in Ref. 28 at frequencies below 5 kHz.

4. Comparisons with experimental data

Several sets of experimental data from different researchers were compared with the computed data for Kemar. The basic conclusion we came to is that computations for the small pinnae model do not agree well with the experiments with the small pinnae at frequencies above 6 kHz or so, while computations for the large pinnae agree good enough with the large pinnae experiments up to 20 kHz. The major reason for this we see in the ways how the ear meshes were produced. In the former case that was done via the laser scans, which loses information on the depth and so the mesh has low fidelity to the real shape. In the latter case the CT-scans were used for mesh generation, which provide much better information on the features of the real shape. The importance for the use of quality scans for meshes and effect on the computation was also discussed in Ref. 9.

The comparisons presented below use experimental data from the database of HRTFs that was presented in Ref. 3, and which is referred to as the "CIPIC" database in the literature. All the data there are standardized in terms of measurement technique and data are presented in the hoop coordinates φ_{hoop} (azimuth) and θ_{hoop} (elevation). These angles are different from the spherical polar angles we used since in

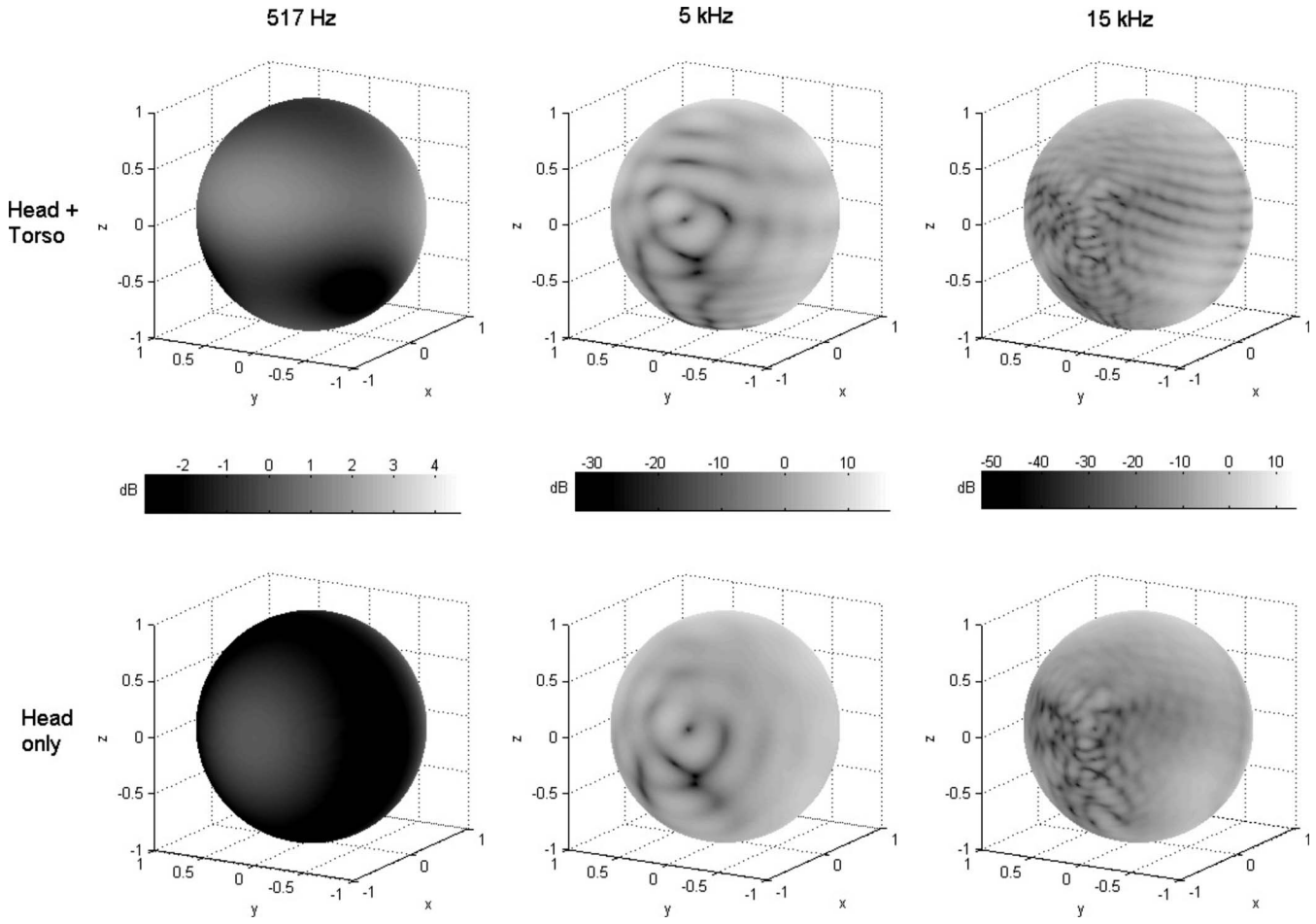


FIG. 7. The plane-wave right ear HRTF ($20\log_{10}|H^{(\mathrm{pw})}|$) computed for the "small pinnae" Kemar with torso (the top row) and without torso (the bottom row) at different frequencies. The HRTF is plotted as a function of spherical angles θ and φ [see Eq. (1)] on a unit sphere. The center of the right ear is located at the $\theta=90°$ and $\varphi=0°$ ($x=1$, $y=0$, $z=0$). The views are from the side of the left ear. Axes (x,y,z) are oriented as shown in Fig. 1.

the experiments the poles of spherical coordinates were located at the ear positions. Relation between the angles (φ,θ) and $(\varphi_{\mathrm{hoop}},\theta_{\mathrm{hoop}})$ follows from Eq. (1) and expression of the Cartesian coordinates via the hoop coordinates

$$\mathbf{r}=(x,y,z)$$
$$=r(\sin\varphi_{\mathrm{hoop}},\cos\varphi_{\mathrm{hoop}}\cos\theta_{\mathrm{hoop}},\cos\varphi_{\mathrm{hoop}}\sin\theta_{\mathrm{hoop}}),$$
$$-\pi/2<\theta_{\mathrm{hoop}}\leqslant 3\pi/2,\quad -\pi/2\leqslant\varphi_{\mathrm{hoop}}\leqslant\pi/2,\quad r=|\mathbf{r}|, \tag{47}$$

so the right ear location, for which all comparisons are made, corresponds to the pole $\varphi_{\mathrm{hoop}}=\pi/2$.

Figure 9 shows comparison of the computation (plane-wave HRTF) and experiments for the frequency range 0.5–20 kHz for frontal direction ($\varphi_{\mathrm{hoop}}=\theta_{\mathrm{hoop}}=0$). It is seen that for frequencies above 6 kHz or so there is a substantial deviation of the computations from experiments for the small pinnae case. On the other hand, computations for the large pinnae produce a HRTF pattern that matches the experiments as all maxima and minima of the frequency dependence are reproduced. Moreover, a good quantitative match is observed for frequencies up to 10 kHz and for the frequency and the depth of the strongest notch. It is also interesting that computations for head alone quantitatively are in better agreement with experiments for frequencies above 10 kHz, while computations with the torso agree only qualitatively (the experimental data on this figure were normalized for 3 dB shift to match maximum value of the HRTF for the large pinnae case). This shows that to reproduce effect of torso quantitatively a better discretization of the torso than a simple fitting ellipsoid should be used. Such a mesh was unavailable to us.

Figure 10 supports these observations for the entire range of azimuths at fixed elevation and for all elevations at fixed azimuths. It is clear that computations for the large pinnae match experiments in terms of the notch structure (dark filamentary patterns), while experiments for the small pinnae show more complex notch structure than the computed one. The effect of torso is well seen on the charts plotted for the fixed azimuth. Both experiments and head-and-torso computations show the ripples due to torso reflections, which are not present in the computations with head alone. The computed and experimental ripple patterns and ripple frequencies agree well, while as we noted above the quantitative agreement for a given point cannot be very good, which happens due to some ripple phase mismatch and slight overall distortion of the ripple pattern due to approximate torso shape modeling.

Figure 11 shows comparisons of computed (head and torso) and experimental data for three fixed frequencies

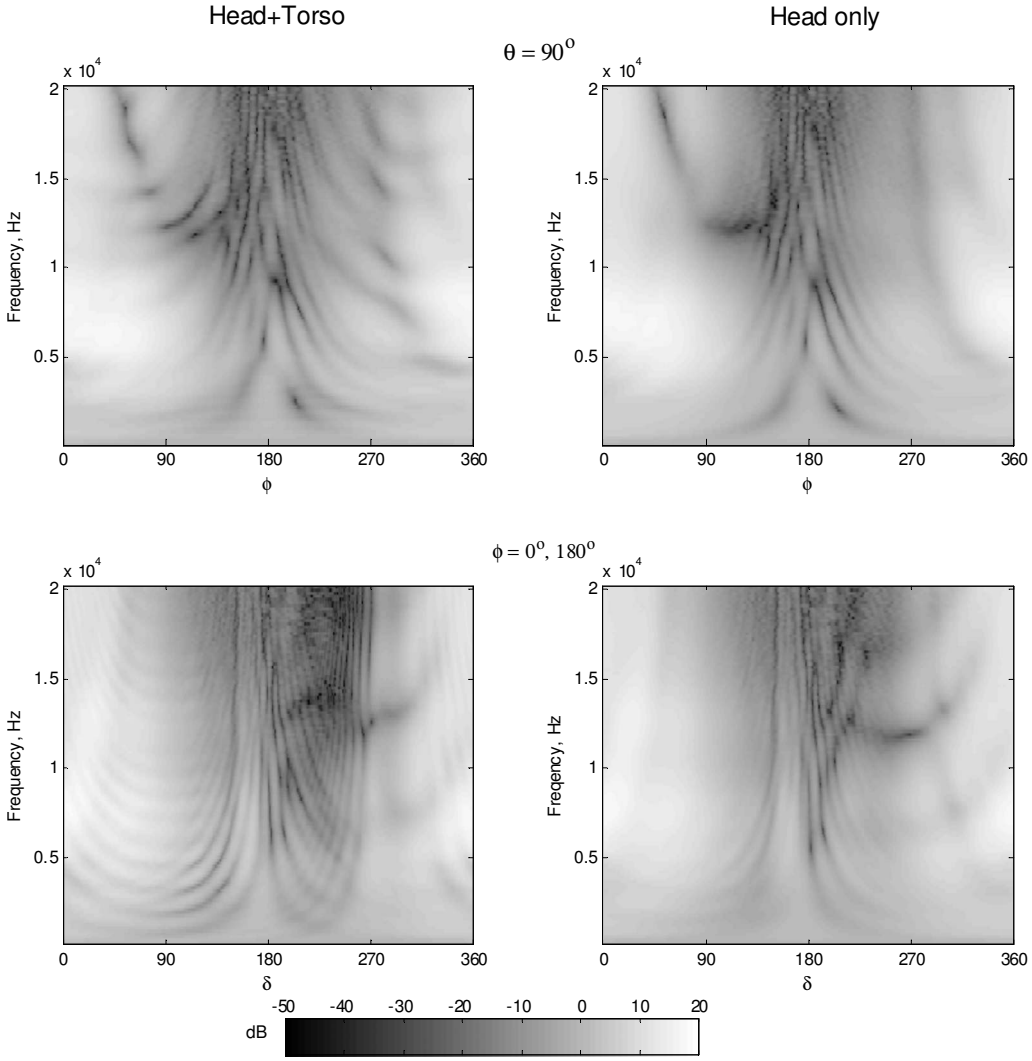


FIG. 8. The plane-wave right ear HRTF ($20\log_{10}|H^{(\mathrm{pw})}|$) computed for the "small pinnae" Kemar with torso (the left column) and without torso (the right column) for fixed elevation (θ=90°, the top row) and for the centerline going through the ears and the top of the head (the bottom row). In the latter case angle δ is measured from the right ear; δ=(90°−θ) mod 360° for φ=0° and δ=90°+θ for φ=180° ($0 \leqslant \theta \leqslant 180°$).

(about 5, 14, and 19 kHz) and all measured azimuths and elevations, which gives an idea on spatial dependence of the HRTF for a given frequency. Again we can see consistency of the computed and experimental data for the "large pinnae case," while such consistency is observed for the small pinnae case only at frequency 5 kHz. At larger frequencies the latter case shows a different notch structure. We did not plot on this graph the computed data for the head alone, which are similar to computations taking into account the presence of the torso in terms of notch locations, but do not have reflection ripples. Such ripples are clearly seen in experiments, and the ellipsoid torso approximation reproduces this structure good enough.

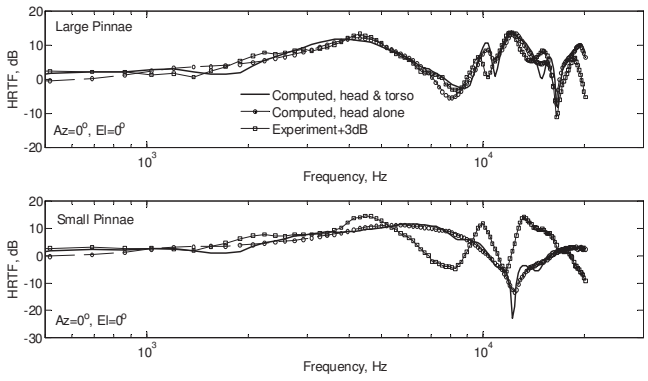


FIG. 9. Comparison of the computed HRTF for head and torso, head alone, and CIPIC experiments (the right ear) for the frontal direction (azimuth =0° and elevation=0°). The top graph corresponds to the "large pinnae" model, while the bottom graph shows comparison for the "small pinnae" model. The experimental data are shifted up by 3 dB.

VI. CONCLUSIONS

The main purpose of the present study was development and testing of an efficient method for computation of the HRTF over the audible frequency range. We believe that this objective was achieved. Indeed, comparing our results with those of the direct boundary element methods, the present FMM accelerated BEM shows a superior performance and is suitable for computation of the HRTF in the full audible range including not only the head but the head-and-torso models. The method appeared to be stable and produces good results for meshes containing hundreds of thousands of

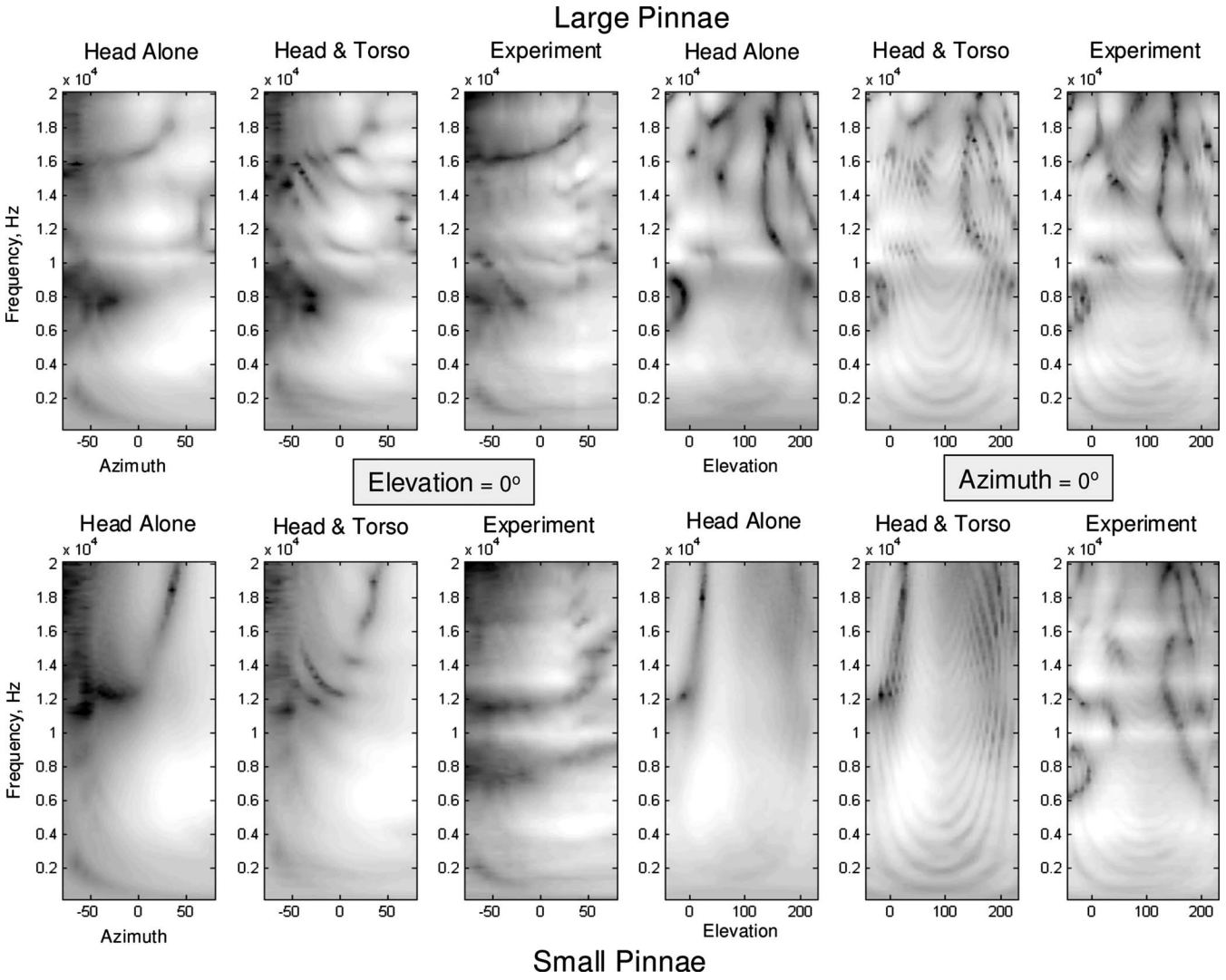


FIG. 10. Comparison of computed (head alone and head and torso) and CIPIC experimental HRTF data (in decibels) for the entire range of audible frequencies for fixed elevation (0°) and all measured azimuths ($-80° \leqslant \varphi_{\text{hoop}} \leqslant 80°$) (the six pictures on the left) and for fixed azimuth (0°) and all measured elevations ($-45° \leqslant \theta_{\text{hoop}} \leqslant 230.625°$) (the six pictures on the right). The top row shows data for the "large pinnae" case and the bottom row shows data for the "small pinnae" case. Experimental data are taken for the right ear. The dynamic range varies from about −60 dB (the darkest spots) to about +10 dB (the brightest areas).

boundary elements, which provides accurate computation and representation of the HRTF at frequencies as high as 20 kHz.

Representation of the HRTF via its discrete spherical spectrum, suggested in this paper (and for fitting of experimental data in Ref. 31), is rather natural. First, this is a relatively compact representation, second, it provides spectral accuracy consistent with the accuracy of computations, and third, it enables fast evaluation of the HRTF for any set of directions. In particular, this is important for rotation transforms, range dependence, and corrections of the HRTF due to body motion in the use of the HRTF in virtual reality.

Comparisons with experiments for the head alone and the head and torso showed that a good agreement can be obtained in the entire audible range. However, to achieve this goal several things should be handled with care. This relates to accurate scans of pinnae (laser scans may provide loss of important features, while the CT-scans are much more accurate though difficult to acquire), careful reproduction of experimental environment [e.g., positioning of the head model, modeling of torso (if present), etc.], and quality computational meshes.

Of course, the availability of a computational tool for the HRTF is only a first step. These should now be used to explore the dependence of HRTFs on shape, to compute HRTFs for different animals/individuals, and to study the influence of things such as clothing and hair. To aid in this effort it would be essential to combine the different research tools such as those developed in this study (the meshing procedures, the fast multipole software, and the software for analysis of the spherical spectra) into a usable tool. Then this tool should be used for computation of the HRTF for other meshes, modeling real human head/torso interactions for comparisons with data in available HRTF databases, and conducting experiments on perceptual experiences with the computed HRTFs. We were able to compute full range HRTF for tens of hours on a PC. However, for parametric studies, further accelerations are needed. These may be provided by

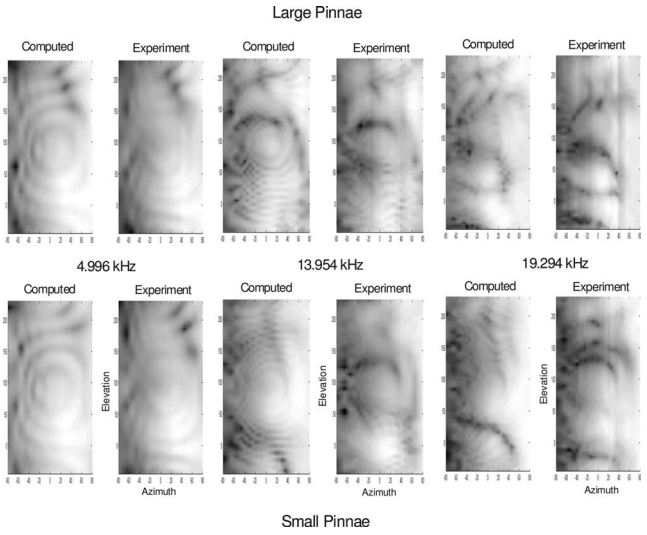


FIG. 11. Comparisons of the computed (head and torso) and CIPIC experimental HRTF data (in decibels) for three fixed frequencies (5.2, 14.1, and 19.2 kHz) and all azimuths and elevations measured in the experiments. The upper row shows the "large pinnae" case, while the bottom row shows the "small pinnae" case. The dynamic range for the highest frequency case varies from about −60 dB (the darkest spots) to about +10 dB (the brightest areas).

supercomputers, but we believe that the method can be further accelerated in the personal computing environments being made available, e.g., using graphic processors. Indeed, some promising results are obtained for the FMM for the Laplace equation on graphical processors,[32] while the method should be developed further for the Helmholtz equation. Also we note that research is needed on the use of HRTF spherical spectra in applications.

ACKNOWLEDGMENTS

These computations are a result of a long term project that has been performed over the past 8 years at the Perceptual Interfaces and Reality Laboratory at the University of Maryland. At various stages our research has been supported by the NSF, DARPA, the state of Maryland, and Fantalgo, LLC; all are thanked for their support. We thank Professor R. O. Duda and Professor V. R. Algazi of the University of California, Davis for several discussions on the HRTF, for providing the CIPIC database, and for information on it. They are also thanked for providing the CT-scans of the Kemar large pinna. We would also like to thank Dr. Kexue Liu for creating the pinna mesh, Dr. Elena Grassi for the measurements of the Fritz HRTFs, Dr. Yuvi Kahana for the original Kemar mesh, and Dr. Brian Katz for the original Fritz mesh. We would like to acknowledge the use of software provided by Fantalgo, LLC for the fast multipole boundary element method.

[1]D. W. Batteau, "The role of the pinna in human localization," Proc. R. Soc. London, Ser. B **168**, 158–180 (1967).
[2]W. M. Hartmann, "How we localize sound," Phys. Today **52**(11), 24–29 (1999).
[3]V. R. Algazi, R. O. Duda, D. M. Thompson, and C. Avendano, "The CIPIC HRTF database," in Proceedings of the 2001 IEEE WASPAA, New Paltz, NY (2001), pp. 21–24.
[4]W. Gardner and K. Martin, "HRTF measurements of a KEMAR," J. Acoust. Soc. Am. **97**, 3907–3908 (1995).
[5]D. N. Zotkin, R. Duraiswami, E. Grassi, and N. A. Gumerov, "Fast head-related transfer function measurement via reciprocity," J. Acoust. Soc. Am. **120**, 2202–2215 (2006).
[6]Y. Kahana, P. A. Nelson, M. Petyt, and S. Choi, "Numerical modelling of the transfer functions of a dummy-head and of the external ear," in AES 16th International Conference on Spatial Sound Reproduction (1999), pp. 1–6.
[7]B. F. G. Katz, "Boundary element method calculation of individual head-related transfer function. I. Rigid model calculation," J. Acoust. Soc. Am. **110**, 2440–2448 (2001).
[8]Y. Tao, A. I. Tew, and S. J. Porter, "The differential pressure synthesis method for efficient acoustic pressure estimation," J. Audio Eng. Soc. **51**, 647–656 (2003).
[9]M. Otani and S. Ise, "Fast calculation system specialized for head-related transfer function based on boundary element method," J. Acoust. Soc. Am. **119**, 2589–2598 (2006).
[10]T. Walsh, L. Demkowicz, and R. Charles, "Boundary element modeling of the external human auditory system," J. Acoust. Soc. Am. **115**, 1033–1043 (2004).
[11]J. L. Aroyan, "Three-dimensional modeling of hearing in delphinus delphis," J. Acoust. Soc. Am. **110**, 3305–3318 (2001).
[12]F. De Mey, J. Reijniers, and H. Peremans, "Simulated head related transfer function of the phyllostomid bat Phyllostomus Discolor," J. Acoust. Soc. Am. **124**, 2123–2132 (2008).
[13]T. Xiao and Q.-H. Liu, "Finite difference computation of head-related transfer function for human hearing," J. Acoust. Soc. Am. **113**, 2434–2441 (2003).
[14]R. Greff and B. F. G. Katz, "Round robin comparison of HRTF simulation results: Preliminary results," Audio Engineering Society Convention Paper, the 123rd Convention (2007).
[15]N. A. Gumerov and R. Duraiswami, "A broadband fast multipole accelerated boundary element method for the 3D Helmholtz equation," J. Acoust. Soc. Am. **125**, 191–205 (2009).
[16]P. M. Morse and K. U. Ingard, *Theoretical Acoustics* (Princeton University Press, Princeton, NJ, 1968).
[17]M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions* (National Bureau of Standards, Washington, DC, 1964).
[18]N. A. Gumerov and R. Duraiswami, *Fast Multipole Methods for the Helmholtz Equation in Three Dimensions* (Elsevier, Oxford, 2004).
[19]A. J. Burton and G. F. Miller, "The application of the integral equation methods to the numerical solution of some exterior boundary-value problems," Proc. R. Soc. London, Ser. A **323**, 201–210 (1971).
[20]W. C. Chew, J. M. Jin, E. Michelssen, and J. Song, *Fast and Efficient Algorithms in Computational Electromagnetics* (Artech House, Boston, MA, 2001).
[21]N. A. Gumerov and R. Duraiswami, "A scalar potential formulation and translation theory for the time-harmonic Maxwell equations," J. Comput. Phys. **225**, 206–236 (2007).
[22]C. F. Bohren and D. R. Huffman, *Absorption and Scattering of Light by Small Particles* (Wiley, New York, 1983).
[23]A. E. O'Donovan, N. A. Gumerov, and R. Duraiswami, "Good meshes for the fast multipole accelerated boundary element method via radial basis function interpolation and relaxation," University of Maryland Department of Computer Scince Technical Report No. CS-TR-4937. Also UMIACS-TR-2009-07 (2009).
[24]N. A. Gumerov and R. Duraiswami, "Fast radial basis function interpolation via preconditioned Krylov iteration," SIAM J. Sci. Comput. (USA) **29**, 1876–1899 (2007).
[25]K. Liu, R. Duraiswami, and L. S. Davis, "A simple and optimal energy surface reconstruction algorithm from volumetric data," Department of Computer Science Technical Report CS-TR-4441 and UMIACS-TR-2003-11, University of Maryland, College Park, MD, 2003.
[26]Y. Saad, "A flexible inner-outer preconditioned GMRES algorithm," SIAM J. Sci. Comput. (USA) **14**, 461–469 (1993).
[27]B. F. G. Katz and D. R. Begault, "Round robin comparison of HRTF measurement systems: Preliminary results," in 19th International Congress on Acoustics (2007).
[28]V. R. Algazi, R. O. Duda, R. Duraiswami, N. A. Gumerov, and Z. Tang, "Approximating the head-related transfer function using simple geometric models of the head and torso," J. Acoust. Soc. Am. **112**, 2053–2064 (2002).
[29]N. A. Gumerov, R. Duraiswami, and Z. Tang, "Numerical study of the

influence of the torso on the HRTF," in Proceedings of the IEEE ICASSP, Orlando, FL, Vol. **2** (2002), pp. 1965–1968.
[30]Y. Tao, A. I. Tew, and S. J. Porter, "A study on head-shape simplification using spherical harmonics for HRTF computation at low frequencies," J. Audio Eng. Soc. **51**, 799–805 (2003).
[31]M. J. Evans, J. A. S. Angus, and A. I. Tew, "Analyzing head-related transfer function measurements using surface spherical harmonics," J. Acoust. Soc. Am. **104**, 2400–2411 (1998).
[32]N. A. Gumerov and R. Duraiswami, "Fast multipole methods on graphics processors," J. Comput. Phys. **227**, 8290–8313 (2008).

Prediction of binaural speech intelligibility against noise in rooms

Mathieu Lavandier[a] and John F. Culling
School of Psychology, Cardiff University, Tower Building, Park Place, Cardiff CF10 3AT, United Kingdom

(Received 30 October 2008; revised 3 November 2009; accepted 5 November 2009)

In the presence of competing speech or noise, reverberation degrades speech intelligibility not only by its direct effect on the target but also by affecting the interferer. Two experiments were designed to validate a method for predicting the loss of intelligibility associated with this latter effect. Speech reception thresholds were measured under headphones, using spatially separated target sentences and speech-shaped noise interferers simulated in virtual rooms. To investigate the effect of reverberation on the interferer unambiguously, the target was always anechoic. The interferer was placed in rooms with different sizes and absorptions, and at different distances and azimuths from the listener. The interaural coherence of the interferer did not fully predict the effect of reverberation. The azimuth separation of the sources and the coloration introduced by the room also had to be taken into account. The binaural effects were modeled by computing the binaural masking level differences in the studied configurations, the monaural effects were predicted from the excitation pattern of the noises, and speech intelligibility index weightings were applied to both. These parameters were all calculated from the room impulse responses convolved with noise. A 0.95–0.97 correlation was obtained between the speech reception thresholds and their predicted value. © *2010 Acoustical Society of America.* [DOI: 10.1121/1.3268612]

PACS number(s): 43.66.Pn, 43.66.Dc, 43.55.Hy, 43.71.Gv [RLF] Pages: 387–399

I. INTRODUCTION

Room acoustical measurements such as the useful-to-detrimental ratio (Bradley, 1986; Bradley *et al.*, 1999) or the speech transmission index (Houtgast and Steeneken, 1985) already exist to predict the potential loss of intelligibility in a room, when the original speech signal is being mixed with its delayed versions associated with the sound reflections on the room boundaries. These measurements can take into account a background ambient noise, but they were not designed to predict intelligibility against localized interferers for which signal statistics at the ears depend on the source and listener positions within the room. When listeners need to separate a target speech from a localized interferer, reverberation impairs intelligibility not only by its direct effect on the target but also by affecting the binaural processing of the interferer (Lavandier and Culling, 2007). The corresponding loss of intelligibility appears at lower levels of reverberation and thus occurs more readily than the loss of intelligibility associated with the smearing effect of reverberation on the target (Lavandier and Culling, 2008).

In anechoic situations, the azimuth separation of sources is a well-established cue helping the segregation of competing sounds (Hawley *et al.*, 2004; Plomp, 1976). The spatial unmasking associated with this azimuth separation arises from two cues (Bronkhorst and Plomp, 1988): head shadow and binaural interaction. When the interferer is moved around the listener's head, its sound level is reduced at the ear which is in the acoustic shadow of the head. The target-to-interferer level ratio is improved at this "better" ear, and this improvement helps the listener to understand the target. Binaural interaction relies on interaural time delays (ITDs). The azimuth separation of interferer and target induces a difference in ITDs for these two sources, which facilitates their segregation. For example, the equalization-cancellation (E-C) theory predicts that the auditory system will be able to "cancel" an interferer if it has a different ITD from the target, thus improving the internal target-to-interferer level ratio (Durlach, 1972).

Levitt and Rabiner (1967b) proposed a model describing the release from masking associated with binaural interaction. Speech intelligibility in noise is predicted from the computation of the articulation index (AI) (Kryter, 1962). Binaural interaction is taken into account by assuming that the effective speech-to-interference ratio in each frequency band is increased by the binaural masking level difference (BMLD) for pure tone detection in noise at the center frequency of the band, using BMLD predictions from Durlach (1963). Predictions based on this model were fairly consistent with previous data collected by the same authors (Levitt and Rabiner, 1967a). From this work, Zurek (1993) proposed a model describing spatial unmasking in anechoic situations. Better-ear listening is simulated by computing the speech-to-interference ratios at the two ears by frequency bands, and taking the better of the left and right ratios in each band. Binaural interaction is then taken into account by increasing the better-ear ratios by the size of the BMLD in each band, this BMLD being estimated for the given set of interaural parameters using a simplified expression proposed by Colburn (1977). The broadband prediction is computed as the AI-weighted sum of the resulting signal-to-noise ratios. The model predictions were compared with measured data from several studies. Predictions of the head-shadow advantage

[a]Author to whom correspondence should be addressed. Present address: Université de Lyon, Ecole Nationale des Travaux Publics de l'Etat, Département Génie Civil et Bâtiment (C.N.R.S.), Rue M. Audin, 69518 Vaulx-en-Velin Cedex, France. Electronic mail: mathieu.lavandier@entpe.fr

were generally larger than the measured effects, whereas binaural-interaction advantages were fairly well predicted.

Many experiments have shown that spatial unmasking is reduced by reverberation (Beutelmann and Brand, 2006; Culling *et al.*, 2003; Plomp, 1976). The sound reflections traveling around the listener reduce head shadow (Plomp, 1976), and reverberation impairs binaural interaction by decorrelating the interferer at the listener's ears (Lavandier and Culling, 2008). The models proposed by Levitt and Rabiner (1967b) and Zurek (1993) cannot be applied to reverberant situations, because they do not take into account the interaural coherence of the interferer. The interaural coherence of a source is the correlation of the sound waveform it produces at the two ears of the listener after interaural delay is taken into account. This coherence is reduced by sound reflections in rooms. Lavandier and Culling (2007, 2008) showed that the target intelligibility decreased when they decreased the interferer coherence, either by increasing the listener-interferer distance or making the room more reverberant. An E-C mechanism would be less effective against an interferer that is not perfectly correlated because a less correlated interferer cannot be fully equalized at the two ears, and hence cannot be fully canceled. As a result, there is more masking and lower speech intelligibility, as shown by previous binaural unmasking experiments (Licklider, 1948; Robinson and Jeffress, 1963).

Different approaches have been proposed to predict the effects of reverberation in multiple-source situations. Wijngaarden and Drullman (2008) extended the speech transmission index method to predict spatial unmasking effects. This approach offers the advantage of taking into account the influence of reverberation on the target. However, it also makes the initial assumption that the speech target is the only source of modulation in the signals reaching the listener's ears. The aim is to look for modulation to measure the transmission of speech. This approach does not offer any chance to be extended to modulated noise or speech interferers because the modulation is then coming from both target and interferer. Zurek *et al.* (2004) proposed a model predicting the detection of a narrow band noise target against a broadband noise interferer in rooms, which was later extended to predict speech intelligibility (Freyman and Zurek, 2008). The model is based on room statistics (surface area and average absorption coefficient of the room, assuming a perfectly diffuse reverberant sound field independent of the direct sound) and an E-C mechanism. Binaural detection of the narrow band noises was quite accurately predicted, even if some discrepancies remained. These discrepancies could be linked to the initial approximations inherent to the use of room statistics rather than room impulse responses, and of a fixed interaural correlation function in the E-C mechanism, independent of the position considered in the room, rather than the measured interaural coherence. Beutelmann and Brand (2006) developed a model predicting the intelligibility of a speech target against a noise interferer in rooms. The stimuli obtained after convolution by the binaural room impulse responses are first processed through an E-C stage which takes into account the deleterious effect of reverberation on binaural interaction and are then resynthesized. The speech intelligibility index (SII) method (ANSI, 1997) is then used to evaluate speech intelligibility in the given configuration. The E-C stage consists of a direct implementation of an E-C process which tests different delays and attenuations for the signals at the ears, and chooses those maximizing the target-to-interferer ratio. The agreement with listening test data was very good, with an overall correlation coefficient of 0.95 between measurement and prediction.

The aim of the present study was to validate another method which quantitatively predicts the loss of intelligibility associated with the impairment of binaural interaction in rooms. Like the model proposed by Beutelmann and Brand (2006), this method is based on the information contained in the room impulse responses. However, whereas Beutelmann and Brand (2006) were primarily concerned with differences between normal-hearing and hearing-impaired listeners in noise and reverberation, we looked for a more computationally efficient prediction method that could be applied in architectural acoustic software. We therefore abandoned the direct implementation of a complex E-C mechanism, in favor of the formula proposed by Culling *et al.* (2005) for the evaluation of BMLDs. This formula has been used to help explain intelligibility in anechoic configurations involving multiple interferers at various azimuths (Culling *et al.*, 2004). Lavandier and Culling (2007, 2008) interpreted the influence of reverberation on binaural interaction as being mediated mainly by the interaural coherence of the interferer, but their experiments were realized with a fixed azimuth separation of target and interferer. The formula used in the present study not only takes into account the dependence of BMLDs on the interaural coherence of the interferer but also on the difference of interaural phase difference between target and interferer. As a result, it seemed to be a good candidate to predict the dependence of binaural unmasking on the level of room reverberation as well as on the azimuth separation of sources. In order to validate the method, predictions were compared to psychoacoustical measurements involving several rooms and source configurations.

Our approach to investigate cocktail-parties in rooms consists of decomposing the various effects taking place simultaneously, in order to predict them individually, and eventually build a global prediction of intelligibility. Because this study was focused on the prediction of binaural interaction, we tried to eliminate any confounding effect associated with better-ear listening. Experiments have attempted to separate the roles of these two components in several ways. Bronkhorst and Plomp (1988) and Culling *et al.* (2004) employed signal-processing techniques in order to create stimuli with only the ITDs that underpin binaural interaction or only the interaural level differences (ILDs) that underpin better-ear listening. In the experiments presented in this paper, the head of the listener was not modeled in order to avoid frequency-dependent head shadow effects which could complicate the interpretation of the effect associated with binaural interaction. Some frequency-dependent ITD effects might have been excluded as well by the removal of the head, but, as discussed at the end of the paper, this limitation should not compromise the validity of the proposed prediction method. The effects of the room and the source-listener distance on

the broadband sound level of sources were also eliminated, by equalizing all stimuli in level independently at the left and right ears, removing broadband interaural level differences.

Speech reception thresholds (SRTs) (i.e., the level of the target compared to that of the interferer for 50% intelligibility of the target) were measured under headphones, using spatially separated targets and interferers simulated in virtual rooms. The simulation allowed variation in the positions of the listener and sources, of the size of the room, and of the absorption coefficient of the room boundaries. To investigate the effect of reverberation on the interferer unambiguously, the target was always anechoic. This experimental design is not realistic, as it implies listening simultaneously to two sources in rooms having different reverberation characteristics, but it offers the advantage of isolating the effect of reverberation on the interferer. The interferer was simulated to be in rooms with different sizes and absorptions, at different distances and azimuths from the listener. Reverberation reduces binaural interaction, but when speech interferers are involved, it impairs intelligibility also by affecting monaural speech segregation mechanisms (Lavandier and Culling, 2008). The sound reflections in the room can disrupt the segregation of competing sentences based on fundamental frequency differences (Culling *et al.*, 2003, 1994). They can also fill the potential silent periods in the speech interferers which otherwise allow one to hear the target better (Bronkhorst and Plomp, 1990; George *et al.*, 2008). To study the influence of reverberation on binaural interaction without mixing it with these additional effects, the experiments presented here used only continuous speech-shaped noise interferers. These continuous noises had no fundamental frequency (F0) and no modulation in their temporal envelope, so that source segregation based on F0 and "dip listening" was eliminated.

The different components of the proposed prediction method were tested in specifically designed experiments. Binaural interaction depends on the interaural phase of the sources and on the interaural coherence of the interferer. Experiment 1 tested whether the proposed method could quantitatively predict the influence of the interaural phase of sources on binaural interaction in rooms. SRTs were measured for four different azimuth separations of target and interferer, at different levels of reverberation. The method based on calculated BMLDs led to good predictions. Experiment 2 tested whether the method could predict the influence of the interaural coherence of the interferer. This coherence was varied in many different ways—changing the absorption and the size of the room, the distance between the listener and the interferer, and the asymmetry of the configuration. The corresponding SRTs were well described by the prediction method. Experiment 2 also showed the importance of taking into account the influence of room coloration on intelligibility.

II. GENERAL METHODS

A. SRT measurements

1. Stimuli

The same voice was used for the target in every experiment. The corpus of sentences was from the Harvard Sentence List (IEEE, 1969). The anechoic recordings of a male voice digitized at 20 kHz with 16-bit quantization were used as the basis of all stimuli. The sentences have low predictability, and each sentence contains five key words. For instance, one sentence was "TAKE the WINDING PATH to REACH the LAKE." The speech-shaped noise interferers were obtained by filtering Gaussian noises with a finite impulse response filter designed to match the speech long-term excitation pattern (Moore and Glasberg, 1983). These interferers all lasted longer than the longest target sentence.

The virtual rooms used in the experiments were simulated using a ray-tracing method (Allen and Berkley, 1979; Peterson, 1986), implemented in the IWAVE signal processing package (Culling, 1996). The absorption coefficients of the room internal surfaces were all set to the same value, and varied together when the reverberation condition was changed. The target was always anechoic, an absorption coefficient of 1 being used for the corresponding room. The listener was modeled as two ears with no head between them, simulated by omnidirectional microphones separated by 18 cm and suspended in space at 1.5 m from the floor. The interferer and target were placed at different positions in the room, also at 1.5 m from the floor. Binaural stimuli were produced by calculating the impulse responses between the source positions and each ear, and convolving the speech and noise samples with these impulse responses. When the interferer was further away from the listener than the target, the arrival time of all interferers was set below the one of the target by deleting part of the silent period at the beginning of their waveforms, to avoid having the target starting before the interferer when stimuli were played to the listeners.

Convolution by a room impulse response can change the sound level of a stimulus differently depending on the source position in the room and the ear considered (Bradley *et al.*, 1999). The root-mean-square (rms) powers of the left and right channels of the convolved stimuli were equalized independently. The broadband target-to-interferer level ratio was then fixed at the ears. Using stimuli equalized before the convolution would have placed control of the target-to-interferer level ratio at the emission of the sound sources. The influence of the room on the sound levels of interferer and target would then have been included in the measured SRTs. In order to interpret the room-induced binaural interaction changes on SRTs unambiguously, all stimuli were equalized after convolution by the room impulse responses.

2. Procedure

SRTs were measured using a 1-up/1-down adaptive threshold method (Plomp and Mimpen, 1979). For each SRT measurement, ten target sentences were presented one after another against the same noise interferer. The target-to-interferer level ratio was initially very low (−32 dB). On the first trial, listeners could either enter a transcript on a computer keyboard or replay the stimuli. If stimuli were replayed, the target level was increased by 4 dB. Stimuli had to be replayed until the target was loud enough to be judged partially intelligible. Listeners were instructed to attempt a transcript of this first target sentence when they believed that they could hear more than half the words of the sentence.

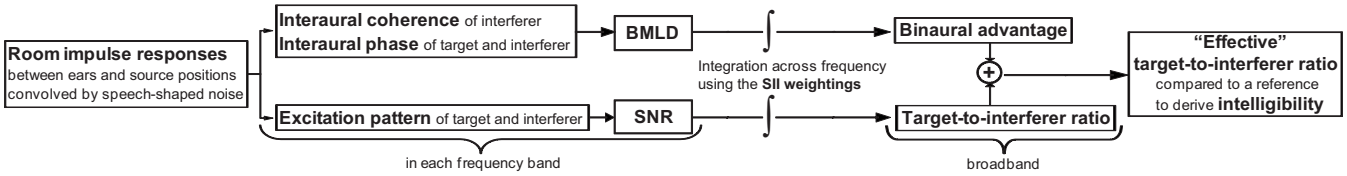


FIG. 1. Proposed method for predicting the intelligibility of a near-field speech target against a localized noise interferer in rooms.

Once the first transcript was entered, the correct transcript was displayed on the computer terminal, with the five key words in capitals. The listener self-marked the number of correct key words. Subsequent target sentences were presented only once, and self-marked in a similar manner. The target level was decreased by 2 dB if the listener correctly identified three or more of the five key words in the previous sentence, and otherwise increased by 2 dB. The SRT for a given condition was taken as the mean target-to-interferer level ratio on the last eight trials.

Each SRT measurement used a different set of ten target sentences and a different noise interferer. The session began with two practice runs using unprocessed stimuli, in order to familiarize listeners with the task. The following runs measured SRTs in each of the N tested conditions in a fixed order randomly chosen ($N=8$ in experiments 1a and 1b, $N=16$ in experiment 2). The order of the conditions was then rotated for successive listeners, while sentence materials remained in the same order. Each target sentence was thus presented to every listener in the same order and, across a group of N listeners, a complete rotation of conditions was achieved. Each experiment therefore used a multiple of N listeners. This procedure also ensured that each condition was presented in each serial position within the experimental session.

Signals were digitally mixed, D/A converted, and amplified using a 24-bit Edirol UA-20 sound card and an MTR HPA-2 headphone amplifier. They were presented to listeners over Sennheiser HD650 headphones in a single-walled IAC sound-attenuating booth within a sound-treated room. A computer terminal screen was visible outside the booth window. A keyboard was inside the booth to gather the transcripts of listeners.

3. Listeners

Listeners all reported normal hearing and English as their first language. They were undergraduate students, paid for their participation. None of them was familiar with the sentences used during the test. Each listener participated in only a single session of a given experiment.

B. Prediction method

Figure 1 presents the method we used to predict the intelligibility of a near-field speech target against a localized noise interferer in rooms. This method requires calculation of the interaural phase and excitation pattern of target and interferer, and of the interaural coherence of the interferer. These acoustical parameters were derived from the room impulse responses of the tested configuration. First, "target" and "interferer" waveforms were created by convolving speech-shaped noise with the binaural impulse responses between the listener and source positions. These signals were equivalent to those delivered to the listeners over headphones in the psychophysical experiments, except that the prediction used noise instead of sentences for the target. The signals were used to evaluate separately the binaural and monaural components of masking by the interferer.

The interaural parameters allow the evaluation of the binaural advantage associated with the azimuth separation of sources, by calculating the predicted BMLD of the tested configuration. The signals were passed through a gammatone filterbank (Patterson *et al.*, 1987, 1988) with two filters per equivalent rectangular bandwidth (ERB) (Moore and Glasberg, 1983) which covered the frequency range 20 to 10 000 Hz.[1] The left and right channels of the filtered waveforms were cross-correlated within a 100-ms exponentially tapering temporal window, with delays determined between + and −5 ms, as implemented in IWAVE (Culling, 1996). The interaural coherence of the interferer and the interaural phase differences of both target and interferer were obtained in each frequency band from the interaural cross-correlation functions. The coherence was computed as the maximum of the cross-correlation, whereas the phase difference was calculated by multiplying the delay corresponding to this maximum by the center frequency of the band. The BMLD was then computed as a function of frequency following the calculation proposed by Culling *et al.* (2004, 2005), using the formula given in Culling *et al.*, 2005,

$$\mathrm{BMLD} = 10\log_{10}([k-\cos(\Phi_T-\Phi_I)]/[k-\rho]), \tag{1}$$

with

$$k = (1+\sigma_\varepsilon^2)\exp(\omega^2\sigma_\delta^2), \tag{2}$$

where ρ is the interaural coherence of the interferer, Φ_T and Φ_I are the interaural phase differences in radians of the target and interferer, ω is the center frequency of the band in rad/s, $\sigma_\delta=0.000\,105$, and $\sigma_\varepsilon=0.25$ (Durlach, 1972). Where the formula returns a negative BMLD, the value is set to zero, following the assumption that binaural thresholds are never below their corresponding monaural thresholds (Durlach, 1963). This computation was realized at four epochs (starting at 0.5, 1, 1.5, and 2 s) during the 4.3-s-duration noises, and the four resulting BMLD values were averaged. Finally, the broadband binaural advantage was obtained by integrating the average BMLD function along the frequency scale, using the SII weightings (ANSI, 1997).

The spectrum of a room impulse response depends on the room and the positions of listener and source. Because this spectrum is not flat, the convolution by a room impulse response alters the noise interferer spectrum (in the present study, the target was always anechoic, so its spectrum was not affected by reverberation). The room-induced coloration

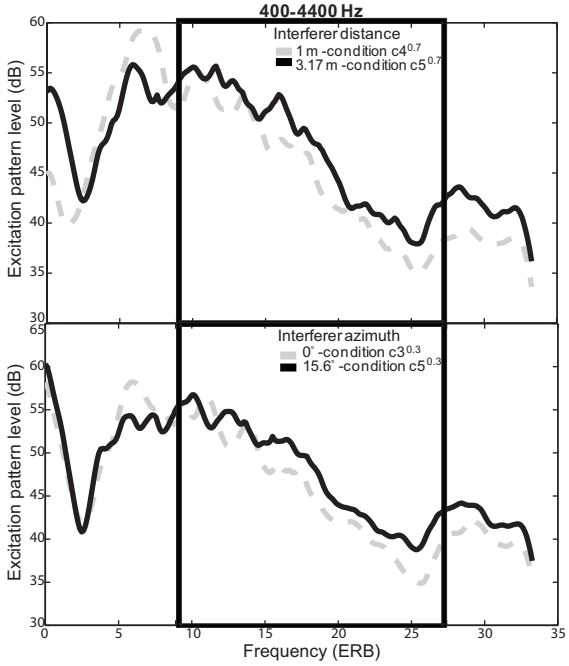


FIG. 2. Minimum excitation pattern levels at the ears for a speech-shaped noise in four interferer conditions of experiment 2, revealing the influence of room coloration. The top panel compares two conditions ($c4^{0.7}$,$c5^{0.7}$) which differed by their listener-interferer distance; the bottom panel compares two conditions ($c3^{0.3}$,$c5^{0.3}$) which differed by their interferer azimuth (see Fig. 7 and Table II for details). Each panel highlights the frequency range most important for intelligibility (400–4400 Hz) according to the SII weightings (ANSI, 1997). In each panel, the dark pattern was almost always above the dashed pattern in this frequency range, indicating an interferer more masking for speech.[2]

of the interferer can influence target intelligibility, even after rms equalization, because frequency bands are not all equivalent regarding intelligibility. An interferer that has more energy in a frequency range important for intelligibility will cause more masking than an interferer that has more energy in a frequency range not important for intelligibility. Figure 2 presents the excitation pattern of the interferer in four conditions of experiment 2. These conditions are described in detail in Sec. IV (Fig. 7, Table II). The top panel compares two conditions differing in their listener-interferer distance; the bottom panel compares two conditions differing in their interferer azimuth. Even though these conditions were equalized in broadband level, the level of the interferer was not the same in every frequency band in the different conditions. The room did not filter the interferer in the same way for different interferer positions. Figure 2 (box) highlights the frequency range most important for intelligibility according to the SII weightings. In each panel, the solid black line is the excitation pattern of an interferer that had more energy than the other (dashed gray line) in this frequency region. The former was therefore a stronger masker for speech.[2]

Even if our experiments were specifically designed to test binaural interaction and eliminate head shadow (no head simulated, broadband ILDs equalized), better-ear listening was reduced but still possible due to frequency-dependent ILDs introduced by room coloration. Our prediction method had to take this influence into account by computing a monaural component of masking by the interferer. Better-ear listening was simulated by frequency bands following the model proposed by Zurek (1993) to describe head shadow in anechoic situations. The target-to-interferer ratio was evaluated using the excitation pattern of the target and interferer waveforms (Fig. 1). The excitation pattern was computed between 0 and 33.25 ERBs (corresponding to the frequency range 0–10 kHz), every 0.13 ERBs, at the left and right ears of the listener. The signal-to-noise ratio (SNR) was evaluated in each frequency band at each ear by computing the difference between the target and interferer patterns.[3] To evaluate a better-ear SNR at each frequency, the maximum of the ratio at the left and right ears was considered. The target-to-interferer ratio was computed by integrating this maximum ratio across frequency using the SII weightings.

The prediction method combines the broadband binaural and monaural components by adding them together, in order to estimate the "effective" target-to-interferer ratio associated with the tested configuration. Although some studies decomposing binaural interaction and better-ear listening found that their combined effect was smaller than the sum of the two isolated effects (Bronkhorst and Plomp, 1988; Culling *et al.*, 2004), additive models proved to work well in other cases (Hawley *et al.*, 2004; Zurek, 1993). Because the influence of better-ear listening was limited by our experimental protocol, we used an additive model as a first approximation. The prediction method is relative; it does not give an absolute evaluation of intelligibility. The target-to-interferer ratio needs to be compared to a reference configuration to derive an intelligibility prediction. In the experiments presented here, this ratio was directly compared to the SRT measured for the given room and source configurations. Because SRTs are target-to-interferer ratio by definition, the comparison of measurement and prediction did not require any index calculation (AI, SII) nor any index-to-intelligibility mapping function (Beutelmann and Brand, 2006; Levitt and Rabiner, 1967b). To allow the comparison, the average SRT of the experiment was chosen as a reference. A constant value was added to the predictions, so that the average prediction was equal to the average SRT. This procedure is equivalent to the normalization of all thresholds to a reference condition, as done in spatial unmasking experiments which usually use a reference condition where target and interferer are co-located (Zurek, 1993; Zurek *et al.*, 2004). The prediction of the threshold difference between two conditions is not affected, and it can be compared to measurement without taking into account the influence of speech material on thresholds (ANSI, 1997).

III. VARYING THE AZIMUTH SEPARATION OF SOURCES IN A GIVEN CONFIGURATION (EXP. 1)

Experiment 1 was designed to test whether the proposed prediction method could describe the influence of the relative interaural phase of sources on binaural interaction in rooms. SRTs were measured for an interferer at a fixed position, with an anechoic target tested at different azimuths around the listener. Different levels of reverberation were applied to the interferer, by varying the absorption coefficient of the room. As interaural phase varies with the target azimuth, and as the coherence of the interferer at a fixed position only

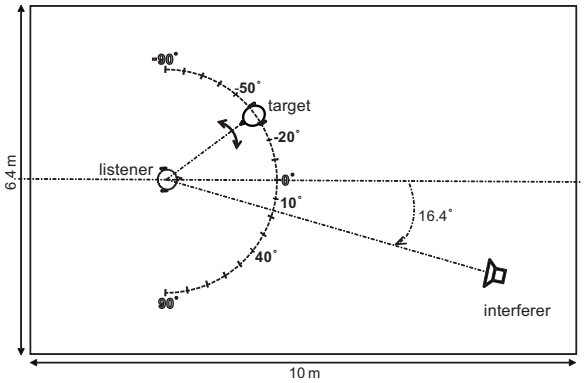


FIG. 3. Virtual room and spatial configuration used in experiment 1, and for the calculations of Fig. 5. The interferer was at 6.16 m from the listener, at 16.4° of azimuth. The target was at 2 m from the listener. In experiment 1, it was tested at −50°, −20°, 10°, and 40° of azimuth. The calculations of Fig. 5 used a target every 10° between −90° and 90° of azimuth, and a target at 16.4° of azimuth.

depends on the level of reverberation, this design produced variation in the BMLDs with target azimuth at several fixed levels of interferer coherence.

A. Design

Experiment 1 was based on the configuration presented in Fig. 3. The target was at 2 m from the listener and was always anechoic. SRTs were measured for four target azimuths: −50°, −20°, 10° and 40°. The room was 10-m long, 6.4-m wide and 2.5-m high. The ears of the listener were placed along an axis parallel to the 6.4-m wall on either side of a center point located at 3.2 m from the 10-m wall and 2.5 m from the 6.4-m wall. The interferer was at 6.16 m from the listener, at 16.4° of azimuth. When computing the room impulse responses used for the interferer, three absorption coefficients were used: 1 (anechoic), 0.7 (mildly reverberant), and 0.2 (very reverberant). These room characteristics and positions were chosen so that they led to different values of interferer coherence at the listener position. Table I shows that the "wide band" coherence of the interferer decreased with increasing reverberation (decreasing absorption coefficient). This form of coherence was considered in two previous studies (Lavandier and Culling, 2007, 2008). It was calculated using the waveforms obtained from the convolution of a speech-shaped noise with the binaural impulse response between the interferer and listener positions. Coherence was computed as the maximum of the cross-correlation of the waveforms.[4] These waveforms were filtered between 20 and 1500 Hz prior to calculation, in order to consider only the frequency range for which binaural unmasking is most effective in broadband noise. For comparison with common room acoustical measurements, Table I also presents the reverberation times (RTs) and direct-to-reverberant ratios involved in experiment 1. They were calculated at both ears, and the results were averaged across ears. The broadband RT was evaluated using a speech-shaped noise convolved with the impulse response between the interferer position and the considered ear, by measuring the time required for the sound level to decrease by 60 dB after the offset of the noise. The direct-to-reverberant energy ratio (D/R ratio) was calculated using the same impulse response. The direct-path and reverberant portions of this response were separated by a time-windowing procedure. This procedure was done by eye and was trivial as our simulated impulse responses were not contaminated by noise. The energy of each portion was computed, and the D/R ratio was expressed in dB. In the anechoic room, the D/R ratio was infinite, the RT was 0, and the coherence was 1, as there was no reverberant field.

TABLE I. Reverberation time, direct-to-reverberant ratio and interaural coherence, for a source at the interferer position of Fig. 3, as a function of the room absorption coefficient used for the interferer in experiment 1.

Absorption coefficient	RT (s)	D/R ratio (dB)	Interaural coherence
1	0	∞	1
0.7	0.22	−3.1	0.89
0.2	1.30	−15.7	0.75

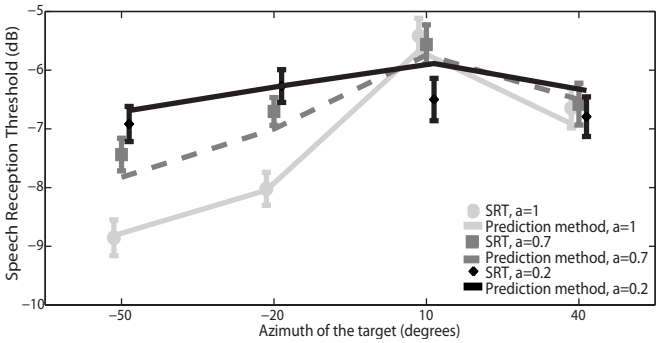


FIG. 4. Mean SRTs with standard errors measured in experiment 1, for four target azimuths and three absorption coefficients (a) used for the interferer. The SRTs are compared to the results of the prediction method.

In experiment 1a, the interferer was either anechoic (absorption coefficient of 1) or very reverberant (absorption coefficient of 0.2). A session of the experiment consisted of measuring the SRTs for the four target azimuths at two levels of reverberation for the interferer, resulting in eight different conditions. Twenty-four listeners took part in a 40-min session of the experiment. The aim of experiment 1b was to test an intermediate level of reverberation for the interferer. A session of the experiment consisted of measuring the SRTs for the four target azimuths with a mildly reverberant interferer (absorption coefficient of 0.7), twice for each listener, resulting in eight conditions. The experiment used the same target sentences as experiment 1a. Sixteen listeners took part in a 40-min session of the experiment; none of them had participated in experiment 1a.

B. Results

Figure 4 presents the mean SRTs measured in experiment 1, for four target azimuths and three reverberation conditions for the interferer. With the anechoic and mildly reverberant interferers, the SRT was maximal when the target azimuth was 10°, close to that of the interferer (16.4°), and decreased when the target was moved away from the interferer. When reverberation was added to the interferer, SRTs tended to increase and became independent of the target azimuth with the very reverberant interferer. Figure 4 also presents the results of the prediction method. The proposed

method led to good predictions of the measured SRTs, with a 0.95 correlation obtained between the SRTs and their predicted value.

Two within-subject analyses of variance confirmed that, on average, reverberation on the interferer impaired intelligibility [$F(1,23)=18.3$, $p<0.001$, Exp. 1a], and that increasing the azimuth separation of sources could produce a significant release from masking [$F(3,69)=15.5$, $p<0.0001$, Exp. 1a], even with a moderately reverberant interferer [$F(3,45)=6.1$, $p<0.01$, Exp. 1b]. The interaction between the effects of reverberation and target azimuth was significant in experiment 1a [$F(3,69)=8.0$, $p<0.001$]. A simple-main-effect analysis of this interaction showed that the effect of reverberation was significant at all target azimuths except 40° [$F(1,23)>14$, $p<0.0015$ in each case]. The azimuth of the target had a significant effect on SRTs when the interferer was anechoic [$F(3,69)=27.5$, $p<0.0001$], but not when it was very reverberant. Tukey pairwise comparisons showed that, in experiment 1a, the target azimuth of −50° led to lower SRTs than the target azimuths of −20°, 10°, and 40°, and that the target azimuth of −20° also led to lower SRTs than the target azimuth of 10° [$q>5.7$, $p<0.001$ in each case]. The SRTs for the target azimuth of 40° were not significantly different from the SRTs for the target azimuths of −20° and 10°. In experiment 1b, the target azimuth of −50° led to lower SRTs than the target azimuth of 10° [$q>6$, $p<0.001$]. The SRTs for the other target azimuths were not significantly different.

C. Discussion

Experiment 1 validated the prediction method for sources with different azimuth separations in a given room. A high correlation was obtained between measurements and predictions. Lavandier and Culling (2007, 2008) showed that the effect of reverberation on binaural interaction was related to the interaural coherence of the interferer. These new results show that BMLD calculations also incorporating the interaural phase of the target and interferer can account for the effects of both azimuth and reverberation level.

To obtain a more detailed picture of the effect of separation in azimuth, Fig. 5 presents the predicted binaural advantage calculated for the target placed every 10° between −90° and 90° of azimuth around the listener, and at 16.4° of azimuth to test for no azimuth separation between target and interferer. This calculation was done at the three levels of reverberation used for the interferer in experiment 1. It should be noted that the binaural advantage curves were not symmetric relative to the interferer azimuth, because the tested configuration was not symmetric within the room, the interferer being on the right of the listener (Fig. 3). The binaural advantage increased when the target azimuth was moved away from that of the interferer and started to flatten for azimuth separations above 50°. These predictions are in agreement with both the anechoic measurements of Bronkhorst and Plomp (1988) and the results of experiment 1. Bronkhorst and Plomp (1988) measured the azimuth dependency of binaural interaction ("dT" condition) using a frontal target at 0°, and moving the interferer by steps of 30° between 0° and 180°. They measured a 3-dB effect of the azimuth at 30°, but no effect with increasing azimuth until the source reached the symmetrical position at the back of the listener, at 150°, where the SRTs became sensitive to the source azimuth again. Our measurements were done between −50° and 40°, with the same 30° step size and the competing source at 16.4°. Therefore, the main difference between our anechoic condition and the dT condition of Bronkhorst and Plomp (1988) is that we limited the range of tested positions to those where the influence of source azimuth was maximal, whereas Bronkhorst and Plomp (1988) measured SRTs at equally spaced azimuths over 180°.

For spatially separated sources, reverberation of the interferer reduced both the predicted binaural advantage and its sensitivity to changes in azimuth separation of sources. These reductions resulted from two effects acting in parallel so that an E-C mechanism would gradually become inefficient with increasing reverberation: the decrease in interferer coherence, ρ, which evaluates the proportion of noise energy that is susceptible to cancellation, and the modification of the interferer interaural phase difference, Φ_I, which needs to be different from that of the target for cancellation to result in an improvement in internal signal-to-noise ratio. To evaluate the relative contributions of reverberant coherence and phase to the reduction in the predicted binaural advantage, Fig. 6 presents model calculations applied to stimuli in which phase and coherence were varied independently. For example, the binaural advantage was computed at all target azimuths using the phase of the anechoic interferer and the coherence of the mildly reverberant interferer (dashed line, panel I) or very reverberant interferer (dashed line, panel III). The three levels of reverberation used in Fig. 5 are considered here by pairs. The predictions at the two levels of each pair are compared with predictions involving the phase of one level and the coherence of the other, and reciprocally. Increasing reverberation on the interferer coherence reduced the binaural advantage for all target azimuths (e.g., light gray line above dashed line in panel III), in agreement with binaural unmasking theories, with a floor effect observed for azimuths close to that of the target. Increasing reverberation on the interferer phase reduced the binaural advantage sensitivity to changes in azimuth separation of sources (e.g., light gray line vs dotted line in panel III). For small azimuth separations, where BMLDs are small in anechoic situations, the room reflections distributed part of the interferer energy at different phases, corresponding to large "apparent" azimuth separations for the associated virtual images of the interferer, so that this energy could be canceled and the binaural advantage increased. For large azimuth separations, where anechoic BMLDs are large, reflections brought part of the interferer energy at phases close to that of the target. This energy could not be canceled and the binaural advantage decreased. Beutelmann and Brand (2006) interpreted the influence of early reflections as responsible for the creation of mirror sources acting as secondary interferers with considerably different azimuths, thus disturbing the binaural system (cafeteria condition). These mirror sources would indeed greatly reduce head shadow if they were on the other side of the head compared to the interferer (Culling *et al.*, 2004; Hawley *et al.*,

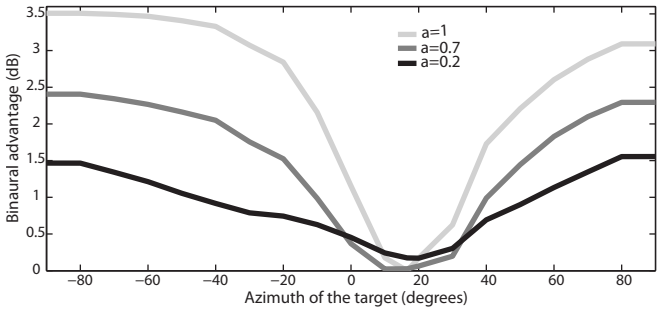


FIG. 5. Predicted binaural advantage as a function of target azimuth in the configuration of Fig. 3, for the three absorption coefficients (a) used for the interferer in experiment 1.

2004), and our results show that they would also influence binaural interaction if they affect the interferer coherence and phase at the ears. It should be noted, however, that mirror sources might not systematically reduce binaural interaction. When considering only interaural phase, if the interferer azimuth is close to that of the target, then a mirror source at a widely different angle could be far from the target azimuth, distributing interfering energy where it can be canceled, so that the binaural advantage might increase depending on the associated effect of reflections on coherence.

When sources have the same interaural phase, reducing the coherence of the interferer can produce a small release

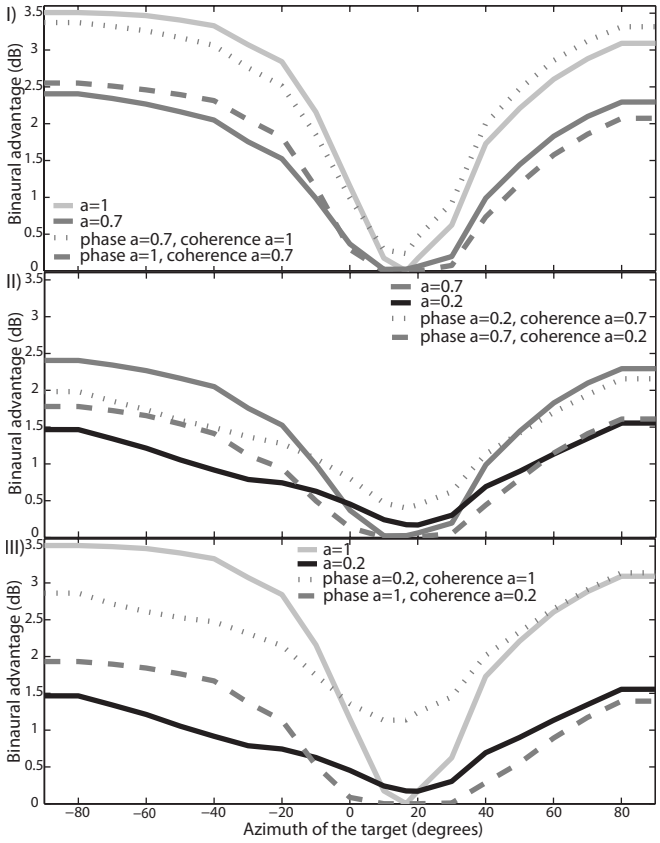


FIG. 6. Relative contributions of reverberant coherence and phase to the reduction of predicted binaural advantage. The predictions of Fig. 5 obtained with three absorption coefficients (a) used for the interferer are presented by pairs ($a=1-0.7$ in I, $a=0.7-0.2$ in II, and $a=1-0.2$ in III). Each pair is associated with the corresponding predictions involving the interferer phase obtained with the first coefficient and the interferer coherence obtained with the second, and reciprocally.

from masking (Licklider, 1948). A corresponding effect occurred in experiment 1a (Fig. 4). For the target azimuth of 10°, close to that of the interferer (16.4°), the SRT significantly decreased when reverberation was added to the interferer. This small effect was somewhat larger than the prediction based on the BMLD calculations, which indicated no effect for the target at 10° and a very limited effect for the target at 16.4° (black line above the light gray line in Fig. 5). No such effect was predicted for the mildly reverberant interferer (dark gray line below the light gray line).

In experiment 1, the three levels of reverberation tested for the interferer led to very limited differences in room-induced coloration. The SII-weighted target-to-interferer ratio in the anechoic room was 0.19 dB lower than in the very reverberant room, and 0.21 dB lower than in the mildly reverberant room. Even without taking into account these small monaural differences, a 0.94 correlation was obtained between the SRTs and their predicted value based only on the binaural component of the method. The coloration did not play an important role here, because only one interferer-listener configuration was tested. More configurations and rooms were involved in experiment 2, so that among other things the monaural component of the prediction method was more thoroughly tested.

IV. VARYING THE CONFIGURATION FOR A GIVEN AZIMUTH SEPARATION OF SOURCES (EXP. 2)

Experiment 2 was designed to further validate the prediction method in a wide variety of configurations. To consider other aspects than those already measured and predicted in experiment 1, the azimuth separation of sources was kept constant. The aim of experiment 2 was to test whether the method could quantitatively predict the influence of the interferer coherence. This coherence was varied in many different ways. SRTs were measured while changing the absorption and the size of the room, the distance between the listener and the interferer, and the asymmetry of the configuration. Because all these design parameters could not be tested systematically here, the aim of the experiment was not to give a general conclusion on the link between each individual parameter and intelligibility. For example, we did not want to link distance or absorption to intelligibility. We were rather interested in the link between interferer coherence and intelligibility. Because the coherence of a source can be modified in many different ways in a room, we did not want to limit ourselves to one way or another. So, we varied the various design parameters that could influence coherence, and then we selected conditions leading to a large range of interferer coherence using these different parameters.

A. Design

In all conditions of experiment 2, the target was anechoic and placed at 2 m from the listener, at 60° on the left of the interferer.[5] The interferer was tested at different positions in different rooms.

Figure 7 and Table II present the room, absorption coefficient, and position used for the interferer in each condition of experiment 2. Room *m* (medium) was the room used in

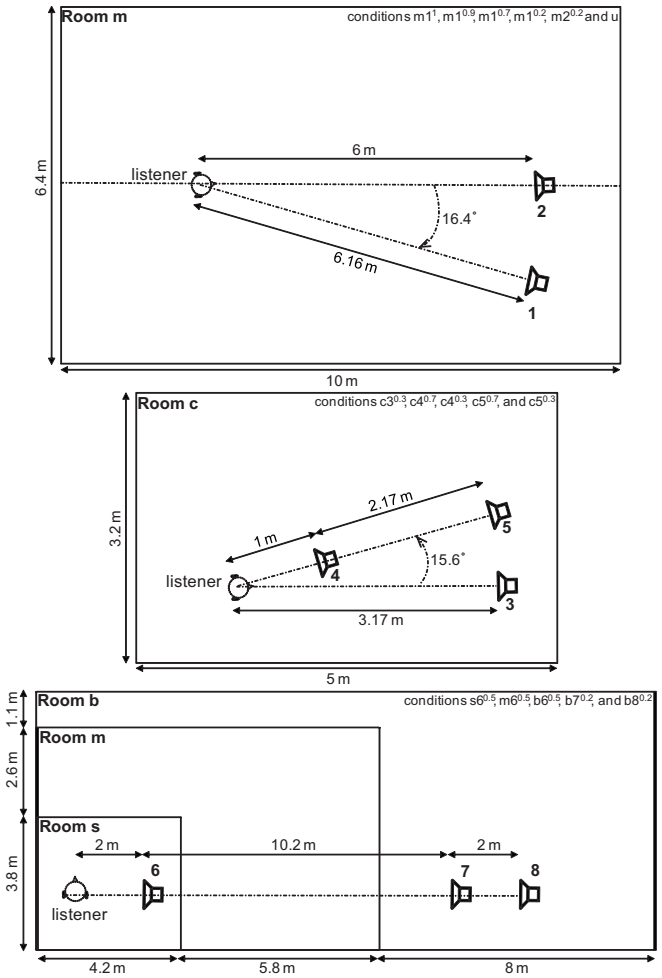


FIG. 7. Virtual rooms and positions used for the interferer in experiment 2. The anechoic target was always at 2 m from the listener, at 60° on the left of the interferer.[5]

experiment 1, with the same listener position used in conditions $m1^{1}$, $m1^{0.9}$, $m1^{0.7}$, $m1^{0.2}$, $m2^{0.2}$, and u. Position 1 was identical to the interferer position of experiment 1. Position 2 was on the median axis of the room, directly in front of the listener. In room *c* [common with Culling *et al.* (2003); Lavandier and Culling (2007)], the ears of the listener were placed along an axis parallel to the 3.2-m wall on either side of a center point located at 0.9 m from the 5-m wall and 1.2 m from the 3.2-m wall. Position 3 was directly in front of the listener, whereas positions 4 and 5 were at 15.6° azimuth, at two different distances. Conditions $s6^{0.5}$, $m6^{0.5}$, $b6^{0.5}$, $b7^{0.2}$, and $b8^{0.2}$ used rooms *s* (small), *m*, or *b* (big), with the ears of the listener placed along an axis parallel to the longer wall on either side of a center point located at 1.6 m from the longer wall and 1.2 m from the shorter wall. Positions 6, 7, and 8 were at different distances from the listener, at 90° azimuth. The sizes of the rooms and the listener-interferer distances are detailed in Fig. 7.

Table II presents the reverberation time, direct-to-reverberant ratio, and interaural coherence for a source at the interferer position, as well as the predicted binaural advantage, for all tested conditions. Because interaural coherence was the parameter of interest here, the conditions were ranked by decreasing interferer coherence. Condition $m1^{1}$ used an anechoic interferer, with a coherence of 1. The configuration of condition $m2^{0.2}$ was symmetric regarding the room, the interferer, and the ears of the listener. In this special case, listener and interferer were both on the median axis of the room, so that the reflections getting to the listener from the interferer were exactly the same at the two ears. The room was very reverberant (D/R ratio of −16.7 dB), but because the reverberation was identical at the two ears, it did not reduce the coherence of the interferer. This coherence was 1 as in the anechoic case, and the predicted binaural advantage was similar in the two conditions. Conditions $m1^{0.9}$ to $c5^{0.3}$ used reverberant interferers with progressively lower interaural coherence. Condition u was introduced to include the reference of an interferer completely uncorrelated

TABLE II. Room, absorption coefficient and position used for the interferer in each condition of experiment 2 (see Fig. 7 for details of rooms and positions). Reverberation time, direct-to-reverberant ratio, and interaural coherence for a source at the interferer position, and predicted binaural advantage in each condition. The conditions are ranked by decreasing coherence.

Condition	Room	Absorption coefficient	Position	RT (s)	D/R ratio (dB)	Interaural coherence	Binaural advantage (dB)
$m1^{1}$	m	1	1	0	∞	1	3.32
$m2^{0.2}$	m	0.2	2	1.23	−16.7	1	3.08
$m1^{0.9}$	m	0.9	1	0.09	3.7	0.96	2.87
$c4^{0.7}$	c	0.7	4	0.12	4.9	0.92	2.40
$m1^{0.7}$	m	0.7	1	0.22	−3.1	0.89	2.12
$c3^{0.3}$	c	0.3	3	0.44	−12	0.83	1.92
$c4^{0.3}$	c	0.3	4	0.47	−3	0.79	1.82
$m1^{0.2}$	m	0.2	1	1.30	−15.7	0.75	1.01
$c5^{0.7}$	c	0.7	5	0.16	−2.7	0.72	1.61
$m6^{0.5}$	m	0.5	6	0.39	−2	0.68	0.34
$b6^{0.5}$	b	0.5	6	0.64	−1.6	0.67	0.43
$b7^{0.2}$	b	0.2	7	2.34	−19.2	0.65	0.18
$s6^{0.5}$	s	0.5	6	0.20	−4.4	0.61	0.42
$b8^{0.2}$	b	0.2	8	2.45	−20.4	0.54	0.51
$c5_{0.3}$	c	0.3	5	0.47	−12.2	0.48	0.97
u	m	1	1[a]	0	∞	0.05	0.52

[a]Using independent noises for each ear.

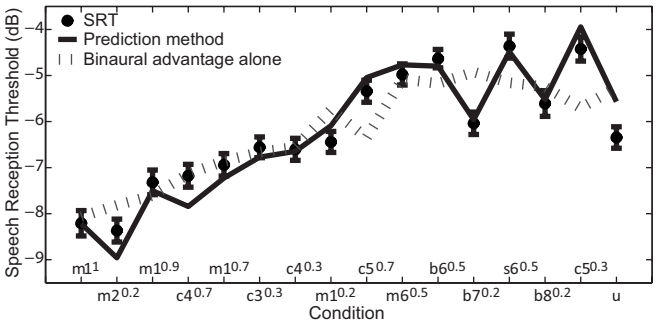


FIG. 8. Mean SRTs with standard errors measured in the sixteen conditions of experiment 2. The SRTs are compared to the results of the prediction method. The predictions based on the binaural advantage alone are also plotted.

at the two ears, with a coherence very close to 0. It used exactly the same configuration as the anechoic condition $m1^{1}$, but instead of convolving one speech-shaped noise with the binaural impulse response, two independent speech-shaped noises were used. One noise was convolved with the impulse response corresponding to the left ear; the other noise was convolved with the impulse response corresponding to the right ear.

A session of experiment 2 consisted of measuring the SRTs for the sixteen conditions. The experiment used different target sentences from experiment 1. Forty-eight listeners took part in a 70-min session of the experiment.

B. Results

Figure 8 presents the mean SRTs measured in the 16 conditions of experiment 2, the conditions being ranked by decreasing coherence. The SRTs initially increased with decreasing coherence (from conditions $m1^{1}$ to $b6^{0.5}$), but the relation was not monotonic at low coherence values (from conditions $b6^{0.5}$ to u). The results of the prediction method in the different conditions are also presented. The proposed method led to good predictions of the measured SRTs, with a 0.97 correlation obtained between the SRTs and their predicted value. The predictions based on the binaural advantage alone are also plotted. They predicted the general trend of increasing SRTs with decreasing coherence, but the correlation between the SRTs and their predicted value was only 0.84; they could not predict the non-monotonic behavior of the SRTs from condition $b6^{0.5}$ to condition u. Here, room coloration had substantial effects on SRTs, according to our model. Some conditions caused more masking than others because the corresponding interferers had more energy in frequency channels important for intelligibility.

Table III compares the SRTs of experiment 2 using a paired t-test analysis. The significant differences are indicated by crosses. The individual significance level was fixed to 0.0004, following the Bonferroni correction (Keppel and Wickens, 2004), in order to keep the family-wise significance level below 0.05 for the 120 comparisons realized. The changes in coherence resulting from the variation in the different design parameters often led to significant differences in measured SRTs. For example, conditions $m1^{1}$, $m1^{0.9}$, $m1^{0.7}$, and $m1^{0.2}$ differed only by the absorption used for the interferer, this absorption decreasing from $m1^{1}$ to $m1^{0.2}$. The anechoic condition $m1^{1}$ led to a significantly lower SRT than the two most reverberant conditions $m1^{0.7}$ and $m1^{0.2}$. Conditions $m2^{0.2}$ and $m1^{0.2}$ differed in the azimuth of the interferer. They used the same absorption and roughly the same level of reverberation (Table II), but condition $m2^{0.2}$ used a symmetric interferer with a coherence unaffected by reverberation. Condition $m1^{0.2}$ with the reduced coherence led to a significantly higher SRT than condition $m2^{0.2}$. The reverberation in condition $m2^{0.2}$ did not impair intelligibility, as shown by its SRT not being significantly different from that of the anechoic condition $m1^{1}$. The three pairs of conditions $c4^{0.7}$–$c5^{0.7}$, $c4^{0.3}$–$c5^{0.3}$, and $b7^{0.2}$–$b8^{0.2}$ correspond to three different rooms, in which the distance between interferer and listener was varied. Increasing this distance led to significantly higher SRTs in the case $c4^{0.7}$–$c5^{0.7}$ and $c4^{0.3}$–$c5^{0.3}$, but not in the case $b7^{0.2}$–$b8^{0.2}$. Conditions $m6^{0.5}$, $b6^{0.5}$, and $s6^{0.5}$ used the same configuration, but the size of the room was changed around the listener and sources. There was no significant effect of the room size on the SRTs.

TABLE III. Comparison of the SRTs measured in the sixteen conditions of experiment 2 using a paired t-test analysis. The significant differences are indicated by crosses.

	$m2^{0.2}$	$m1^{0.9}$	$c4^{0.7}$	$m1^{0.7}$	$c3^{0.3}$	$c4^{0.3}$	$m1^{0.2}$	$c5^{0.7}$	$m6^{0.5}$	$b6^{0.5}$	$b7^{0.2}$	$s6^{0.5}$	$b8^{0.2}$	$c5^{0.3}$	u
$m1^{1}$	0	0	0	×	×	×	×	×	×	×	×	×	×	×	×
$m2^{0.2}$	.	0	×	×	×	×	×	×	×	×	×	×	×	×	×
$m1^{0.9}$	.	.	0	0	0	0	0	×	×	×	0	×	×	×	0
$c4^{0.7}$	.	.	.	0	0	0	0	×	×	×	0	×	×	×	0
$m1^{0.7}$	.	.	.	.	0	0	0	×	×	×	0	×	0	×	0
$c3^{0.3}$	.	.	.	.	.	0	0	×	×	×	0	×	0	×	0
$c4^{0.3}$	.	.	.	.	.	.	0	×	×	×	0	×	0	×	0
$m1^{0.2}$	.	.	.	.	.	.	.	0	×	×	0	×	0	×	0
$c5^{0.7}$	.	.	.	.	.	.	.	.	0	0	0	×	0	0	0
$m6^{0.5}$	.	.	.	.	.	.	.	.	.	0	0	0	0	0	×
$b6^{0.5}$	.	.	.	.	.	.	.	.	.	.	×	0	0	0	×
$b7^{0.2}$	.	.	.	.	.	.	.	.	.	.	.	×	0	×	0
$s6^{0.5}$	.	.	.	.	.	.	.	.	.	.	.	.	0	0	×
$b8^{0.2}$	.	.	.	.	.	.	.	.	.	.	.	.	.	0	0
$c5^{0.3}$	.	.	.	.	.	.	.	.	.	.	.	.	.	.	×

C. Discussion

While experiment 1 showed that the prediction method correctly took into account the influence of the interaural phases of target and interferer, experiment 2 showed that it accurately described the influence of reverberation on the coherence and spectrum of the interferer, in a wide variety of configurations. A high correlation was obtained between measurements and predictions. The predictions based on the binaural advantage alone described the general trend of increasing SRTs with decreasing coherence, but they could not predict the non-monotonic behavior of the SRTs measured at high levels of reverberation. At low levels of reverberation, SRTs seemed to be predominantly influenced by interferer coherence and binaural interaction. At high levels of reverberation, the binaural advantage was very limited and it became important to take into account room coloration.

If the room reflections coming from the interferer were different at the listener's ears, then reverberation reduced the interferer coherence, resulting in lower binaural advantages and higher SRTs. No matter how late or energetic they were, these reflections did not impair intelligibility if they were identical at the two ears[6]: intelligibility in the reverberant but symmetric configuration of condition $m2^{0.2}$ was as good as in the anechoic condition $m1^{1}$. The reverberation on the interferer impaired intelligibility only if it reduced its coherence, in agreement with Lavandier and Culling (2008). Measuring target relevant characteristics such as the direct-to-reverberant ratio or the reverberation time could not predict this result, whereas a prediction based on the E-C theory can: similar values of binaural advantage led to similar SRTs.

In experiment 2, the binaural advantage and interferer coherence were reduced in various ways, which led to significant increases in SRT. Decreasing the room absorption gave more energy to the sound reflections, which made the interferer less coherent if these reflections were different at the two ears (conditions $m1^{1}$ vs $m1^{0.7}$ and $m1^{0.2}$). When the distance between interferer and listener increased, the reverberant sound, which was incoherent, became more important relative to the direct sound, which was coherent, so that the interferer coherence decreased (conditions $c4^{0.7}$ vs $c5^{0.7}$ and $c4^{0.3}$ vs $c5^{0.3}$). These effects of absorption and distance were consistent with previous measurements (Lavandier and Culling, 2007). Varying the azimuth of the interferer could also decrease its coherence and increase the SRT (condition $m2^{0.2}$ vs $m1^{0.2}$ and $c3^{0.3}$ vs $c5^{0.3}$). There was no significant effect of the size of the room used for the interferer in conditions $m6^{0.5}$, $b6^{0.5}$, and $s6^{0.5}$; but it should be noted that the predicted binaural advantage was very similar in these conditions.

Table II shows that the main influence of interferer coherence on the predicted binaural advantage took place at low levels of reverberation, for coherence between 1 and 0.75 (conditions $m1^{1}$ to $m1^{0.2}$). Further decreases in coherence had less influence. This is in agreement with the results of Licklider (1948), who showed that most of the variation in intelligibility for speech in noise occurs for a noise coherence between 1 and 0.75. This reduced sensitivity at low coherence might explain why the SRTs in conditions $c5^{0.7}$ vs $c5^{0.3}$ and $b7^{0.2}$ vs $b8^{0.2}$ did not differ significantly. Conditions $m1^{0.9}$ vs $m1^{0.7}$ and $c4^{0.7}$ vs $c4^{0.3}$ did not lead to significant differences either, probably because their coherences and resulting binaural advantages were not sufficiently different. Conditions $m1^{0.9}$ and $m1^{0.2}$ were expected to result in different SRTs, due to the magnitude of their difference in interferer coherence (Lavandier and Culling, 2007). However, the 1.86-dB predicted difference of binaural advantage was slightly reduced by room coloration, so that the overall difference predicted from the model was only 1.41 dB, and the results of experiment 2 showed a non-significant difference of 0.9 dB (Fig. 8).

V. GENERAL DISCUSSION

The method proposed here to predict binaural interaction in rooms is an extension of the models of Levitt and Rabiner (1967b) and Zurek (1993). Binaural interaction is modeled by increasing the speech-to-interference ratio by the size of the BMLD for pure tone detection in noise in each frequency band. BMLDs are estimated for the given set of interaural parameters using the formula proposed by Culling *et al.* (2005). Unlike the BMLD estimations used by Levitt and Rabiner (1967b) and Zurek (1993), this formula depends on the interaural coherence of the interferer, so it can describe the decorrelation of the interferer at the listener's ears by reverberation and the corresponding impairment of binaural interaction, allowing us to extend the predictions to reverberant situations. Like the model proposed by Beutelmann and Brand (2006), the method is based on the information contained in the room impulse responses rather than room statistics (Zurek *et al.*, 2004), but the direct implementation of a complex E-C mechanism is replaced by the more computationally efficient BMLD estimation. It should be noted that the threshold difference between two compared configurations is predicted without any model parameter being fitted to the data.

The method correctly took into account the influence of the azimuth separation of sources and the importance of their interaural phases in experiment 1, with a 0.95 correlation obtained between the SRTs and their predicted value. It also accurately described the influence of reverberation on the coherence and spectrum of the interferer in experiment 2, leading to a 0.97 correlation between measurement and prediction. These results are comparable with those of Beutelmann and Brand (2006), who obtained an overall correlation of 0.95 between measured and predicted SRTs (0.97 in the anechoic room, 0.94 in the office and cafeteria). In terms of prediction errors, the maximum and mean prediction errors were 0.6 and 0.2 dB in experiment 1; they were 0.8 and 0.3 dB in experiment 2. The maximum and mean prediction errors obtained by Beutelmann and Brand (2006) were 3 and 1.6 dB in the anechoic room, about 1 and 0.5 dB in the office, and about 1 and 0.3 dB in the cafeteria. Zurek *et al.* (2004) obtained maximum prediction errors ranging between 2 and 5 dB for the binaural detection of narrow band noises in rooms. These discrepancies could be due to the approximations inherent to the use of room statistics and of a fixed correlation function in the E-C model.

Reverberation impairs binaural interaction for noise and speech interferers (Lavandier and Culling, 2008), and room coloration also affects both types of interferer. The proposed prediction method is not limited to noise interferers and can be used to evaluate part of the effect of reverberation on speech interferers. However, the prediction would not be complete, because additional speech segregation mechanisms are involved with speech interferers (Lavandier and Culling, 2008). Room reflections can disrupt the segregation of competing sentences based on fundamental frequency differences (Culling *et al.*, 2003, 1994), and they can also fill the potential silent periods in the speech interferers (Bronkhorst and Plomp, 1990; George *et al.*, 2008). These additional effects need to be described to refine the intelligibility prediction. The direct effect of reverberation on the target needs to be considered as well (Bradley, 1986; Houtgast and Steeneken, 1985).

In the experiments presented in this paper, the head of the listener was not modeled. The aim was to avoid head shadow effects which could complicate the interpretation of the effects associated with binaural interaction. The head would lead to extra interaural decorrelation in asymmetric configurations, particularly for sources to the side (Lindevald and Benade, 1986). Because of the floor effect associated with coherence, the maximum intelligibility loss with increasing reverberation should be little affected by the presence or absence of a head, but the additional decorrelation associated with the head would make the interferer coherence drop at lower levels of reverberation. There is no reason to believe that the binaural component of our prediction method would not be able to take this additional decorrelation into account. Adding the head between the ears would also introduce a frequency-dependent head shadow. This head shadow can be seen as a coloration component which is different at the two ears. The better-ear target-to-interferer ratio of our prediction correctly described the coloration introduced by the room, so it should be able to predict this head coloration as well. Zurek (1993) used a similar SNR calculation to predict the head shadow in anechoic situations. Though it seems plausible that our prediction method will describe the influence of the head, it remains to be tested. Moreover, by simply adding the binaural and monaural components of masking, the method neglects any potential interaction between better-ear listening and binaural interaction. It led to very good results in this study where the influence of better-ear listening was limited by our experimental protocol. With the extra ILDs introduced by the head, the assumption of additivity might not hold anymore. The interaction would then need to be modeled. The applicability of the method to real-life situations also requires it to be tested in real rooms. It could be done using binaural room impulse responses similar to those measured by Watkins (2005) instead of our simulated room responses.

Several studies have shown that binaural interaction can still be efficient to partly unmask a target among multiple interferers placed at different positions in an anechoic environment (Bronkhorst and Plomp, 1992; Carhart *et al.*, 1969; Culling *et al.*, 2004). Culling *et al.* (2004) showed that BMLD calculations could be used to describe SRTs measured in such configurations. The prediction method proposed in this paper might generalize to a setting with more than one interfering source in rooms. The number of these sources should, in principle, be immaterial to the model, since it is already handling multiple virtual images of one source. The parameters—interaural phase and coherence, and excitation pattern—of the single interferer of Fig. 1 would need to be replaced by the same parameters calculated on the overall interfering signals, obtained by adding the signals produced at the ears by each interferer. The prediction method does not need to "know" if the interaural phase corresponds to the position of a single interferer, to multiple interferers at different positions, or to room reflections mixed with the direct sound at the ears. In the same way, it does not matter if the source spectra at the ears result from head shadow, room coloration, or multiple sources. A prediction of intelligibility should be possible in each case because the auditory mechanisms involved remain the same.

VI. CONCLUSION

Using SRT measurements and acoustical analyses done in several virtual rooms, a method was proposed to quantitatively predict the influence of reverberation on the binaural interaction mechanism used to improve speech intelligibility in the presence of noise interferers. This method, based on predicted BMLDs, took into account the decorrelation of the interferer at the listener's ears by room reflections, and the dependence of binaural interaction on the azimuth separation of sources. To fully predict the measured losses of intelligibility caused by reverberation, the coloration introduced by the room also had to be considered. A 0.95–0.97 correlation was obtained between the SRTs and their predicted value. The method requires calculation of the excitation pattern and interaural phase of target and interferer, and of the interaural coherence of the interferer. These parameters were all calculated from the room impulse responses convolved with noise. At low levels of reverberation, SRTs seemed predominantly influenced by binaural interaction, so that the azimuth separation of sources and the interferer coherence were critical. At high levels of reverberation, the binaural advantage was very limited and it became important to take into account room coloration.

ACKNOWLEDGMENTS

This work was supported by the UK EPSRC. The authors are grateful to the associate editor Richard Freyman and three anonymous reviewers for their helpful comments on the first version of this paper, and to all listeners who took part in the experiments.

[1]The upper limit of the filterbank was unnecessarily high, given that BMLDs occur mostly below 1500 Hz (Culling *et al.*, 2004). We choose to do the calculation up to 10 000 Hz to use the same frequency range as the one used for the computation of the excitation patterns.

[2]The peak of each excitation pattern observed at very low frequencies (below 2.5 ERBs) was due to an artifact inherent to the room simulation program. This dc component increasing with the level of reverberation was not audible, and this irrelevant energy was not taken into account while equalizing the rms level of the stimuli.

[3]In this study, the target was always anechoic and equalized in rms level at

the two ears, so that the excitation pattern of the target waveform was constant across ears and conditions. Therefore, the target-to-interferer ratio only depended on the excitation pattern of the interferer.

[4]Coherence was not computed as the maximum of the absolute value of the cross-correlation, as indicated by mistake in Lavandier and Culling, 2007, 2008.

[5]The azimuth separation of sources was 60° in all conditions of experiment 2 except in the configuration used for conditions $m1^{1}$, $m1^{0.9}$, $m1^{0.7}$, $m1^{0.2}$, and u (which correspond to the configuration of Fig. 3 with a target azimuth of −50°), where it was 66.4°, due to a calculation error. It resulted in an azimuth separation of sources which was not perfectly constant across conditions, as had been planned, but this does not affect the validity of the results obtained and does not affect the conclusions of the study. With such a separation, following the calculations of Fig. 5, the binaural advantages in the tested configurations were mainly influenced by the changes in the interferer interaural coherence.

[6]Of course, if the configuration is such that the reflections coming from the interferer are different at the two ears, the more energetic they are, the more deleterious effect they should have on interferer coherence and target intelligibility.

Allen, J. B., and Berkley, D. A. (**1979**). "Image method for efficiently simulating small-room acoustics," J. Acoust. Soc. Am. **65**, 943–950.

ANSI (**1997**). "Methods for calculation of the speech intelligibility index," ANSI S3.5-1997, American National Standards Institute, New York.

Beutelmann, R., and Brand, T. (**2006**). "Prediction of speech intelligibility in spatial noise and reverberation for normal-hearing and hearing-impaired listeners," J. Acoust. Soc. Am. **120**, 331–342.

Bradley, J. S. (**1986**). "Predictors of speech intelligibility in rooms," J. Acoust. Soc. Am. **80**, 837–845.

Bradley, J. S., Reich, R. D., and Norcross, S. G. (**1999**). "On the combined effects of signal-to-noise ratio and room acoustics on speech intelligibility," J. Acoust. Soc. Am. **106**, 1820–1828.

Bronkhorst, A. W., and Plomp, R. (**1988**). "The effect of head-induced interaural time and level differences on speech intelligibility in noise," J. Acoust. Soc. Am. **83**, 1508–1516.

Bronkhorst, A. W., and Plomp, R. (**1990**). "A clinical test for the assessment of binaural speech perception in noise," Audiology **29**, 275–285.

Bronkhorst, A. W., and Plomp, R. (**1992**). "Effect of multiple speechlike maskers on binaural speech recognition in normal and impaired hearing," J. Acoust. Soc. Am. **92**, 3132–3139.

Carhart, R., Tillman, T. W., and Greetis, E. S. (**1969**). "Release from multiple maskers: Effects of interaural time disparities," J. Acoust. Soc. Am. **45**, 411–418.

Colburn, H. S. (**1977**). "Theory of binaural interaction based on auditory-nerve data. II. Detection of tones in noise," J. Acoust. Soc. Am. **61**, 525–533.

Culling, J. F. (**1996**). "Signal processing software for teaching and research in psychoacoustics under UNIX and X-windows," Behav. Res. Methods Instrum. Comput. **28**, 376–382.

Culling, J. F., Hawley, M. L., and Litovsky, R. Y. (**2004**). "The role of head-induced interaural time and level differences in the speech reception threshold for multiple interfering sound sources," J. Acoust. Soc. Am. **116**, 1057–1065.

Culling, J. F., Hawley, M. L., and Litovsky, R. Y. (**2005**). "Erratum: The role of head-induced interaural time and level differences in the speech reception threshold for multiple interfering sound sources," J. Acoust. Soc. Am. **118**, 552.

Culling, J. F., Hodder, K. I., and Toh, C. Y. (**2003**). "Effects of reverberation on perceptual segregation of competing voices," J. Acoust. Soc. Am. **114**, 2871–2876.

Culling, J. F., Summerfield, Q., and Marshall, D. H. (**1994**). "Effects of simulated reverberation on the use of binaural cues and fundamental-frequency differences for separating concurrent vowels," Speech Commun. **14**, 71–96.

Durlach, N. I. (**1963**). "Equalization and cancellation theory of binaural masking-level differences," J. Acoust. Soc. Am. **35**, 1206–1218.

Durlach, N. I. (**1972**). "Binaural signal detection: Equalization and cancellation theory," in *Foundations of Modern Auditory Theory*, edited by J. Tobias (Academic, New York), Vol. **II**, pp. 371–462.

Freyman, R. L., and Zurek, P. M. (**2008**). "Effects of reverberation on spatial release from masking (A)," J. Acoust. Soc. Am. **123**, 2977.

George, E. L. J., Festen, J. M., and Houtgast, T. (**2008**). "The combined effects of reverberation and nonstationary noise on sentence intelligibility," J. Acoust. Soc. Am. **124**, 1269–1277.

Hawley, M. L., Litovsky, R. Y., and Culling, J. F. (**2004**). "The benefit of binaural hearing in a cocktail party: effect of location and type of interferer," J. Acoust. Soc. Am. **115**, 833–843.

Houtgast, T., and Steeneken, H. J. M. (**1985**). "A review of the MTF concept in room acoustics and its use for estimating speech intelligibility in auditoria," J. Acoust. Soc. Am. **77**, 1069–1077.

IEEE (**1969**). "IEEE recommended practice for speech quality measurements," IEEE Trans. Audio Electroacoust. **17**, 227–246.

Keppel, G., and Wickens, T. D. (**2004**). *Design and Analysis. A Researcher's Handbook*, 4th ed. (Pearson Prentice-Hall, Englewood Cliffs, NJ).

Kryter, K. D. (**1962**). "Methods for the calculation and use of the articulation index," J. Acoust. Soc. Am. **34**, 1689–1697.

Lavandier, M., and Culling, J. F. (**2007**). "Speech segregation in rooms: Effects of reverberation on both target and interferer," J. Acoust. Soc. Am. **122**, 1713–1723.

Lavandier, M., and Culling, J. F. (**2008**). "Speech segregation in rooms: Monaural, binaural, and interacting effects of reverberation on target and interferer," J. Acoust. Soc. Am. **123**, 2237–2248.

Levitt, H., and Rabiner, L. R. (**1967a**). "Binaural release from masking for speech and gain in intelligibility," J. Acoust. Soc. Am. **42**, 601–608.

Levitt, H., and Rabiner, L. R. (**1967b**). "Predicting binaural gain in intelligibility and release from masking for speech," J. Acoust. Soc. Am. **42**, 820–829.

Licklider, J. C. R. (**1948**). "The influence of interaural phase relations upon masking of speech by white noise," J. Acoust. Soc. Am. **20**, 150–159.

Lindevald, I. M., and Benade, A. H. (**1986**). "Two-ear correlation in the statistical sound fields of rooms," J. Acoust. Soc. Am. **80**, 661–664.

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.

Patterson, R. D., Nimmo-Smith, I., Holdsworth, J., and Rice, P. (**1987**). "An efficient auditory filterbank based on the gammatone function," presented to the Institute of Acoustics speech group on auditory modelling at the Royal Signal Research Establishment.

Patterson, R. D., Nimmo-Smith, I., Holdsworth, J., and Rice, P. (**1988**). "Spiral VOS final report, Part A: The auditory filterbank," Cambridge Electronic Design, Contract Report No. APU 2341.

Peterson, P. M. (**1986**). "Simulating the response of multiple microphones to a single acoustic source in a reverberant room," J. Acoust. Soc. Am. **80**, 1527–1529.

Plomp, R. (**1976**). "Binaural and monaural speech intelligibility of connected discourse in reverberation as a function of azimuth of a single competing sound source (speech or noise)," Acustica **34**, 200–211.

Plomp, R., and Mimpen, A. M. (**1979**). "Improving the reliability of testing the speech-reception threshold for sentences," Audiology **18**, 43–52.

Robinson, D. E., and Jeffress, L. A. (**1963**). "Effect of varying the interaural noise correlation on the detectability of tonal signals," J. Acoust. Soc. Am. **35**, 1947–1952.

Watkins, A. J. (**2005**). "Perceptual compensation for effects of reverberation in speech identification," J. Acoust. Soc. Am. **118**, 249–262.

Wijngaarden, S. J., and Drullman, R. (**2008**). "Binaural intelligibility prediction based on the speech transmission index," J. Acoust. Soc. Am. **123**, 4514–4523.

Zurek, P. M. (**1993**). "Binaural advantages and directional effects in speech intelligibility," in *Acoustical Factors Affecting Hearing Aid Performance*, edited by G. Studebaker and I. Hochberg (Allyn and Bacon, Needham Heights, MA), pp. 255–276.

Zurek, P. M., Freyman, R. L., and Balakrishnan, U. (**2004**). "Auditory target detection in reverberation," J. Acoust. Soc. Am. **115**, 1609–1620.

Effect of age at onset of deafness on binaural sensitivity in electric hearing in humans

Ruth Y. Litovsky,[a] Gary L. Jones, and Smita Agrawal
University of Wisconsin Waisman Center, 1500 Highland Avenue, Madison, Wisconsin 53705

Richard van Hoesel
The HEARing CRC, C/O Audiology, Hearing and Speech Sciences, The University of Melbourne, 550 Swanston Street, Victoria 3010, Australia

(Received 23 January 2009; revised 1 October 2009; accepted 2 October 2009)

Sensitivity to binaural cues was studied in 11 bilateral cochlear implant users, all of whom received both of their cochlear implants as adults, but who varied in the age at onset of deafness, from pre-lingual to childhood-onset to adult-onset. Sensitivity to interaural timing difference (ITD) and interaural level difference (ILD) cues was measured at basal, middle, and apical pitch-matched places of stimulation along the cochlear arrays, using a stimulation rate of 100 Hz. Results show that there is a trend for people whose onset of deafness occurred during adult life or late childhood to retain at least some sensitivity to ITDs, whereas people with onset of deafness earlier in life were insensitive to ITDs. In contrast, ILD cue sensitivity was present in all subjects. There were no effects of place of stimulation on binaural sensitivity, suggesting that there is no indication of a dependence of ITD sensitivity on apical vs basal electrode location.
 [DOI: 10.1121/1.3257546]

PACS number(s): 43.66.Pn, 43.66.Ts, 43.66.Qp [JCM] Pages: 400–414

I. INTRODUCTION

Cochlear implants (CIs) offer a means of providing hearing to deaf individuals through electrical stimulation of the auditory nerve. Traditionally, a single implant in either ear was provided; however, in recent years increasing numbers of patients have received bilateral cochlear implants (van Hoesel, 2004; Nopp *et al.*, 2004; Litovsky *et al.*, 2006a). Bilateral implantation has been motivated by the fact that, in normal-hearing people, binaural hearing plays an important role, for example, by facilitating sound localization and speech understanding in noise. Studies done in the sound field using loudspeakers have shown that at least some of the advantages of having two ears extend to bilateral cochlear implant users. For example, on tasks of sound source location identification, root-mean-square errors can be as low as 10°–20° when listening with bilateral implants and ≥60° when listening monaurally (van Hoesel and Tyler, 2003; Litovsky *et al.*, 2004; Nopp *et al.*, 2004). Similarly, the ability to understand speech in the presence of either noise or competing speech is improved in the majority of patients when they use both of their implants vs one implant alone (Nopp *et al.*, 2004; Schleich *et al.*, 2004; van Hoesel, 2004; Litovsky *et al.*, 2006a, 2009).

Despite these notable improvements, both sound localization and speech understanding in noise abilities of bilateral cochlear implant users are overall significantly worse than performance measured in normal-hearing people. There may be a biological limitation at fairly peripheral levels in the system due to degradation of neural ganglion cells following a prolonged period of auditory deprivation (Leake *et al.*, 1999; Coco *et al.*, 2007), and thus loss of fidelity with which information can be transmitted to binaural circuits. Another factor that is likely constraining performance is hardware-based limitation of access to binaural information.

In normal-hearing people, sound localization abilities in the horizontal plane depend on the extent to which listeners are able to extract and utilize differences in arrival time and level of stimuli at the two ears. Localization of un-modulated signals up to approximately 1500 Hz is known to depend on the interaural time difference (ITD) arising from disparities in the fine-structure of the waveform (for review see Blauert, 1997). The prominent cue for localization of high-frequency signals is the interaural level difference (ILD) cue (Blauert, 1997). However, it has also now been well established that for higher-frequency signals ITD information can be transmitted by imposing a slow modulation, or envelope, on the high-frequency carrier (e.g., Bernstein, 2001). The use of such modulated signals with high-frequency carriers is highly relevant to the coding of speech in cochlear implant processors (Seligman *et al.*, 1984; McDermott *et al.*, 1992; Wilson *et al.*, 1991; Skinner *et al.*, 1994; Wilson and Dorman, 2008a).

Preservation and presentation of binaural cues to bilateral cochlear-implant users pose a unique challenge. Today's clinical processors are not engineered in a way that allows for coordination between the devices in the two ears; thus the binaural cues that would be available to normal-hearing people are not preserved and presented to the auditory system with fidelity. Bilateral cochlear implant users are essentially fitted with two separate monaural systems. Speech processing strategies in clinical processors utilize pulsatile, non-simultaneous multi-channel stimulation, whereby a bank of

[a]Author to whom correspondence should be addressed. Electronic mail: litovsky@waisman.wisc.edu

bandpass filters is used to filter the incoming signal into numerous frequency bands (ranging from 12 to 22) and to send specific frequency ranges to individual electrodes. The envelope of the signal is extracted from the output of each band and is used to set stimulation levels for each frequency band; however, fine-structure information is discarded. An additional factor is that the rate at which electrical pulses are presented to selected electrodes is fixed, and not necessarily related to the characteristics of the acoustic stimulus. While it is possible that the envelope cues can provide some interaural difference cues that are usable and that lead to binaural advantages, interaural difference cues that arise from fine-structure information in the signal are not available (for review see van Hoesel, 2004; Wilson and Dorman, 2008b).

We investigated binaural processing in electrical hearing using an approach that bypasses the clinical processors and provides direct control over stimuli that are provided to each electrode, and the relative timing and levels of these stimuli. Previous experiments with electrical pulse-trains applied to electrode pairs, in which implant users were required to detect changes in ITD, have reported that discrimination thresholds for stimuli presented at low rates of about 100 Hz can be as low as 50–200 μs in some patients but are up to an order of magnitude larger in others (van Hoesel *et al.*, 1993; Lawson *et al.*, 1998; Majdak *et al.*, 2006; van Hoesel, 2007). Thresholds from best-performers in the cochlear implant population overlap with thresholds seen in moderately trained normal-hearing people ($\sim$70 μs) presented with low-frequency tones carrying ITD information (Blauert, 1997; Bernstein, 2001; Wright and Zhang, 2006).

Our work differs from prior studies in a number of ways. First, rather than testing discrimination abilities for stimuli presented to the right vs left, binaural parameters that are associated with a range of locations in space in acoustic hearing were used in a lateralization task in which listeners reported a perceived intracranial position of the sound source for various ITD or ILD values. This approach, which has been used to some extent over the years by van Hoesel and colleagues (van Hoesel *et al.*, 1993; van Hoesel and Tyler, 2003; van Hoesel, 2008), offers a more direct estimate of the degree to which binaural cues contribute to perceived source locations.

Second, prior studies in this area have utilized small numbers of patients. Furthermore, most recent studies have the added criterion of deliberate exclusion of subjects with poorer than average ITD sensitivity, the reason being that effects of electrical stimulation per se, rather than unknown individual factors, were of interest (e.g., Laback *et al.*, 2007; van Hoesel, 2008). Although these criteria are reasonable for studies that seek to focus on best performance achieved by CI users, they offer less opportunity for understanding the applicability of such findings to the general population of bilateral CI users. Individual variability in performance is one of the most challenging hallmarks of research in this area. One of the known sources of variability is the age at which the onset of deafness occurred. Patients can vary from having experienced onset of deafness prior to language acquisition, after language acquisition but still during childhood, or during adulthood. In addition, in the childhood- and adult-acquired groups, the activation of hearing with CIs can often occur many years after onset of deafness; thus patients vary in the amount of auditory deprivation experienced between the time of cessation of acoustic hearing and time of activation of electric hearing. Typically, patients with adult-onset of deafness have been studied more extensively with regard to binaural sensitivity. In the present study, by also including people with early-onset and childhood-onset of deafness, we were able to examine the extent to which early vs late onset of deafness is an important determinant for successful use of binaural cues with electric hearing.

Third, few published data are available on the importance of place of stimulation along the cochlear electrode array for binaural sensitivity. It may be possible that variation in place of stimulation is a potential factor in the variability observed in ITD sensitivity in prior studies. In this study, place of stimulation was varied to include basal, middle, and apical regions along the electrode array. Since the electrode arrays are essentially confined to the basal turn, the use of the terms *apical*, *middle*, and *basal* indicate relative positions within the basal turn. This approach enabled us to evaluate whether ITD coding at different places of stimulation is processed similarly. In normal-hearing listeners "transposed stimuli" have been used to compare ITD sensitivity for different places of stimulation along the cochlea (Bernstein and Trahiotis, 2002; Oxenham *et al.*, 2004), although mechanical frequency coding differences may preclude sufficient matching of the spatio-temporal response patterns in different regions, an issue that cannot be easily resolved with acoustic stimulation. The use of electric stimulation circumvents this issue by eliminating systematic difference in peripheral responses at high- and low-frequency regions along the cochlea (van Hoesel *et al.*, 2009).

II. METHODS

A. Subjects

Eleven adults participated in the study. All were successful users of bilateral cochlear implants who relied on their implants for everyday communication and provided verbal reports regarding noticeable changes in quality of life with bilateral implants compared to single-implant use. All subjects wore Nucleus-24 or 24M devices in both ears. Table I shows demographic information for each subject, including age at the time of testing, etiology, age at onset of hearing loss, and duration of cochlear implant use in each ear, or of hearing aid use prior to implantation. All subjects received both of their cochlear implants as adults, and as Table I indicates, many of the subjects had experienced hearing loss for numerous years prior to being implanted. The most relevant category for this study is that of age at the onset of deafness, divided into sub-groups of adult, mid-childhood, and pre-lingual. All subjects were native speakers of English and had bilateral listening experience for a minimum of 6 months. Speech processors in everyday use were either bilateral body-worn SPrint or bilateral ear-level ESPrit processors. All subjects traveled to Madison, WI where testing took place for 6–8 h a day for 2–3 days at a time.

TABLE I. Subject details are organized by age-at-deafness group. For each subject the following details are included, if available: age at time of testing, age at onset of hearing loss in each ear, type of hearing loss, etiology if known, duration of cochlear implant use in each ear, and whether a hearing aid was used in each ear.

Age at deafness	Subject code	Age at testing (yr)	Age at onset of hearing loss		Type of hearing loss	Etiology	Duration of cochlear implant use		Hearing aid use	
			Right	Left			Right	Left	Right	Left
Adult	IAD	48	18 yrs	46 yrs	Sudden	Unknown	16 mo	16 mo	No	No
	IAK	58	27 yrs	27 yrs	Sudden	Ototoxicity	14 yr	2.2 yr	Yes	Irregular
	IAN	55.5	38 yrs	38 yrs	Progressive	Hereditary	6 mo	7 yr	Yes	Yes
	IAP	64	24 yrs	24 yrs, deaf by 35 yrs	Progressive	Meniere's syndrome, hereditary	5 yr	5 yr	Yes	Yes
Mid-childhood	IAF	58	Mid-childhood, deaf by mid-20s	Mid-childhood, deaf by mid-20s	Progressive	Mumps/measles/ chicken pox	1 yr	7 yr	Until age 47	Until age 47
	IAH	48.5	Mid-childhood	Mid-childhood	Progressive	Hereditary	2.3 yr	2.3 yr	Yes	Yes
	IAJ	59	Mid-childhood to adolescence	Mid-childhood to adolescence	Progressive	Unknown	9 mo	7.7 yr	Until age 15	Until age 15
	IAR	50	Mid-childhood	Mid-childhood	Unknown	Unknown	3.3 yr	4.7 yr	Yes	Yes
Pre-lingual	IAE	45	3 yrs	3 yrs	Unknown	Unknown	7 mo	6 mo	Yes	10–12 yrs
	IAG	49	At birth	At birth	N/A	Unknown	11 mo	5.7 yr	Yes	Yes
	IAI	49	At birth	At birth	N/A	Unknown	3 yr	7.8 yr	Yes	Yes

B. Stimuli

All the tests were done by directly activating the electrodes using current levels appropriate for each individual subject. Stimulation was achieved using a custom built research processor (Spear III, The Hearing CRC) that transmits electrical stimulation to the patient's receiver coils and that is able to control binaural pulse timing within 2.5 μs for chosen pairs of electrodes. Figure 1 shows a schematic of the electrode numbers, 1 and 22 being the most basal and apical electrodes on each electrode array, respectively. In the cochlear implant system used here, physical spacing between adjacent electrodes is 0.75 mm. Stimuli consisted of monopolar constant-amplitude biphasic current-pulses with extra-cochlear reference electrodes. Current pulses were gated on and off instantaneously and presented at a rate of 100 pulses per second (pps) for a total duration of 300 ms per stimulus. Pulse width was 25 μs per phase for 8/11 subjects (40 for IAI and 50 for IAF and IAG). The 100 pps stimuli were selected based on previous research showing that ITD sensitivity is likely to be strongest at lower stimulation rates (e.g., van Hoesel *et al.*, 2009). It is likely that at the low rate of 100 pps used here ongoing cues provide a strong cue for ITD discrimination in listeners displaying sensitivity to ITDs (e.g., van Hoesel, 2007; Laback *et al.*, 2007; van Hoesel *et al.*, 2009).

C. General procedure

All experimental procedures followed the regulations set by NIH and were approved by the University of Wisconsin's Human Subject IRB. Subjects reside outside of the Madison, WI area and were thus required to travel to Madison, at which point they stayed in town for several days and spent

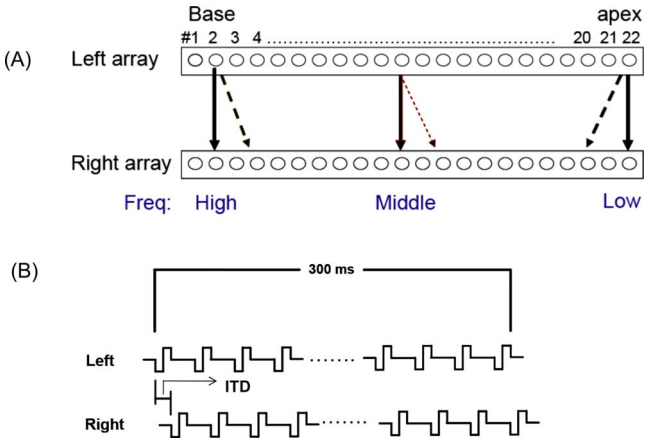


FIG. 1. (Color online) Schematics of the stimuli used in this study are shown. Panel (A) shows the left and right cochlear arrays with 22 intracochlear electrodes. The potential combinations of stimulation places along the array (high-frequency at the base, mid-frequency in the middle, and low-frequency at the apex) are denoted by the arrows. Dashed arrows indicate that pitch-matched electrodes could consist of pairs of electrodes that were not always the same electrode number along the left and right arrays. Panel (B) shows a schematic of the electrically pulsed signals that were presented on each trial. Left- and right-ear stimuli are shown on the top and bottom rows as a function of time. Stimuli were 300 ms in duration, with biphasic pulses presented at a rate of 100 Hz. The interaural time difference can be seen in the difference between onsets of the left- and right-ear pulses.

the majority of each day in the laboratory. Testing was conducted in blocks lasting 20–40 min, with breaks between blocks and long breaks for lunch.

Each experiment began by establishing threshold and maximal comfort levels before discomfort for 20 electrodes (even-numbered electrodes in the two cochleae). Current levels in the 20 electrodes were adjusted to yield equal loudness at a stimulation level near 90% of the dynamic range, comparable to everyday speech levels according to subjective reports. Pairs of electrodes in the two cochleae were selected based on subjective pitch-matching. Subjects were trained to perform a pitch-magnitude-estimation (PME) task, with training lasting approximately 1–2 h, whereby they labeled the perceived pitch of electrical stimulation at various cochlear locations on a numerical scale. Once it was established that PME performance was stable and consistent, PME was measured experimentally for the even-numbered electrodes in each ear, ranging from 2 to 22. A blocked randomized design was used (2 ears $\times$ 10 electrodes per ear $\times$ 10 repetitions per electrode). Once testing was completed, all PME data were sorted, and the electrodes with the most similar PME values in the right and left cochleae were selected as pitch-matched electrode pairs. Prior work by van Hoesel (van Hoesel *et al.*, 1993; van Hoesel and Tyler, 2003; van Hoesel *et al.*, 2009) has shown that this method for selecting electrode pairs yields similar results for binaural sensitivity to methods in which place-matching based on x-ray views are used. Selection of best-matched electrode pairs was further adjusted and sharpened using feedback from the participants about pitch similarity for various electrode combinations. The outcome of this elaborate procedure was the identification of three pairs of pitch-matched electrodes for each subject located in the basal, middle, or apical regions of the electrode array within the basal turn. For each electrode pair subsequently used in testing a second stage of stimulus selection took place. When pairs of electrodes that were subjectively reported to be loudness-balanced were activated simultaneously with an ITD of 0 μs, the extent to which the subject heard a fused auditory image with an intracranial position at the center of the head was determined. In the event that the image was un-fused testing was not conducted using that electrode pair. When there was an audible fused image, the perceived location was typically displaced intra-cranially toward the right or left. This perceived position could be shifted by lowering the amplitude of the electrode on the side of the head dominating perception. For instance, if the sound was perceived as right of the midline and the current level unit (CLU) on the right was then lowered by a few current level units the auditory image would shift toward center. Once a fused centered image was established for a given electrode pair, non-zero ITDs or ILDs were imposed, and lateralization judgments were obtained for each cue separately.

Prior to initiating testing on the lateralization task, subjects received training lasting approximately 1 h. There was emphasis placed on determining whether stimulation of each electrode pair produced an auditory image whose positions varied for different values of ITDs and ILDs. Initially, stimuli with large values of ITDs or ILDs were presented one at a time, and subjects were asked to report the perceived intracranial location of the auditory image. This was repeated for smaller values of ITDs and ILDs so that subjects understood that perceived locations could vary. However, feedback was not provided since there is no correct or incorrect response on these types of measures. As is described below, some subjects under some conditions were unable to perceive fused auditory images whose intracranial position could be reported.

During testing, on each trial, the perceived location of the auditory image within the head was measured using a visual pointer on the screen. Subjects faced a computer monitor and viewed a display depicting a horizontal line spanning about 20 cm across the center of the screen. The horizontal line was bisected in the middle with a small vertical line and was marked with the letters L and R on the left-most and right-most edges. Subjects were instructed to treat the horizontal line as the range of intracranial auditory images that span from their left ear to the center of the head and over to the right ear. Subjects used a computer mouse to click on a spot of their choice on the horizontal line that corresponded to perceived intracranial image. The response options on the horizontal lines were continuous, and responses were stored as numerical values that ranged from 0 (perceived near left ear) to 50 (perceived in center of the head) to 100 (perceived near right ear).

D. Testing sensitivity to binaural cues

Testing was conducted in blocks of trials in which the pair of electrodes being stimulated was held constant. The order for selecting electrode pairs and for testing ITD and ILD sensitivity for a given electrode pair was random. At the beginning of testing with each electrode pair, the general procedure described above for ensuring a fused, centered auditory image near 90% of the dynamic range was followed. For each electrode pair, either ITD or ILD values were varied within a block, with 20 trials per cue value. Prior to testing with the randomized block design, extensive pilot testing was conducted to determine what range of ITDs and ILDs would maximize listeners' perception of lateralized images with each pair of electrodes. It was sometimes observed that values that were completely lateralized to one ear or the other prior to testing were no longer completely lateralized during testing. Some listeners (in the prelingual group with ITDs, in particular) had difficulty determining left vs right even with ITDs on the order of several milliseconds, but these values were used nonetheless when the experiments were carried out. ITD cue sizes typically ranged from ± 25 to ± 1000 μs. For subjects that could clearly perform the task at values lower than 800, the typical set of values was 25, 50, 100, 200, 400, and 800. If subjects' performance was not above chance at ITD=1000 μs we used values greater than 1000 μs. ILD values generally varied from ± 1 to ± 20 CLUs, translating to ± 0.17 to ± 3.5 dB in stimulation current, noting that a change of 1 CLU corresponds to $20 \log 10(175)/255 = 0.176$ dB in stimulation current.[1] In some cases, very large values were used. In one case (subject

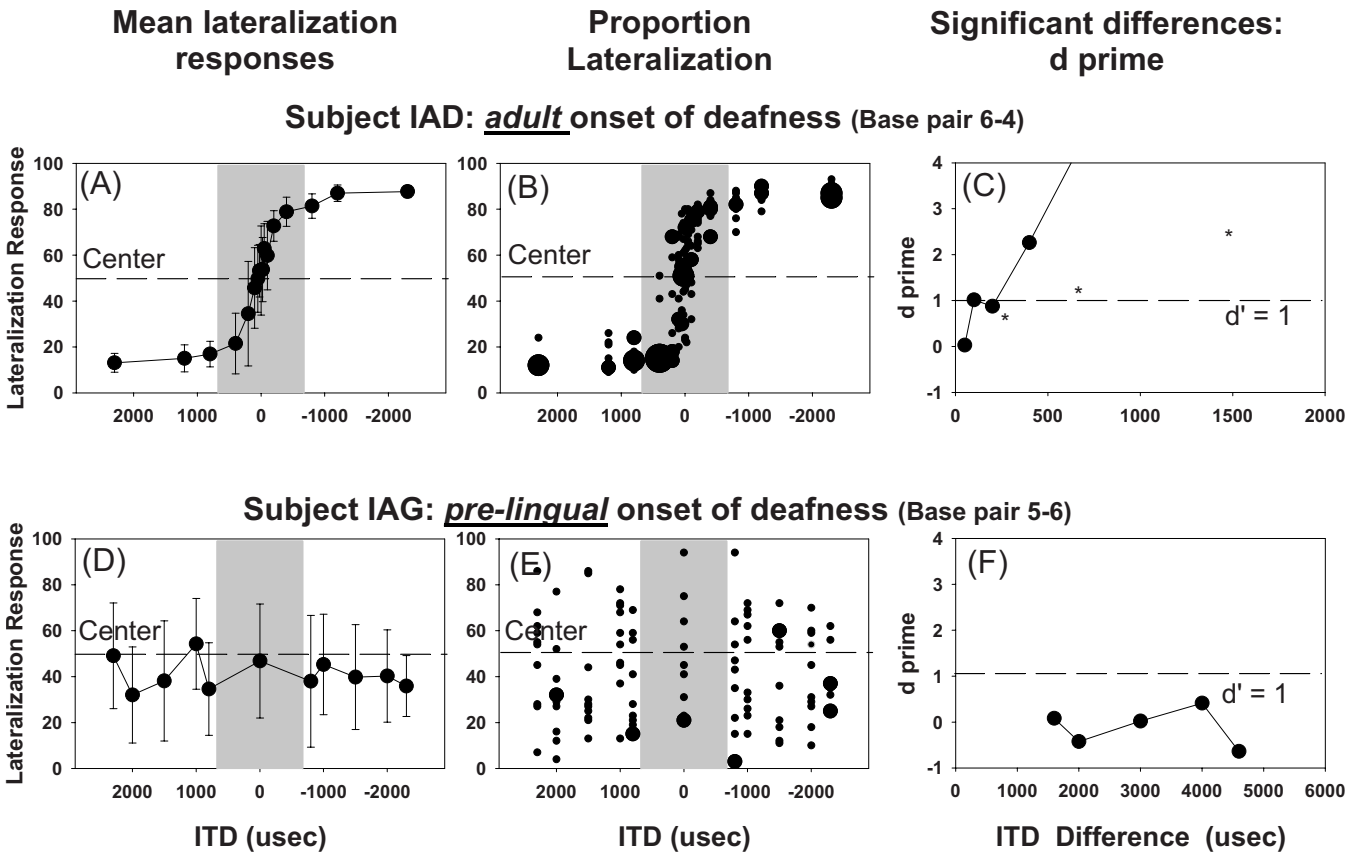


FIG. 2. (Color online) Results from ITD experiments are shown for two subjects, one with adult-onset of deafness [(A)–(C)] and one with pre-lingual onset of deafness [(D)–(F)]. Panels (A) and (D) show the average perceived position (±SD) for each interaural timing difference value tested on the lateralization pointer task. Panels (B) and (E) show the proportion of trials on which each perceived location (lateralization response) was selected for each value of ITD. In (A), (B), (D), and (E) the convention is to use positive and negative values on the abscissa to represent left and right leading stimuli, respectively; the shaded area represents the range of stimulation values that are physiologically relevant and are known to produce auditory images that are fully lateralized for normal-hearing listeners (±700 μs). On the ordinate, lateralization response refers to the value representing the perceived intracranial location of the auditory image as indicated during the pointer task during testing. Panels (C) and (F) show the d' values obtained at each ITD magnitude, that is, absolute difference in right vs left ITD. The dashed line is drawn at $d'=1$ which is the value used for threshold criterion throughout the manuscript.

IAI, apical pair), data were not collected because the subject was unable to return for further testing due to personal circumstances.

The data consist of individual listeners' lateral position judgments. The "raw" data are presented as the average (±SD) perceived lateral position for specific values of binaural cues. In addition, these data were subjected to statistical analyses in which comparisons were made for each condition (combinations of pairs of ITD values, pair of electrodes, for every participant). In order to conduct these analyses, responses were first normalized by applying an arcsin transformation; then for each left-right stimulus value pair (e.g., +400 and −400 μs) the distance between the two distributions of responses for the left and right stimulus values was calculated. At each left-right stimulus value the statistic d' was computed as the difference between the two means divided by a pooled estimate of their standard deviations. For each condition, a best-fit line was calculated for the d' values for that condition calculated as described above, with the fit constrained to pass through zero. At high values d' becomes more variable; thus only d' values less than or equal to 4 were included in the calculation of the best-fit line. Threshold was defined as the point where the line from the best fit intersected with the value $d'=1$. It is important to note that the values plotted here refer to two times the magnitude of ITD and ILD relative to values for which an auditory image that was perceived to be centered. In the case of ITDs the relative value was 0 μs and known to be perceived in the center of the head from initial testing and stimulus parameter determination (see above). In the case of ILDs the value of 0 refers to the current levels in the two ears that produced a perceptually centered image. A threshold of ITD=400 μs on one of our plots indicates that the value $d'=1$ was reached when the left and right ITDs each had a value of 200 μs. An ILD threshold of 4 CLUs on our plots indicates that the value $d'=1$ was reached when the left and right ILDs each had a value of 2 CLUs.

III. RESULTS

ITD lateralization data are shown for two listeners in Fig. 2. What differentiates these listeners from one another is the age of onset of deafness, with subject IAD (top) in the childhood-onset group and subject IAG (bottom) in the prelingual-onset group. For subject IAD there is a clear indication that varying the ITD values within the physiologically relevant range of ±700 μs [see shaded gray area, panel (A)] results in perceived locations that spread throughout the

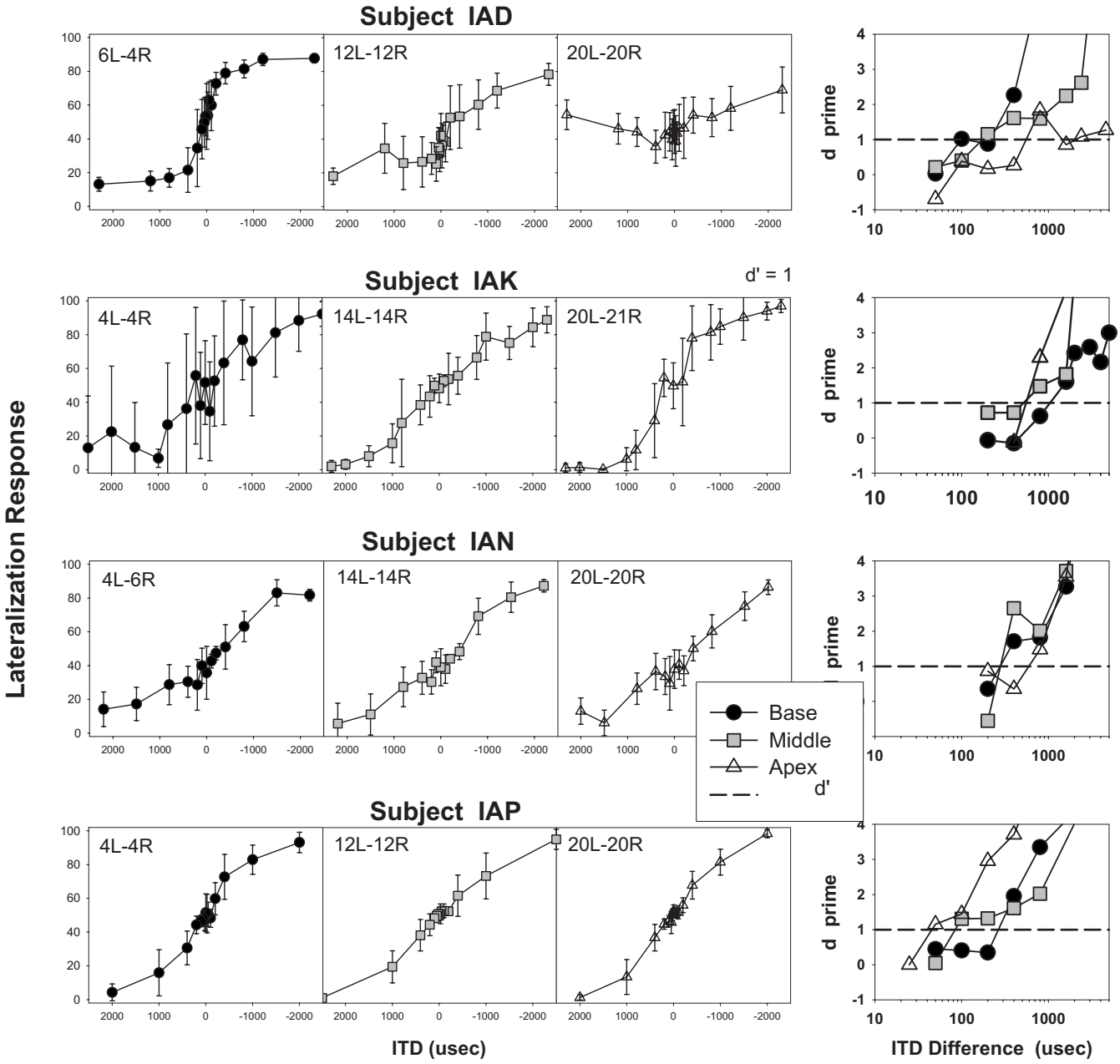


FIG. 3. Results from the ITD task are shown for the four subjects with adult-onset deafness. For each subject the average perceived position (±SD) for each ITD value tested is plotted separately for the three electrode pairs tested, approximately in the base, middle, and apical regions of the arrays; the paired numbers within each panel indicate the left-right electrode numbers for each pair tested. The right-most panels for each subject show d' values as a function of the left-right ITD tested for the three electrode pairs.

range of intracranial positions, from far left to far right. Performance is markedly different for subject IAG; although this listener reported that when the pair of electrodes used here (14L and 14R) was activated a fused auditory image was perceived, the location of the image does not vary systematically along an intracranial dimension as a function of ITD. Panels (C) and (F) plot the d' values obtained at each ITD size tested. Subject IAD's performance hovers near $d'=1$ at small ITD values and reaches this value consistently when the absolute difference in right vs left ITD is 400 μs, with threshold estimated as 213 μs based on a best-fit line calculation. In contrast, the fact that subject IAG is unable to perform the task regardless of the ITD cue size is reflected in d' values smaller than 1 at all ITDs tested, consistent with there being no discriminable difference between the ITDs favoring the right vs left ears. This was true even at "super" ITD values on the order of a few milliseconds. For this listener, whose onset of deafness was thought to have been present at birth, we observed that ITDs do not evoke a lateralized image nor one with a perceived location. Anecdotal reports from this subject suggest that stimuli with ITDs generated images that were broadly distributed within the head and thus not consistently lateralized in either direction.

The variability in performance on the ITD lateralization task occurs both amongst and within people who use bilateral cochlear implants, as is illustrated in greater detail in Figs. 3–5, where lateralization data are plotted for people with onset of hearing loss during adulthood (Fig. 3), mid-

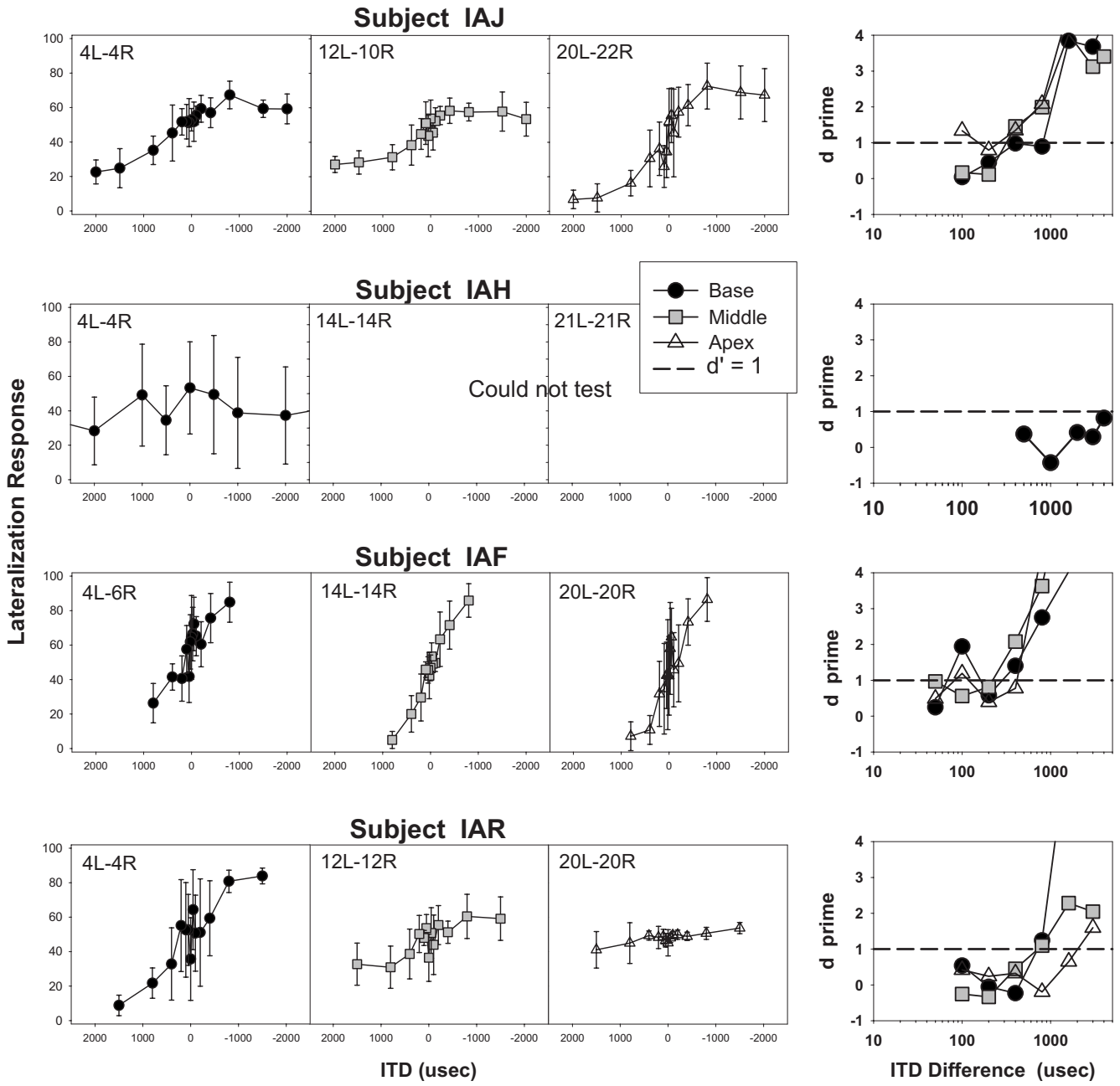


FIG. 4. Same as Fig. 3 for subjects with mid-childhood-onset of deafness.

childhood (Fig. 4), or prelingually (Fig. 5). In each figure lateralization data are plotted for basal, middle, and apical places of stimulation. The right-most plots in each figure also show d' calculated for each value of the absolute difference in right vs left ITD (see captions for detail). All four people with bilateral CIs whose onset of deafness occurred during adulthood are able to use ITDs to perceive changes in the lateral position of the intracranial source image. Except for subject IAD's performance with apical stimulation, lateralization spanned nearly the entire range of responses, suggesting that subjects were able to use ITDs effectively to distinguish intracranial source positions.[2] Subject IAD also had trouble with the middle electrode pair but had excellent performance with basal stimulation. It is worth noting that of the four subjects with adult-onset of deafness, this was the only person with long-term asymmetric hearing loss (see Table I).

With regard to the effect of place of stimulation, it appears that performance on this task is not systematically best or worst at any of the places. Unlike IAD whose performance deteriorated as stimulation progressed from base to apex, for other listeners there were assorted best/worst trends; these can be viewed most clearly in the right-most panels where d' values are plotted for all three conditions for each listener. For instance, listener IAP's performance with apical stimulation is notably better than that with basal stimulation. These findings suggest that ITD lateralization is not restricted to regions along the cochlea that respond best to low-frequency signals.

Regarding subjects whose onset of deafness occurred in mid-childhood, it is clear that they are generally able to use ITDs for lateralization. Inter-subject variability is a bit greater than those with adult-onset; whereas subject IAF

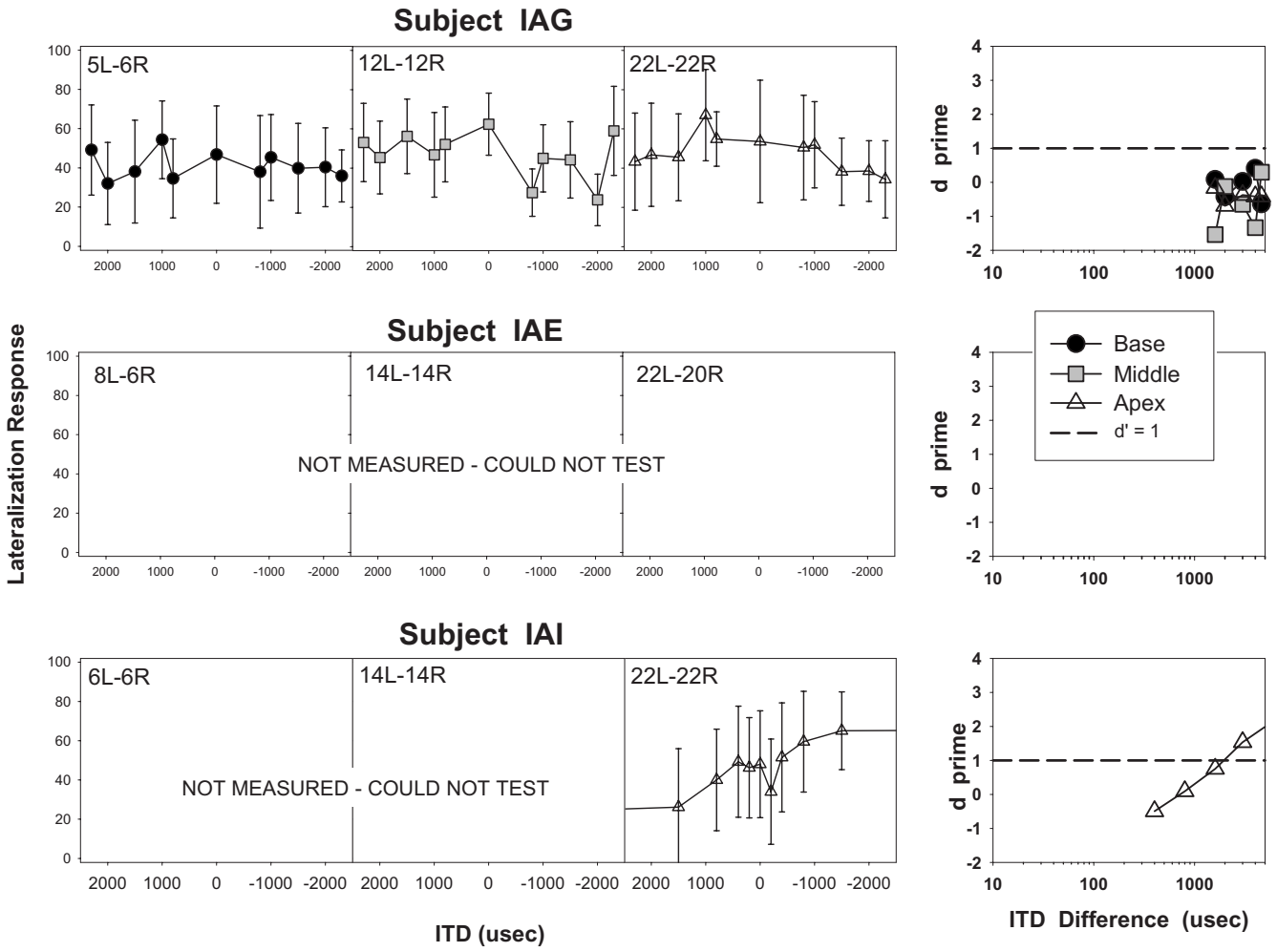


FIG. 5. Same as Fig. 3 for subjects with prelingual-onset of deafness.

(Figs. 2 and 4) was able to perform as well as people with adult-onset of deafness, IAH performed poorly at the only electrode pair for which testing was possible.[3] IAJ and IAR were able to perform the task but the range over which the auditory images were perceived to be lateralized was more restricted. Finally and most notably, subjects whose onset of deafness occurred at birth or prior to age 3 (Fig. 5) were virtually unable to use ITDs on this task, even with "super size" ITD values on the order of several milliseconds. As described above in the statistical analyses section, lateralization thresholds defined as $d'=1$ were computed for each electrode pair. These values are plotted by subject group in Fig. 6, where the effect of age at onset of deafness is clear, and the lack of trend for place of stimulation is evident as well. Individual thresholds from Fig. 6 are listed in Table II, along with ILD thresholds (see below). ITD data from the prelingual group fall into one of two categories. As described in Sec. II, some subjects under some conditions were able to perform the task; however, thresholds were >3000 and were thus not accurately determined. These data points are identified in Fig. 6 as "could not determine" (CND). A good example is the apical data point for subject IAI; Fig. 5 contains a curve from measured data, but the ITD at which d' reached 1 was above 3000 μs. In other cases, testing was attempted but no data were obtained due to extreme difficulty of the

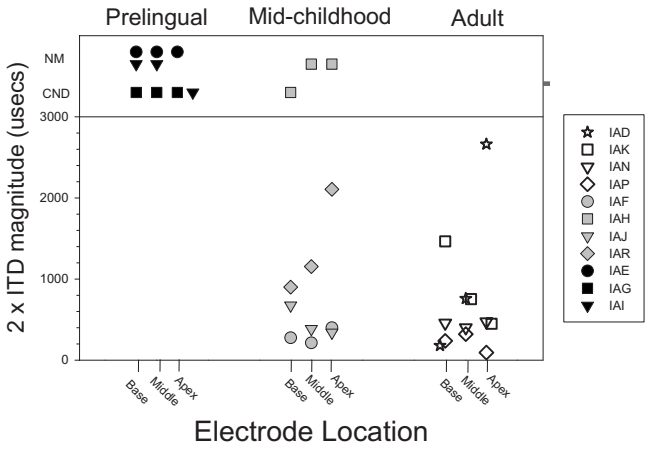


FIG. 6. Summary of results from ITD sensitivity measures is shown grouped by age at onset of deafness from left to right: prelingual, mid-childhood, and adult. Each data point represents a single subject's ITD threshold defined as the sum of the microsecond values for the right and left stimuli at which performance reached $d'=1$; in other words, the ITD value favoring either right or left, multiplied two times. Within each group, results are shown for the three places of stimulation along the cochleae tested (base, middle, and apex). The dashed horizontal line at the top demarcates between data points for which exact threshold was obtained and data points that could not be determined (CND), and data points that were not measured (NM) due to task difficulty (see text).

TABLE II. Thresholds obtained for each subject at each condition tested. Values denoted as CND refer to conditions under which testing was conducted but thresholds were extremely high and could not be determined. Values denoted as NM refer to conditions under which testing was initiated, but thresholds were not obtainable due to extreme difficulty of the task (see text).

	ITD threshold ($2\times\mu$s)			ILD threshold ($2\times$CLU)		
	Basal	Middle	Apical	Basal	Middle	Apical
			Adult onset			
IAD	180	759	2662	1.41	0.53	0.64
IAK	1464	753	448	1.43	1.98	1.42
IAN	458	400	474	0.92	0.97	1.00
IAP	237	321	94	0.29	0.61	0.48
			Childhood onset			
IAF	276	213	399	1.07	0.71	1.11
IAH	NM	NM	CND	0.81	1.86	1.32
IAJ	675	384	340	1.42	1.12	0.51
IAR	900	1154	2106	1.77	0.87	0.69
			Prelingual			
IAE	NM	NM	NM	0.29	0.50	0.66
IAG	CND	CND	CND	5.61	3.02	4.05
IAI	NM	NM	CND	1.36	1.53	Not measured

task; hence these data are identified as "not measured" (NM). From a statistical perspective CND and NM are essentially the same as in both cases the data are censored, i.e., values of these observations fell outside the range of values that could be determined. Nonetheless, we consider censored data as being meaningful; while they do not provide information regarding thresholds they are informative regarding the subject's ability to utilize the cues that were manipulated in the experiment.

Statistical analyses consisted of non-parametric tests because there are known limitations to parametric statistics when the N size is small as is the case here and when non-measured or undetermined data points are considered as missing. Subjects were categorized into a 3×2 table of group membership by censorship. Fisher's exact test for dependence was significant [two-tailed p-value$=0.030$], thus establishing a relationship between group and censorship, whereby data points from patients with early-onset of deafness were censored more than the other patients. In addition, a test of linear-by-linear association, conducted to evaluate whether there is a linear trend across groups, revealed a significant effect [$X^2(1)=6.433$, $p=0.015$], implying increased likelihood of censorship as subject group progressed from adult-onset to childhood-onset to prelingual-onset of deafness. To evaluate possible condition effects, Friedman's test was applied to the data collected for uncensored subjects, revealing non-significant findings [$X^2(2)=0.286$, $p=0.867$]. In summary, the statistical analyses conducted here reveal significant effects of age at onset of deafness on ITD sensitivity, but no effects of place of stimulation.

Parallel data from the ILD experiments are shown in Figs. 7–11, where lateralization scores are plotted as a function of CLU offsets to the right or left; 1 CLU corresponds to 0.176 dB in stimulation current (see also footnote [1]). Examples from the same two individuals for which ITD data were shown in Fig. 2 are seen in Fig. 7 for ILDs. Subject IAD performed remarkably well on this task, as can be clearly seen in Fig. 7. Worth noting is that error bars for this listener were smaller with ILDs than ITDs, suggesting that perhaps ILDs provided a stronger cue for lateralization. For this listener, ILDs at which $d'=1$ could barely be measured due to limitations of stimulus presentation at very small CLUs. In contrast, subject IAG showed poorer ability to utilize ILDs for lateralization. On the one hand, although this listener reported hearing a fused auditory image that was perceived to be generally lateralized in the presence of ILDs, her ability to pinpoint specific intracranial positions was poor. In addition, even at large ILDs the auditory event was not greatly shifted in either direction. Nonetheless, it is worth noting that subject IAG was at least able to perform the task with ILD cues, albeit not well, whereas she had been unable to perform the task with ITD cues.

While subject IAG was selected as an example of a listener with prelingual-onset of deafness whose use of ILDs was poor, there are cases of subjects in this study with prelingual onset of deafness who demonstrate very good lateralization ability using ILDs. Results from the three subject groups are shown in Figs. 8–10 in the same order of appearance as the ITD data: adult-, childhood-, and prelingual-onsets of deafness. Unlike the ITD task which depended on the age of onset of deafness, all subjects were able to perform the ILD task. Because the CLU values required to achieve lateralization were somewhat large for some of the listeners, we chose to plot the results with a large range of values on that axis, which compresses the data points for one of the subjects. Nonetheless the equal-range approach allows for direct comparison of performance across all listeners. One notable difference between ITD and ILD lateralization abilities is the greater variation in performance across electrodes with ITDs. For instance, in the adult-onset group, subject IAD who struggled to lateralize the auditory images with middle and apical stimulation using ITDs (see Fig. 3) is able to hear the sound images at the full range of intracranial positions for all places of stimulation with ILDs (see Fig. 8). Similarly, in the childhood-onset group subject IAH is a good example of someone for whom ITDs were not easily accessible on this task (see Fig. 4), whereas ILDs provided an excellent lateralization cue (see Fig. 9). For IAJ and IAF ITDs and ILDs are similarly useful, and IAR seems to improve with ILDs at least with apical stimulation. The change in performance with the ILD task is most markedly seen in the subject group with prelingual onset of deafness. Results in Fig. 10 suggest that all three subjects were able to utilize ILD cues on the lateralization task (note that apical data are missing for subject IAI). In contrast with the other two groups of subjects, however, listeners in the prelingual group never seemed to lateralize the auditory image all the way toward either ear, even at very large interaural CLU differences. ILD thresholds are summarized in Fig. 11 to parallel the summary presented for ITDs in Fig. 6. Individual thresholds from Figs. 6 and 11 are listed in Table II. As with the

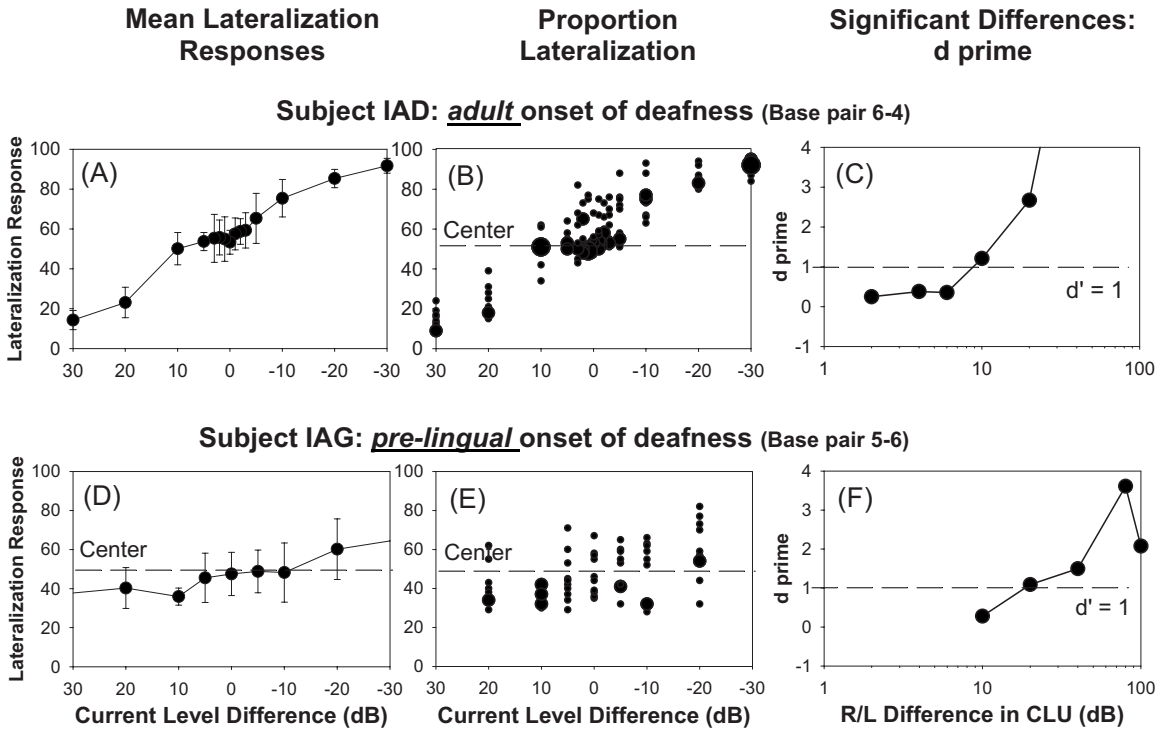


FIG. 7. Results from ILD experiments are shown for two subjects, one with adult-onset of deafness [(A)–(C)] and one with pre-lingual onset of deafness [(D)–(F)]. The ordinate refers to CLUs, noting that a change of 1 CLU corresponds to $20\log 10(175)/255=0.176$ dB in stimulation current (see also footnote [1]). Panels (A) and (D) show the average perceived position ($\pm$SD) for each interaural level difference value tested on the lateralization pointer task. These are the same subjects for whom ITD data are shown in Fig. 2. Panels (B) and (E) show the proportion of trials on which each perceived location (lateralization response) was selected for each value of ILD. In (A)–(E) the convention is to use positive and negative values on the abscissa to represent left and right leading stimuli, respectively. Panels (C) and (F) show the d' values obtained at each ILD magnitude, that is, absolute difference in right vs left ILD. The dashed is drawn at $d'=1$ which is the value used for threshold criterion throughout the manuscript.

ITD data, analyses were conducted using non-parametric tests. Effect of condition was analyzed using the Friedman test. All 11 subjects were included with the one missing data point excluded, revealing no significant effect [$X^2(2)=0.545$, $p=0.761$]. A Kruskal–Wallis test to establish whether there exists a relationship between group and rank mean ILD score revealed a non-significant effect [$X^2(2)=1.326$, $p=0.553$]. The results suggest that ILD thresholds measured here did not appear to depend on the age at onset of deafness nor on the place of stimulation along the cochlear array. One subject (IAG) however, had very high ILD thresholds compare to the rest of the subjects; thus the statistical analyses will be interpreted with caution.

IV. DISCUSSION

In this study we investigated sensitivity to binaural cues using an experimental approach that enabled us to control ITD and ILD cues presented to selected pairs of electrodes along the cochlear array of bilaterally implanted CI users. Two primary observations are worth noting. First, participants in whom the onset of deafness occurred at a very young age were generally able to use ILD cues, but their ability to use ITD was severely compromised. In contrast, there was a trend for listeners with onset of deafness during childhood or adulthood to be able to utilize both ITD and ILD cues. Second, while there appeared not to be a systematic effect of place of stimulation on lateralization of ITD or ILD cues, individual subjects had better performance with some electrode pairs than others.

A. Binaural sensitivity and effects of auditory deprivation

In general, binaural sensitivity measures from the lateralization data are in agreement with published results using the discrimination paradigm (e.g., Long *et al.*, 2006; van Hoesel and Tyler, 2003; van Hoesel, 2007; Laback *et al.*, 2007) in which ITD thresholds in bilaterally implanted CI users range from tens of microseconds to a few hundred microseconds. Here we examined this variability in the context of the age at the onset of deafness. Sensitivity to ITDs seemed to be more affected by long-term deprivation of auditory input than was sensitivity to ILD. The three prelingually deafened individuals were unable to utilize binaural timing cues to subjectively judge the perceived intracranial position of stimuli containing ITDs; sensitivity to ITDs was also problematic for one individual with childhood-onset of deafness. However, all patients with adult-onset deafness were able to lateralize sounds with right-left differences in ITDs ranging from ~90 to 400 μs for at least one place of stimulation. When considering ILD sensitivity, subjects who had experienced auditory deprivation early in life did not show overall more susceptibility to disruption compared to subjects with onset of deafness later in life. This trend is in agreement with a previous report with five subjects who had

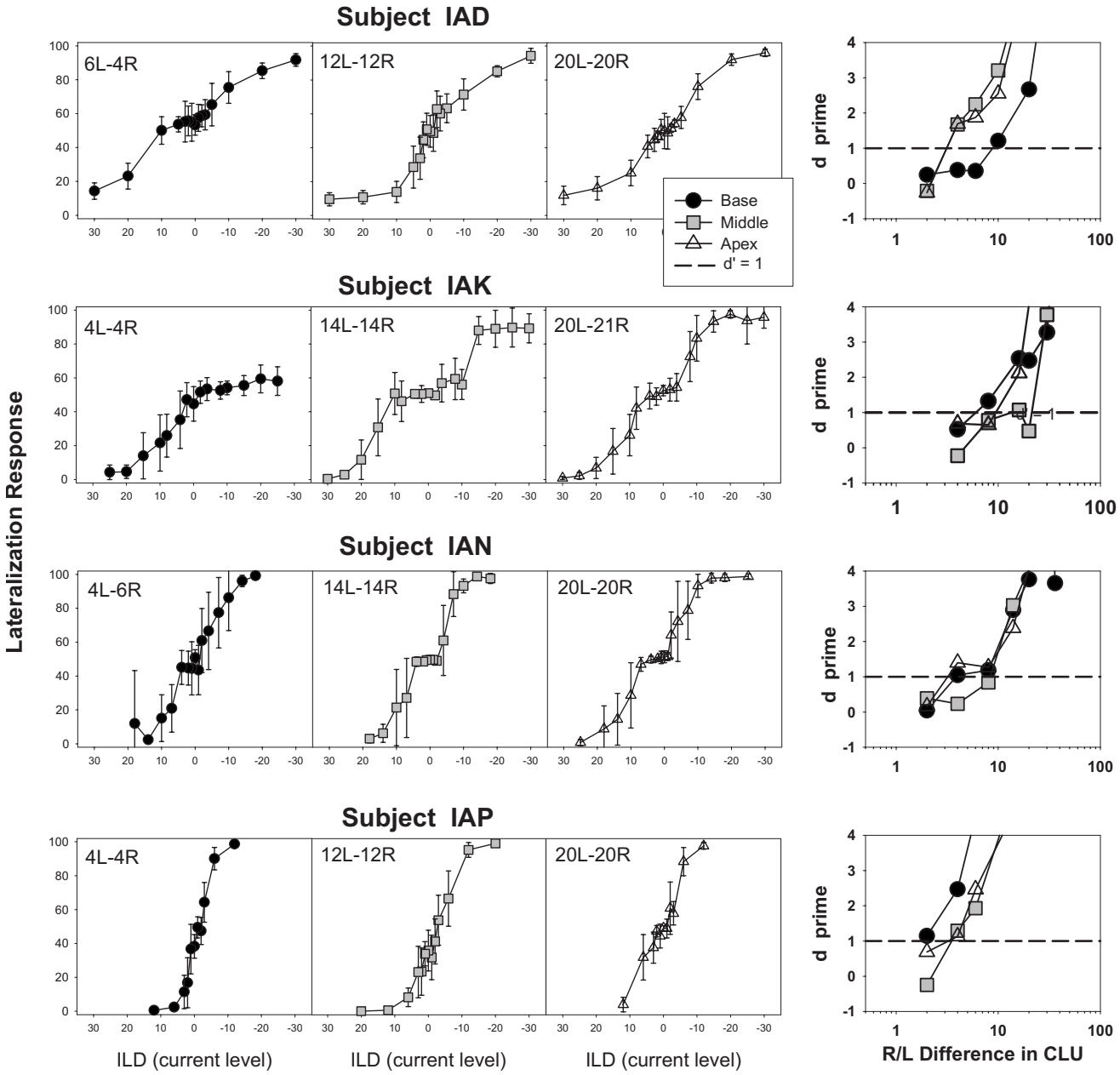


FIG. 8. Results from the ILD task are shown for the four subjects with adult-onset deafness. For each subject the average perceived position (±SD) for each ILD value in right-left CLU interaural difference is shown, noting that change of 1 CLU corresponds to 0.176 dB in stimulation current. Each ILD tested is plotted separately for the three electrode pairs tested, approximately in the base, middle, and apical regions of the arrays; the paired numbers within each panel indicate the left-right electrode numbers for each pair tested. The right-most panels for each subject show d' values as a function of the left-right ILDs tested for the three electrode pairs.

adult-onset deafness in which subjects demonstrated good sensitivity to ILD but variable sensitivity to ITD (van Hoesel and Tyler, 2003). While it is tempting to conclude that the mechanisms involved in processing ITD cues could be more susceptible to hearing loss than are the mechanisms associated with ILDs, these findings must be interpreted with caution, and some caveats are worth noting. First, while the total N size of subjects in this study is relatively large ($N=11$) compared with prior studies (N size 2–5), the number of subjects in each onset-of-deafness sub-group is too small to draw overarching conclusions. The trends observed here advocate for further research on this issue. Second, whereas the ITD is well defined in both acoustic and electrical stimulations, the electrical ILD is arbitrarily related to acoustic decibels in a processor through a loudness mapping function, and in the present experiment, we do not know what the ILDs used would correspond to in acoustic hearing. Thus, while performance across groups was not different, the ability of patients in our sub-groups to use ILDs in real-world situations may differ. Third, it is possible that in the ILD task subjects had access to some overall monaural level cues. That is, they might have been able to extract information regarding the levels in each ear separately and use those cues when making their judgment regarding intracranial position. Informal observations under monaural conditions suggest otherwise, but proper testing under these conditions is required prior to determining whether monaural cues were used, and if so, by which subjects under what conditions.

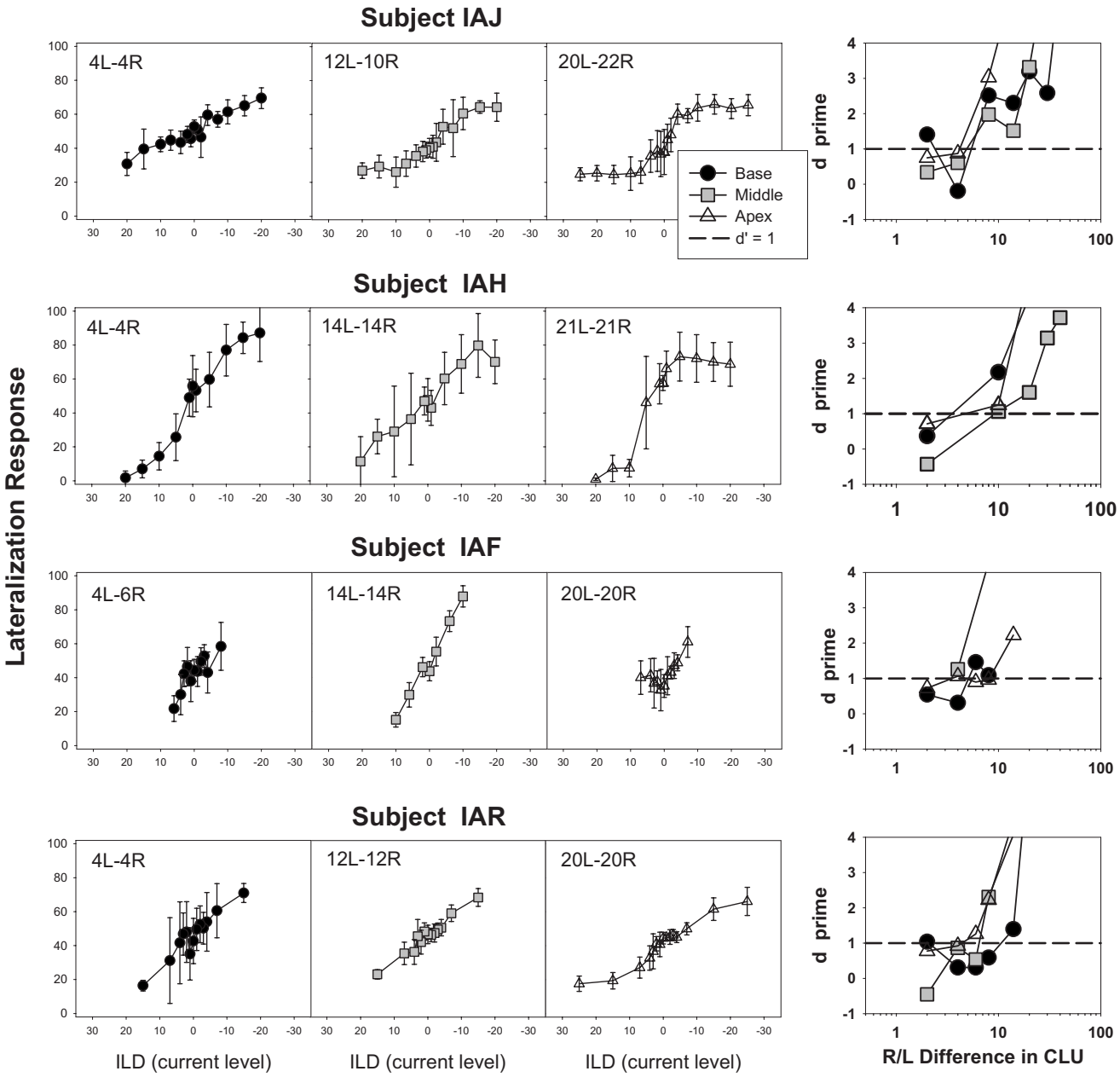


FIG. 9. Same as Fig. 8 for subjects with mid-childhood-onset of deafness.

Nonetheless, we specifically avoided another manipulation in which overall level is roved, as is done in many studies in free field, because it could potentially introduce uncontrolled ILD shifts due to unequal changes in loudness in the two ears, even when the same CLU increments/decrements are applied.

Results from this study are consistent with much of the prior work in the area of cochlear implantation in which individual variability is a clear hallmark. Here, the notable factor that might have contributed to variability is the age at onset of deafness. It is worth considering whether plasticity following altered sensory experience is particular to the binaural system or represents a more general predisposition of auditory system functionality following deprivation and subsequent activation. There is ample evidence from research with cochlear implant users to suggest that in other areas, including speech and language abilities, adults whose deafness occurred during adult life experience better outcomes than adults whose onset of deafness was early in life (Skinner *et al.*, 1994; Freisen *et al.*, 2001; Busby *et al.*, 1993). This evidence, supporting the general predisposition of neural systems to function best with early exposure, has long-standing roots in other sensory systems (Blakemore and Cooper, 1970; Rakic and Goldman-Rakic, 1982; Kaas *et al.*, 1983), and there is little reason to suspect that the findings with regard to effect of age at onset of deafness on performance are unique to audition.

Several of the subjects experienced deprivation of a different nature. Thus, our results also speak to the long-term potency of functional connections in the binaural auditory pathway, which are particularly notable in the individuals who became deaf as young adults, spent numerous years, even decades, being deprived of hearing, and subsequently had their auditory pathways reactivated with electric hearing.

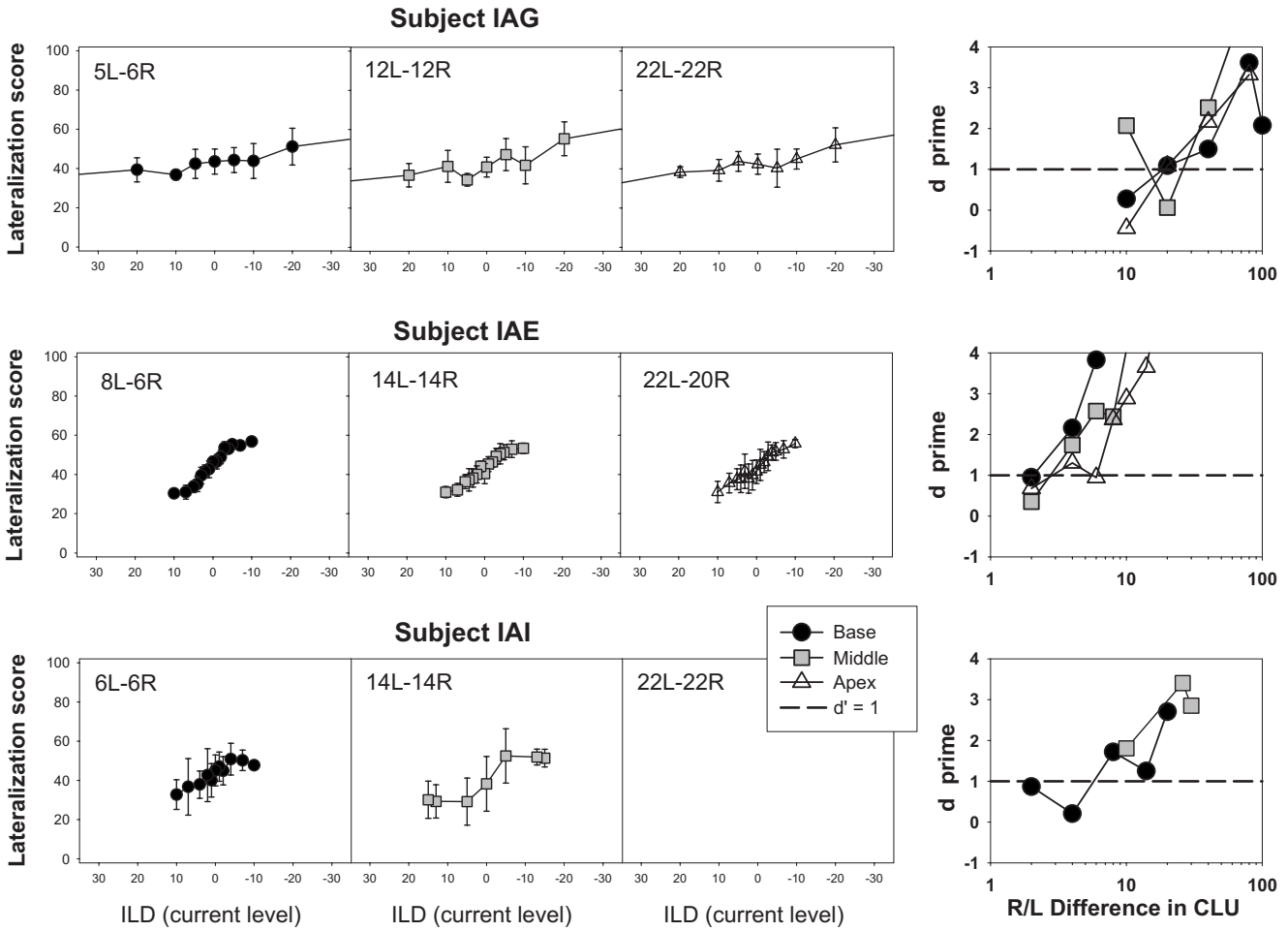


FIG. 10. Same as Fig. 8 for subjects with prelingual-onset of deafness.

Subject IAK in our study is a good example of someone who was profoundly deaf for 30 years, and within 1 year of having received bilateral cochlear implants shows binaural lateralization thresholds of ~450 μs for ITDs and 1.2 CLUs for ILDs. Subject IAD is another example of someone who lost hearing in the right ear at age 18 and in the left ear at age 46. Thus, he was deprived of binaural hearing for 28 years. Upon activation with bilateral CIs, he had an ITD threshold of 180 μs in the basal electrode pair and ILD thresholds of 1.4 CLUs or less at all stimulation places. Relative to binaural sensitivity of normal-hearing people, where ITD thresholds can be as low as 10–20 μs, thresholds seen here are more than an order of magnitude worse. Nonetheless, the ITD values at threshold are still potentially usable in real life. For instance, if ILD cues are absent then segregation of sources that are widely separated across the right and left hemifields or source location identification could be achieved with such ITDs.

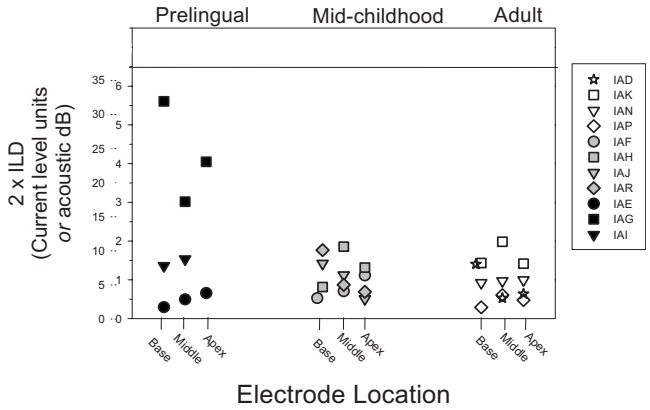


FIG. 11. Summary of results from ILD sensitivity measures are shown, grouped by age at onset of deafness from right to left: adult, mid-childhood, and prelingual. Each data point represents a single subject's ILD threshold defined as the difference in CLUs between the right and left ILDs at which performance reached $d' = 1$; in other words, the ILD value toward either right or left multiplied two times. Within each group, results are shown for the three places of stimulation along the cochleae tested (base, middle, and apex). There are no data points above the dashed horizontal line at the top because there were not data points that could not be measured or that could not be determined (CND), as was the case in the ITD results.

There may be some insights to be gained from the auditory deprivation literature in which animals underwent periods of monaural occlusion during various stages in development. It appears that neural circuits involved in binaural hearing can be recalibrated throughout life in reaction to altered inputs (Kacelnik *et al.*, 2006). What is not clear from that literature is the extent to which the remapping involves alterations in sensitivity to specific binaural cues. Hence, the extent to which the circuits that mediate ITD and ILD, respectively, are affected by experience remains to be studied with greater precision. When plasticity and recalibration of sensitivity to auditory cues are considered, in particular, at the level of the auditory cortex, there appears to be a protracted period of plasticity in the adult animal, which might

help to explain the retention or re-establishment of binaural sensitivity in adults tested in the present study. A factor that is most clearly potent in driving plasticity of neural circuits involved in spatial hearing is the shaping that takes place by training and experience (Kacelnik *et al.*, 2006; Keuroghlian and Knudsen, 2007). These findings from the animal literature can be viewed as potentially encouraging with regard to the role of training and rehabilitation of cochlear implant and hearing aid users and the possibility that, with experience, their spatial maps may be altered in ways that will lead to functional improvement in performance. Also noteworthy is the potential role of cues provided in clinical sound processors. If ITDs are not well coded by sound processors, but ILDs are, then listeners are being trained in their everyday situations with ILDs, but not ITDs.

Finally, recent studies in children who are born deaf and receive bilateral CIs as their first mode of auditory stimulation suggest the important role of early activation in the emergence of binaural interaction components at the level of the brainstem (Gordon *et al.*, 2007) and cortex (Bauer *et al.*, 2006). Behavioral results on this topic come from a population of children who are born deaf and who receive bilateral CIs either at a very young age or at older ages. The former group is more likely to reach age-appropriate spatial hearing resolution (Grieco-Calub *et al.*, 2008) than children who are stimulated bilaterally at a later age (Litovsky *et al.*, 2006b, 2006c). An interesting comparison can be made in future studies between the prelingually deaf children and adults in the prelingual group of the present study. While they share the common trait of prelingual deafness, early stimulation of these children may lead to better performance when they reach adulthood. What remains to be seen is whether these children will have better access to binaural cues, in particular, ITDs, given that the type of stimulation they are receiving provides no obligatory coordination of inputs to the two ears.

B. Effect of place of stimulation along the cochlear array

Data from the investigation of the effect of place of stimulation reported here suggest that there can be good sensitivity to timing differences with stimulation in the basal region of the cochlea, where innervating auditory nerve fibers are tuned to high-frequencies. This finding lends support to the view that ITDs can be processed by neurons with best sensitivity to high-frequency signals (Joris and Yin, 1995; Joris *et al.*, 2004). ITDs may be categorized as differences in timing of the onset, fine-structure (carrier), and envelope. Generally, in normal-hearing listeners, ITD-based lateralization is dominated by onset and fine-structure cues (Neutzel and Hafter, 1976; Bernstein and Trahiotis, 1985). With high-frequency carriers, where basal stimulation predominates, the ITDs in envelopes can also provide reliable binaural cues for lateralization (Bernstein, 2001; Joris *et al.*, 2004). Our data provide further support for the ability of the auditory system to use ITDs under conditions in which high-frequency auditory nerve fibers are stimulated at the basal region of the cochlea.

C. Remaining gaps and issues in bilateral CI users

Results from 11 subjects whose data are presented here show a broad range of performance and reflect the type of variability that is a hallmark of performance with CIs. Approaches for improving performance can include a number of avenues. Here, direct tests of sensitivity to binaural cues suggest that ILDs are available to all subjects, and that right-left discrimination thresholds can be as low as the smallest ILD (~0.1–0.2 dB) allowed by clinical implant systems (Lawson *et al.*, 1998; Long *et al.*, 2003; van Hoesel and Tyler, 2003). In contrast, ITD thresholds tested here and by others are highly variable, ranging from ~50 μs (rare) to several milliseconds depending on the subject, pulse rate, and electrode pair tested (Lawson *et al.*, 1998; van Hoesel and Tyler, 2003; van Hoesel, 2007). This reflects a clear deviation from performance in normal-hearing listeners, in whom sensitivity to ITD for acoustic stimuli can be as good as thresholds of 10–20 μs, which is clearly better, is also more consistent across subjects, and less dependent on stimulus parameters (Klumpp and Eady, 1956; Yost *et al.*, 1971). In addition, ITD sensitivity in bilateral CI users becomes much worse as the rate of stimulation increases from 100 to 1000 pps (e.g., van Hoesel *et al.*, 2009), which contrasts with pure-tone sensitivity in normal-hearing listeners, which actually improves over the same range of rate increase (Bernstein, 2001; Bernstein and Trahiotis, 2002). Thus, it may be the case that ITD coding is where there may be room for improvement in current bilateral implants. There are caveats, however, in the design of bilateral implants that must be overcome prior to this improvement being possible, which include preservation of binaural information by the clinical sound processors. One area that should be investigated is the possibility that stimulation across the two ears can be more selective so as to ensure that both auditory nerves receive information that is frequency-matched in addition to being well-timed.

ACKNOWLEDGMENTS

The authors are grateful to the bilateral cochlear implant users for their participation. This work was supported by the NIH-NIDCD (Contract No. R01 DC 003083 to R.Y.L.). Support for R.v.H. was provided by The Hearing CRC, Australia. Portions of this work were presented at Association for Research in Otolaryngology Mid-winter Meeting, 2005.

[1]The current produced by the device is $I=K\times 10\times 175^{(CL/255)}$, where I is the current in microamperes and K is a calibration factor that can be disregarded in this description. Taking $20\log 10(x)$ of each side, the product is $20\log 10(I)=20\log 10(10K)+20(CL/255)\log 10(175)$. Thus, for a change of 1 CLU, $\Delta[20\log 10(I)]=20\log 10(175)/255=0.176$ dB in terms of current. Note that the definition of clinical level units is independent of acoustic SPL. The number of CLUs corresponding to a given SPL change is determined by a separate mapping function in the processor (the loudness growth map). The same approach has been used in previous studies (e.g., van Hoesel, 2004, 2008; van Hoesel *et al.*, 1993, 2009).

[2]Extensive discussions with subjects convinced the authors of this paper that the ITD cues were indeed being mapped to specific perceived source locations, rather than some other arbitrary percept.

[3]The term "could not test" refers to extensive attempts during testing to find combinations of stimuli that produced binaural fused auditory images that the listeners would also report to be lateralized with large ITD values.

Bauer, P. W., Sharma, A., Martin, K., and Dorman, M. (**2006**). "Central auditory development in children with bilateral cochlear implants," Arch. Otolaryngol. Head Neck Surg. **132**, 1133–1136.
Bernstein, L. R. (**2001**). "Auditory processing of interaural timing information: New insights," J. Neurosci. Res. **66**, 1035–1046.
Bernstein, L. R., and Trahiotis, C. (**1985**). "Lateralization of low-frequency, complex waveforms: The use of envelope-based temporal disparities," J. Acoust. Soc. Am. **77**, 1868–1880.
Bernstein, L. R., and Trahiotis, C. (**2002**). "Enhancing sensitivity to interaural delays at high frequencies by using 'transposed stimuli'," J. Acoust. Soc. Am. **112**, 1026–1036.
Blakemore, C., and Cooper, G. F. (**1970**). "Development of the brain depends on the visual environment," Nature (London) **228**, 477–478.
Blauert, J. (**1997**). *Spatial Hearing* (The MIT Press, Cambridge, MA).
Busby, P. A., Tong, Y. C., and Clark, G. M. (**1993**). "Electrode position, repetition rate, and speech perception by early- and late-deafened cochlear implant patients," J. Acoust. Soc. Am. **93**, 1058–1067.
Coco, A., Epp, S. B., Fallon, J. B., Xu, J., Millard, R. E., and Shepherd, R. K. (**2007**). "Does cochlear implantation and electrical stimulation affect residual hair cells and spiral ganglion neurons?," Hear. Res. **225**, 60–70.
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.
Gordon, K. A., Valero, J., and Papsin, B. C. (**2007**). "Auditory brainstem activity in children with 9–30 months of bilateral cochlear implant use," Hear. Res. **233**, 97–107.
Grieco-Calub, T. M., Litovsky, R. Y., and Werner, L. A. (**2008**). "Using the observer-based psychophysical procedure to assess localization acuity in toddlers who use bilateral cochlear implants," Otol. Neurotol. **29**, 235–239.
Joris, P. X., and Yin, T. C. (**1995**). "Envelope coding in the lateral superior olive. I. Sensitivity to interaural time differences," J. Neurophysiol. **73**, 1043–1062.
Joris, P. X., Schreiner, C. E., and Rees, A. (**2004**). "Neural processing of amplitude-modulated sounds," Physiol. Rev. **84**, 541–577.
Kaas, J. H., Merzenich, M. M., and Killackey, H. P. (**1983**). "The reorganization of somatosensory cortex following peripheral nerve damage in adult and developing mammals," Annu. Rev. Neurosci. **6**, 325–356.
Kacelnik, O., Nodal, F. R., Parsons, C. H., and King, A. J. (**2006**). "Training-induced plasticity of auditory localization in adult mammals," PLoS Biol. **4**, e71.
Keuroghlian, A. S., and Knudsen, E. I. (**2007**). "Adaptive auditory plasticity in developing and adult animals," Prog. Neurobiol. **82**, 109–1021.
Klumpp, R. G., and Eady, H. R. (**1956**). "Some measurements of interaural time difference thresholds," J. Acoust. Soc. Am. **28**, 859–860.
Laback, B., Majdak, P., and Baumgartner, W. D. (**2007**). "Lateralization discrimination of interaural time delays in four-pulse sequences in electric and acoustic hearing," J. Acoust. Soc. Am. **121**, 2182–2192.
Lawson, D. T., Wilson, B. S., Zerbi, M., van den Honert, C., Finley, C. C., Farmer, J. C., Jr., McElveen, J. T., Jr., and Roush, P. A. (**1998**). "Bilateral cochlear implants controlled by a single speech processor," Am. J. Otol. **19**, 758–761.
Leake, P. A., Hradek, G. T., and Snyder, R. L. (**1999**). "Chronic electrical stimulation by a cochlear implant promotes survival of spiral ganglion neurons after neonatal deafness," J. Comp. Neurol. **412**, 543–562.
Litovsky, R. Y., Parkinson, A., and Arcaroli, J. (**2009**). "Spatial hearing and speech intelligibility in bilateral cochlear implant users," Ear Hear. **30**, 419–431.
Litovsky, R. Y., Parkinson, A., Arcaroli, J., Peters, R., Lake, J., Johnstone, P., and Yu, G. (**2004**). "Bilateral cochlear implants in adults and children," Arch. Otolaryngol. Head Neck Surg. **130**, 648–655.
Litovsky, R. Y., Parkinson, A., Arcaroli, J., and Sammeth, C. (**2006a**). "Simultaneous bilateral cochlear implantation in adults: A multicenter clinical study," Ear Hear. **27**, 714–731.
Litovsky, R. Y., Johnstone, P. M., and Godar, S. P. (**2006b**). "Benefits of bilateral cochlear implants and/or hearing aids in children," Int. J. Audiol. **45**, S78–S91.
Litovsky, R. Y., Johnstone, P. M., Godar, S., Agrawal, S., Parkinson, A., Peters, R., and Lake, J. (**2006c**). "Bilateral cochlear implants in children: Localization acuity measured with minimum audible angle," Ear Hear. **27**, 43–59.
Long, C. J., Carlyon, R. P., Litovsky, R. Y., and Downs, D. H. (**2006**). "Binaural unmasking with bilateral cochlear implants," J. Assoc. Res. Otolaryngol. **7**, 352–360.
Long, C. J., Eddington, D. K., Colburn, H. S., and Rabinowitz, W. M. (**2003**). "Binaural sensitivity as a function of interaural electrode position with a bilateral cochlear implant user," J. Acoust. Soc. Am. **114**, 1565–1574.
Majdak, P., Laback, B., and Baumgartner, W. D. (**2006**). "Effects of interaural time differences in fine structure and envelope on lateral discrimination in electric hearing," J. Acoust. Soc. Am. **120**, 2190–2201.
McDermott, H., McKay, C., and Vandali, A. (**1992**). "A new portable sound processor for the University of Melbourne/Nucleus Limited multi electrode cochlear implant," J. Acoust. Soc. Am. **91**, 3367–3371.
Nuetzel, J. M., and Hafter, E. R. (**1976**). "Lateralization of complex waveforms: Effects of fine structure, amplitude, and duration," J. Acoust. Soc. Am. **60**, 1339–1346.
Nopp, P., Schleich, P., and D'Haese, P. (**2004**). "Sound localization in bilateral users of MED-EL COMBI 40/40+ cochlear implants," Ear Hear. **25**, 205–214.
Oxenham, A. J., Bernstein, J. G., and Penagos, H. (**2004**). "Correct tonotopic representation is necessary for complex pitch perception," Proc. Natl. Acad. Sci. U.S.A. **101**, 1421–1425.
Rakic, P., and Goldman-Rakic, P. S. (**1982**). "The development and modifiability of the cerebral cortex," Neurosci. Res. Program Bull. **20**, 433–438.
Schleich, P., Nopp, P., and D'Haese, P. (**2004**). "Head shadow, squelch, and summation effects in bilateral users of the MED-EL COMBI 40/40+ cochlear implant," Ear Hear. **25**, 197–204.
Seligman, P. M., Patrick, J. F., Tong, Y. C., Clark, G. M., Dowell, R. C., and Crosby, P. A. (**1984**). "A signal processor for a multiple-electrode hearing prosthesis," Acta Oto-Laryngol., Suppl. **411**, 135–139.
Skinner, M. W., Clark, G. M., Whitford, L. A., Seligman, P. M., Staller, S. J., Shipp, D. B., Shallop, J. K., Everingham, C., Menapace, C. M., Arndt, P. L. (**1994**). "Evaluation of a new spectral peak coding strategy for the Nucleus 22 channel cochlear implant system," Am. J. Otol. **15**, 15–27.
van Hoesel, R. J. (**2004**). "Exploring the benefits of bilateral cochlear implants," Audiol. Neuro-Otol. **9**, 234–246.
van Hoesel, R. J. (**2007**). "Sensitivity to binaural timing in bilateral cochlear implant users," J. Acoust. Soc. Am. **121**, 2192–2206.
van Hoesel, R. J. M. (**2008**). "Observer weighting of level and timing cues in bilateral cochlear implant users," J. Acoust. Soc. Am. **124**, 3861–3872.
van Hoesel, R. J., and Tyler, R. S. (**2003**). "Speech perception, localization, and lateralization with bilateral cochlear implants," J. Acoust. Soc. Am. **113**, 1617–1630.
van Hoesel, R. J. M., Jones, G. L., and Litovky, R. Y. (**2009**). "Interaural time-delay sensitivity in bilateral cochlear implant users: Effects of pulse rate, modulation rate, and place of stimulation," J. Assoc. Res. Otolaryngol. **10**, 557–567.
van Hoesel, R. J., Tong, Y. C., Hollow, R. D., and Clark, G. M. (**1993**). "Psychophysical and speech perception studies: A case report on a binaural cochlear implant subject," J. Acoust. Soc. Am. **94**, 3178–3189.
Wilson, B. S., and Dorman, M. F. (**2008a**). "Cochlear implants: A remarkable past and a brilliant future," Hear. Res. **242**, 3–21.
Wilson, B. S., and Dorman, M. F. (**2008b**). "Cochlear implants: Current designs and future possibilities," J. Rehabil. Res. Dev. **45**, 695–730.
Wilson, B. S., Finley, C. C., Lawson, D. T., Wolford, R. D., Eddington, D. K., and Rabinowitz, W. M. (**1991**). "Better speech recognition with cochlear implants," Nature (London) **352**, 236–238.
Wright, B. A., and Zhang, Y. (**2006**). "A review of learning with normal and altered sound-localization cues in human adults," Int. J. Audiol. **45**, S92–S98.
Yost, W. A., Wightman, F. L., and Green, D. M. (**1971**). "Lateralization of filtered clicks," J. Acoust. Soc. Am. **50**, 1526–1531.

A relation between electrode discrimination and amplitude modulation detection by cochlear implant listeners

Monita Chatterjee[a] and Jian Yu
Cochlear Implants and Psychophysics Laboratory, Department of Hearing and Speech Sciences, University of Maryland, College Park, Maryland 20742

(Received 22 March 2009; revised 6 October 2009; accepted 7 October 2009)

The objective of this study was to examine the relation between measures of spectral and temporal resolutions in cochlear implant listeners at a particular electrode location. The hypothesis was that a common underlying factor, such as the health of local groups of neurons, might partially determine patients' sensitivity to both spectral and temporal cues at specific tonotopic locations. Participants were adult cochlear implant listeners. A significant correlation was found between electrode discrimination measured at soft levels (20% and 30% of the dynamic range) and modulation sensitivity at those levels, for stimulation in bipolar mode and a 100 Hz modulation rate. Correlations between the two measures were weaker under monopolar stimulation, or when the modulation rate was 10 Hz. At a higher stimulation level (40% of the dynamic range), no significant correlations between these measures were observed. It is hypothesized that the more restricted excitation pattern at lower levels and/or with a narrower stimulation mode allows the measurement of locally driven sensitivity to spectral and temporal cues, particularly under more challenging listening conditions. Thus, psychophysical measures obtained under conditions that evoke a narrower excitation pattern may serve as a useful indicator of the functional health of local neural populations. © *2010 Acoustical Society of America.* [DOI: 10.1121/1.3257591]

PACS number(s): 43.66.Ts [JCM] Pages: 415–426

I. INTRODUCTION

Speech perception with a cochlear implant (CI) is thought to require good resolution in both spectral and temporal domains. As far as the relation between speech perception and spectral resolution is concerned, the picture is not entirely clear. While earlier studies exploring the relation between place-pitch coding and speech perception by CI listeners have reported mixed results (Nelson *et al.*, 1995; Collins *et al.*, 1997; Throckmorton and Collins, 1999; Zwolan *et al.*, 1997; Donaldson and Nelson, 2000; Henry *et al.*, 2000), more recent investigations employing more complex stimuli have demonstrated significant correlations between spectral pattern processing abilities and speech perception by CI listeners (Henry and Turner, 2003; Henry *et al.*, 2005; Litvak *et al.*, 2007; Won *et al.*, 2007). It seems reasonable to assume that spectral pattern processing would require good electrode discrimination, but direct evidence for this is not clearly present in the literature.

On the other hand, temporal resolution has been found to have a positive relation with speech perception by CI listeners. Cazals *et al.* (1994) found that vowel and consonant recognition scores were significantly correlated with the slope of the high-frequency cutoff of the modulation transfer function measured in CI listeners at the most apical electrode. Fu (2002) showed that the mean modulation detection threshold (of 100 Hz modulated pulse trains) of CI listeners, calculated across the dynamic range, was strongly correlated with their phoneme recognition scores. As modulation sensitivity in CI listeners tends to be strongly level-dependent, the mean of modulation thresholds measured at various levels provides a more representative measure of listeners' performance than a single measurement at a fixed level. The correlations observed by Fu (2002) and Cazals *et al.* (1994) between the measures of modulation sensitivity and vowel recognition are somewhat difficult to explain, given that fast temporal cues have been shown to play a relatively minor role in vowel recognition (Xu *et al.*, 2005). As correlation does not necessarily imply causation, interpretation of the results should be undertaken with some caution. It is possible, for instance, that CI listeners who have excellent modulation sensitivity, also have excellent spectral (spatial) resolution: the overall health and number of surviving auditory nerve neurons might dictate both. Thus, the correlations observed between either spectral resolution or temporal sensitivity and speech perception measures might well be a reflection of the contribution of overall spectro-temporal sensitivity to speech perception, rather than the importance of temporal sensitivity or spectral resolution alone.

Work by Shepherd and colleagues indicates that as duration of deafness increases, auditory nerve neurons in the cat decrease in number; further, the surviving neurons show poorer phase-locking to the temporal patterns of electrical stimuli owing to longer refractory periods (Shepherd and Javel, 1997; Shepherd *et al.*, 2004). Smaller numbers or irregular patterns of surviving neurons may be expected to result in weak or uneven pitch salience, thus contributing to poorer electrode discrimination. At the same time, poorer phase-locking of the surviving neurons would result in poorer temporal discrimination. It is thus possible that the pattern of nerve survival and the health of surviving neurons

[a]Author to whom correspondence should be addressed. Electronic mail: mchatterjee@hesp.umd.edu

TABLE I. Relevant information about subjects. Note that "Early/perilingual" onset of deafness indicates cases in which the subjects was not congenitally deaf; but may have had some auditory exposure prior to the onset of deafness. "Stimulation Mode" refers to the experimental mode used in the current study.

Subject	Onset of deafness	Stimulation mode	Device	Gender	Age at implantation	Age at testing
S1	Early/perilingual	MP1+2, BP+1	N-24	F	46	49
S2	Prelingual	MP1+2, BP+2	N-24	M	16	23
S3	Postlingual	MP1+2, BP+2	N-24	M	57	60
S4	Postlingual	BP+2	N-22	M	35	49
S5	Postlingual	BP+2	N-22	M	49	65
S6	Postlingual	BP+2	N-22	M	51	65
S7	Early/perilingual	MP1+2, BP+2	N-24	F	47	56
S8	Postlingual	MP1+2, BP+2	N-24	F	60	64
S9	Postlingual	MP1, BP+1	Freedom	F	63	64
S10	Postlingual	MP1, BP+2	Freedom	F	66	67
S11	Postlingual	MP1, BP+1	Freedom	F	69	71
S12	Postlingual	MP1	Freedom	M	64	67
S13	Postlingual	MP1	Freedom	F	63	64

is related to electrode discrimination (sensitivity to changes in the spectral pattern) as well as modulation detection (sensitivity to changes in the temporal pattern). Thus, measures of modulation sensitivity may be correlated with measures of electrode discrimination.

The strength of such a relationship may also depend on the stimulation mode. With monopolar (MP) stimulation (the mode of choice in present-day devices), the electric field is thought to be broader than with bipolar (BP) stimulation. By stimulating smaller groups of neurons using the bipolar stimulation mode, it should be possible to evoke responses from more localized regions of the auditory nerve array than in monopolar mode. In this case, the strength of the relation, if any, between spatial discrimination and temporal sensitivity may be stronger in bipolar mode than in monopolar mode.

In the present study, we explored the relation between electrode discrimination and sensitivity to amplitude modulation at a medial electrode location in a group of cochlear implant listeners. Effects of stimulation mode, stimulus level, and the rate of amplitude modulation were also examined.

Both electrode discrimination and amplitude modulation sensitivity improve with stimulus level in CI users, reaching ceiling performance at moderate loudness levels (Pfingst *et al.*, 2007; Galvin and Fu, 2005; Fu, 2002; Pfingst *et al.*, 1999; McKay *et al.*, 1999). The excellent performance achieved by most CI listeners in these tasks at moderate stimulation levels (e.g., at 40% of the dynamic range and above) means that the relative range of variation in performance across subjects is reduced as stimulation level is increased. It is therefore more likely that if there is indeed a relation between listeners' performance in these two tasks, it might be observed at low stimulation levels rather than at high levels. In support of this expectation, preliminary results indicated a relation between amplitude modulation sensitivity and electrode discrimination at soft stimulation levels, but not at high stimulation levels. The goal of the study was to examine the relations between measures of electrode discrimination and amplitude modulation sensitivity under conditions that were likely to produce the greatest intersubject variability. Therefore, the focus of this investigation was placed on performance at relatively soft levels of stimulation (i.e., 20%–40% of the dynamic range).

Bipolar stimulation requires higher current levels than monopolar stimulation. To obtain a measurable dynamic range while remaining within the compliance limits of the device, it was necessary to (i) increase the pulse phase duration and inter-phase gap in bipolar mode over that in monopolar mode and (ii) in some instances, to increase the distance between the active and return electrodes in bipolar mode (to BP+1 or BP+2 as needed). These steps, while necessary, may have reduced the effective difference between the two modes of stimulation in this study. These constraints have likely weakened the strength of the observations reported here. This suggests that studies conducted under more ideal conditions might yield even stronger findings than the present study.

Sensitivity to amplitude modulation at both a slow (10 Hz) and a fast (100 Hz) rate was measured in this study. While slow envelope fluctuations are likely to be primarily limited by intensity discrimination limens, fast fluctuations are additionally limited by the temporal resolution of the system (phase-locking of auditory nerve neurons, etc.). As phase-locking has been shown to be compromised in animals with prolonged duration of deafness, it is expected that the 100 Hz modulation would be more poorly encoded than the 10 Hz modulation by the auditory system in locations of damage.

II. METHODS

A. Subjects

A total of 13 adult CI users (S1–S13) participated in this study. Of these, 11 were post-lingually deafened, and the remaining two (S2 and S7) were early deafened. Relevant information about the subjects is provided in Table I. All participants used CIs manufactured by Cochlear Corporation (N-22, N-24, or Freedom devices). Eight participants were

involved in measurements with both monopolar and bipolar modes. An additional three were stimulated in bipolar mode only. Two others were stimulated in monopolar mode only.

B. Stimuli

The nucleus cochlear implant arrays described in this study comprise 22 intracochlear electrodes, numbered 1–22 from base to apex. Henceforth, stimulation in BP+1 or BP+2 modes will be collectively referred to as BP or bipolar stimulation, while stimulation in MP1 or MP1+2 modes will be collectively referred to as MP or monopolar stimulation. A custom research interface (House Ear Institute Nucleus Research Interface: Shannon *et al.*, 1990; Robert, 2002) was used to generate electrical stimuli. All stimuli were 300-ms long trains of biphasic current pulses, presented at 1000 pulses/s. For modulation detection tasks, stimuli were presented to electrode 10 (a centrally located electrode in the array) in either bipolar or monopolar stimulation mode. For electrode discrimination tasks, the reference stimulus was presented on electrode 10 and the comparison stimuli were presented on neighboring electrodes. In bipolar mode, pulses were 100–200 μs in phase duration, with a 20 μs interphase gap. In monopolar mode, pulses were either 25 or 40 μs in phase duration, with interphase gaps of 6 or 8 μs. In the electrode discrimination experiment, stimuli were unmodulated pulse trains. In the modulation detection experiment, a sinusoidal modulation was applied to the amplitude of the pulses. Stimuli were presented in BP+1 or BP+2 mode for the BP experiments, and in MP1 or MP1+2 mode in the MP experiments. In the BP experiments, the narrowest mode that allowed us to measure the full dynamic range was selected. In the MP experiments, the N-24 users were tested in MP1+2 mode. For technical reasons, the Freedom users were tested in MP1 mode.

C. Procedures

1. Threshold and dynamic range

Threshold was obtained using a two-down, one-up, two-interval, forced-choice, adaptive procedure, converging at the 70.7% correct point on the psychometric function (Levitt, 1971). The mean of two runs was calculated as the threshold. The maximum acceptable level (MAL) was measured using a procedure in which the subject adjusted the level of the stimulus by pressing the "up arrow" or "down arrow" keys in the computer keyboard until the level of the sound was at the upper end of comfortable loudness. The dynamic range (DR) was calculated as the difference in microamperes between the MAL and the threshold. Stimuli were presented at various % DRs.

2. Modulation detection thresholds

The threshold modulation depth (MDT) was measured using a two-down, one-up, three-interval, forced-choice (3IFC) adaptive procedure, converging at the 70.7% correct point on the psychometric function (Levitt, 1971). Two of the three intervals (randomly) contained the unmodulated reference stimulus, while the remaining interval contained the modulated one; the subject's task was to indicate which interval sounded different. The MDT was measured at levels of the carrier ranging from 20% to 80% of the DR, in 10% DR increments. Measurements were made at modulation frequencies of 100 and 10 Hz. In each condition (e.g., bipolar mode and 100 Hz modulation frequency), measurements at different carrier levels were made in random order. At least two independent measurements were made at each level, and the mean of the two was calculated as the modulation threshold at that carrier level.

3. Electrode discrimination

Electrode discrimination was measured using a three-interval, forced-choice procedure, in which two of the intervals (randomly) contained the stimulus on the reference electrode, and the remaining interval contained the comparison electrode. The subject's task was to indicate which of the three intervals sounded different from the other two. Each comparison electrode was presented 20 times. Blocks were created with a number of comparison electrodes, and the order of presentations was randomized. Results were obtained as percent correct. The percent correct scores were converted into d' measures for further analyses (Macmillan and Creelman, 2005). For a particular trial, stimuli were presented at the same % DR level.

It is to be noted that stimulus level was not roved in the electrode discrimination task. In preliminary experiments, we observed large variations in performance when small level variations were introduced, perhaps because performance asymptotes rapidly with level in this task. However, as several electrode comparisons were included in each block and the presentations were randomized, stimuli presented in successive trials would have had slightly different loudness levels (some slightly louder and some slightly softer), which would have resulted in a natural rove. Subjects were asked about loudness cues, and all except two (subjects S7 and S8) reported that the stimuli were equally loud. In their case, loudness was reported by the subject to be an important cue in the task, and the stimuli were loudness-balanced (adjusted until the subjects reported them to be equally loud).

Initially, electrode discrimination was measured at a range of levels in some of the subjects (S2, S5, S6, and S7). Preliminary analyses indicated a correlation between electrode discrimination at soft levels and measures of modulation sensitivity in the bipolar stimulation mode. Therefore, we elected to constrain the remaining measures to the lower levels (20%–40% DR). At higher levels of stimulation, performance in electrode discrimination approached ceiling levels, and inter-subject variation was relatively reduced.

III. RESULTS

A. Level dependence of the MDT

Consistent with previous findings (Fu, 2002; Shannon, 1992; Chatterjee and Robert, 2001; Chatterjee and Oba, 2005; Pfingst *et al.*, 2007), modulation sensitivity at both 100 and 10 Hz showed considerable level dependence in all subjects. The top left-hand panel of Fig. 1 shows the modulation detection thresholds obtained in all 11 subjects in the medial location (electrode 10) using the bipolar stimulation mode

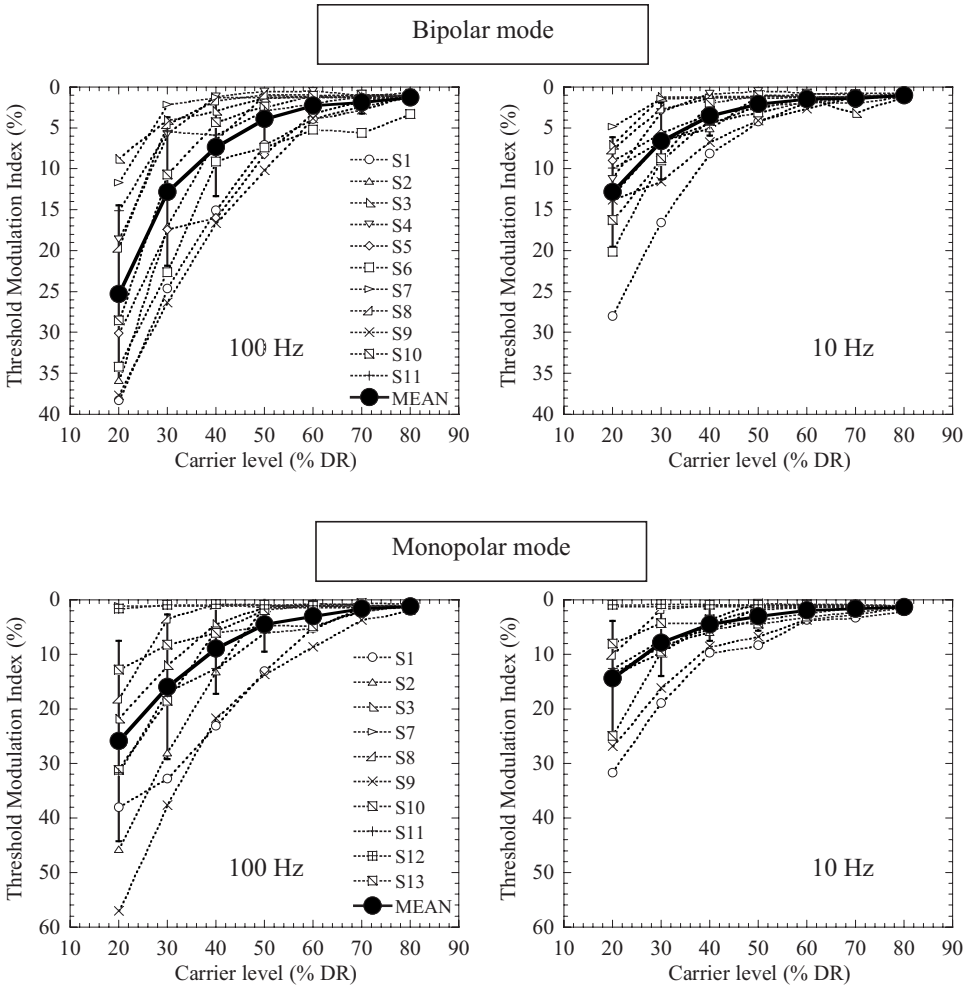


FIG. 1. Threshold modulation index (in %) plotted as a function of stimulus level. Upper and lower panels show results obtained in bipolar and monopolar modes, respectively. Left- and right-hand panels show results obtained using the 100 Hz and the 10 Hz modulation rates, respectively. Within each panel, different symbols refer to different subjects. The thick solid line indicates the mean calculated across the seven subjects, in each case. Error bars show ±1 s.d. Note that the subject pools for bipolar and monopolar modes are overlapping, but not identical.

and the 100 Hz modulation rate, plotted as a function of level. The top left-hand panel shows data obtained using identical conditions, but with a 10 Hz modulation rate. Results in monopolar mode, but with a somewhat different group of subjects, are shown in the lower panels. Note that the vertical scale is different between the upper and lower panels. Although there was considerable inter-subject variation in the absolute modulation detection thresholds, the shape of the function relating modulation sensitivity to carrier level was similar across subjects and conditions, and is captured by the mean function (thick solid line in each panel). Thresholds decreased (sensitivity increased) with increasing carrier level following a nonlinear curve, asymptoting at generally excellent performance at moderate to high carrier levels. The carrier level at which performance reached the asymptote varied across subjects. As expected from the low-pass-filter shape of the temporal modulation transfer function (Shannon, 1992), the mean sensitivity of subjects to the 10 Hz modulator was greater than the mean sensitivity to the 100 Hz modulator in each stimulation mode. A two-way repeated measures analysis of variance (ANOVA) was conducted on the results obtained in each mode. In BP mode (using a Greenhouse–Geisser adjustment for violation of the sphericity assumption), the results showed a significant main effect of level [$F(1.243, 12.429) = 40.067$, $p < 0.001$] and modulation frequency [$F(1, 10) = 18.49$, $p < 0.005$], and a significant interaction between the two [$F(1.748, 17.476) = 22.136$, $p < 0.001$]. An identical analysis conducted on the results obtained in MP mode (also using the Greenhouse–Geisser adjustment) also showed significant main effects of level [$F(1.147, 10.324) = 16.897$, $p < 0.001$] and modulation frequency [$F(1, 9) = 10.074$, $p < 0.02$] and a significant interaction between the two was again observed [$F(1.426, 12.831) = 8.866$, $p < 0.001$]. The interaction between carrier level and modulation frequency stems from the greater sensitivity to the slower modulation rates at lower carrier levels. At high carrier levels, the two mean functions converge. The temporal modulation transfer function (TMTF) of cochlear implant listeners reflects these differences as well: at low carrier levels, the TMTF often appears to be more low-pass (larger differences in sensitivity to low and high modulation rates), while at high carrier levels, the function tends to flatten out (smaller differences between sensitivity to low and high modulation rates: Shannon, 1992; Chatterjee and Robert, 2001).

Although the above results obtained with monopolar and bipolar stimulations were not obtained with the same set of subjects, eight of the participants provided data with both stimulation modes. The effect of stimulation mode, and its interactions with modulation frequency and level, could thus be directly tested on the results obtained with these subjects. Mean results are shown in Fig. 2, with open and closed symbols representing bipolar and monopolar stimulations, re-

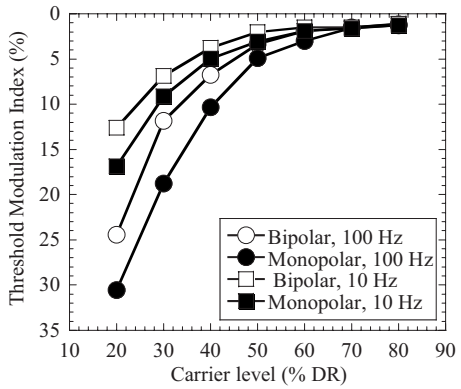


FIG. 2. Mean modulation detection thresholds plotted as a function of stimulus level, for bipolar and monopolar stimulation modes (open and filled symbols, respectively), and 100 and 10 Hz modulation rates (circles and squares respectively). Results are shown for the eight subjects who were tested in both stimulation modes.

spectively, while circles and squares represent 100 and 10 Hz data. Error bars are omitted for clarity. A three-way repeated measures ANOVA was conducted on these results. Despite considerable variation in the results, analyses showed a small but statistically significant main effect of stimulation mode [$F(1,7)=8.61$, $p=0.049$], a significant main effect of modulation frequency [$F(1,7)=11.655$, $p<0.02$], and a significant main effect of level [$F(1.15,8.03)=23.796$, $p<0.005$]. The Greenhouse–Geissler correction was applied for violation of sphericity. Significant interactions were observed between stimulation mode and level [$F(1.32,9.26)=4.73$, $p<0.05$] and between modulation frequency and level [$F(1.423,9.963)=12.27$, $p<0.005$]. No interactions were observed between stimulation mode and modulation frequency, and no three-way interactions were observed. As is seen in Fig. 2, mean thresholds in bipolar mode were consistently higher than mean thresholds in monopolar mode: however, these differences were not large. There was also considerable inter-subject variation in the data, likely contributing to the small level of significance for the effect of mode.

B. Level dependence of electrode discrimination

Figure 3 shows results obtained in the electrode discrimination task by the 11 subjects using bipolar stimulation. The abscissa shows the location of the comparison electrode, and the ordinate shows percent correct in the 3IFC task. The horizontal solid line shows chance performance. Figure 4 shows similar results obtained with the ten subjects using the monopolar stimulation mode. Performance improved moderately with level in all subjects and stimulation modes, with one exception: subject S7's performance declined with increasing level in the monopolar stimulation mode. The dependence of electrode discrimination on level is consistent with previous observations by Pfingst *et al.* (1999) and McKay *et al.* (1999). 8 of the 11 subjects who provided data with the bipolar stimulation mode also participated in the measurements using the monopolar stimulation mode. Figure 5 provides a side-by-side comparison of the results obtained in these eight subjects with the two stimulation modes. The ordinate represents the average of the values of d' obtained on electrodes located at equal distances from the reference electrode on the basal and apical side. The abscissa represents the relative distance of the comparison electrodes from the reference electrode. Within each panel, the parameter is the stimulus level (in % DR). The left-hand column shows results obtained with bipolar stimulation, and the right-hand column shows results obtained with monopolar stimulation.

For purposes of analysis, sensitivity indices were developed for electrode discrimination. As most subjects reached excellent performance at a three-electrode distance from the reference electrode, the one- and two-electrode distances were selected as ones that would yield poor to excellent performance without reaching ceiling effects in most subjects. Performance (in percent correct) at a distance of two electrodes (both in the apical end and in the basal direction) from the reference electrode was converted into a d' score for each electrode location and stimulation mode tested. The mean of the two values of d' (basal and apical) was taken as the measure of electrode discrimination. This measure will be henceforth referred to as the EDI_2 (electrode discrimination index at the two-electrode distance). A similar measure of sensitivity at the one-electrode distance was also developed (EDI_1). Figure 6 shows the mean d' obtained at distances of one and two electrodes away from the reference electrode in the bipolar and monopolar stimulation modes, at 20%, 30%, and 40% DR. The measures of EDI_1 were somewhat larger with bipolar than with monopolar stimulation. A paired t-test showed a statistically significant difference between the one-electrode d' obtained at 20% and 40% DR ($p<0.05$) but not at 30% DR. No significant difference was observed between the two-electrode d' obtained in the two modes at any of the levels.

C. Measures of dispersion for modulation detection and electrode discrimination

Inspection of the results suggested that at higher amplitudes of stimulation, the inter-subject variation in electrode discrimination and modulation detection would be reduced as a result of performance approaching ceiling levels. The coefficient of variation was selected as an appropriate, normalized measure of the dispersion in the data in each case. Analysis showed that indeed, the coefficient of variation in both the EDI_1 and EDI_2 measures reduced systematically with increasing level in both stimulation modes [Fig. 7(a)]. The coefficients of variations computed for modulation sensitivity ($1-m$, where m=modulation index at threshold) were much smaller than those calculated for electrode discrimination [Fig. 7(b)]. However, these coefficients also declined with increasing level in both stimulation modes [inset of Fig. 7(b): expanded vertical axis]. The increased inter-subject variation in the results at the lower stimulus levels suggests that any correlations between performance in the two tasks may be stronger at lower levels.

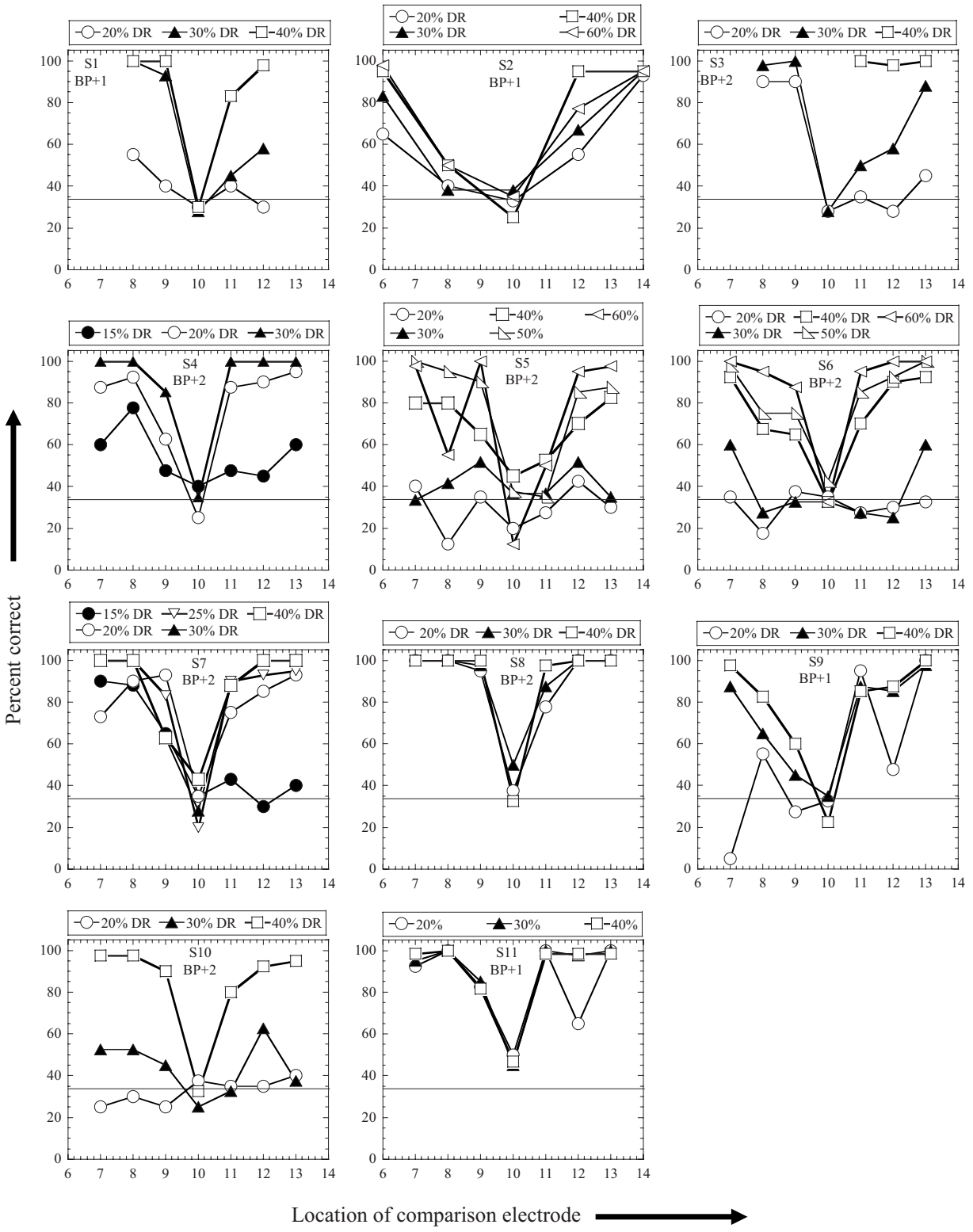


FIG. 3. Percent correct in the electrode discrimination task, obtained using the bipolar stimulation mode. Each panel shows results obtained with a different subject. Within each panel, the parameter is stimulus level (in % DR). The abscissa shows the location of the comparison electrode (the reference was always at electrode 10).

D. Relation between measures of electrode discrimination and modulation sensitivity

The EDI measures described above were used as indices of sensitivity in electrode discrimination. The MDT obtained at the corresponding level served as an index of modulation sensitivity. The data were analyzed for correlations between the EDI obtained at a particular level and the MDT obtained at that same level.

Figure 8 shows results obtained in bipolar stimulation mode. Figure 8(a) shows a scatterplot of the EDI_2 obtained at

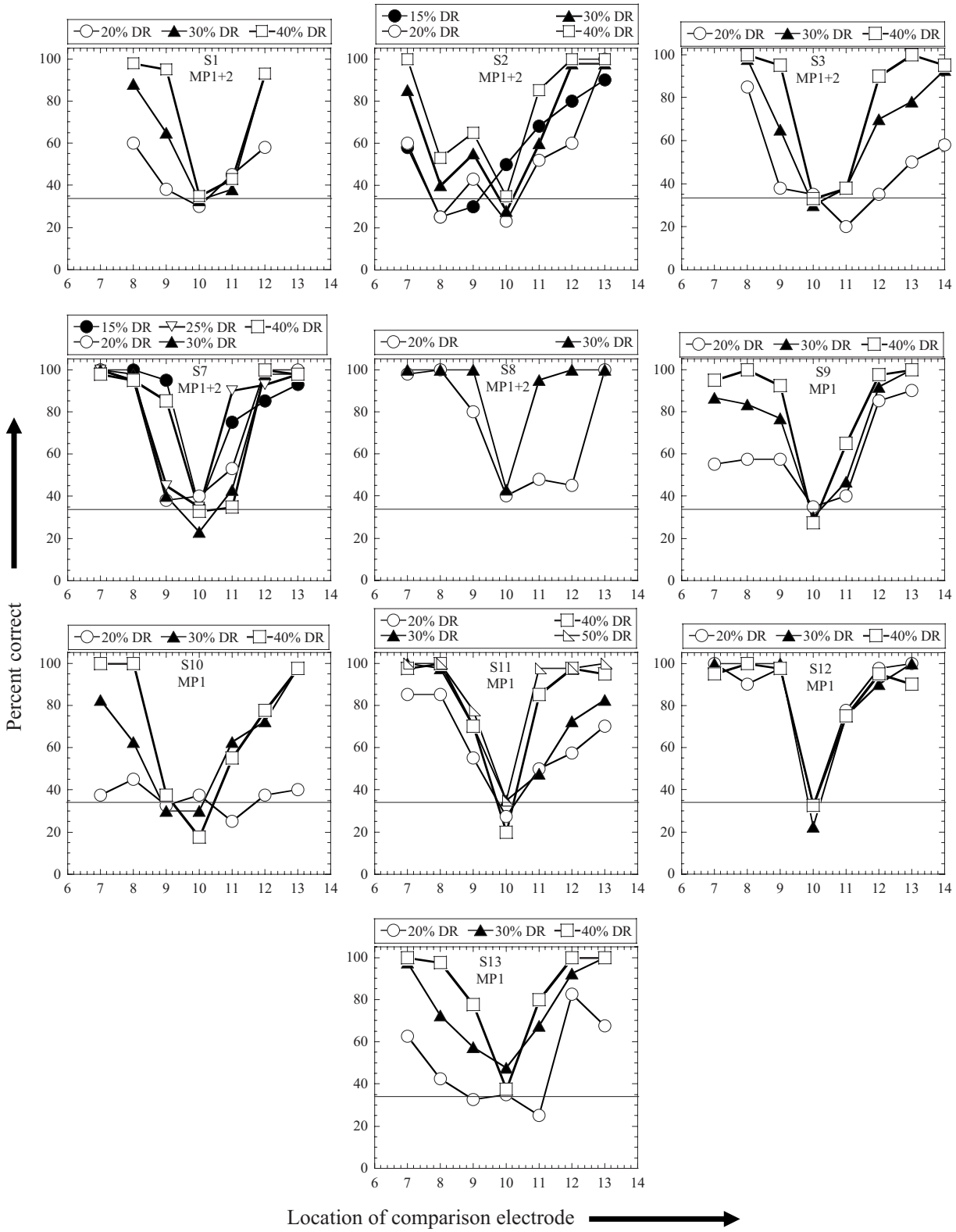


FIG. 4. Percent correct scores in the electrode discrimination task obtained using the monopolar stimulation mode. Otherwise, identical to Fig. 2.

20% DR against the 20% DR MDT obtained at the 100 Hz modulation frequency. The intermittent line indicates the result of a linear regression. The two measures were significantly correlated ($r=-0.621$, $p=0.042$). Figure 8(b) shows the 30% DR EDI_2 measure plotted against the 30% DR 100-Hz MDT. These two measures were again significantly correlated ($r=-0.709$, $p=0.015$). When the modulation frequency was 10 Hz, the correlations fell below significance at both the 20% and 30% DR levels.

Similar analyses performed using the EDI_1 mode showed a significant correlation at the 20% DR level: that is, the 20% DR EDI_1 and the 20% DR 100 Hz MDT were sig-

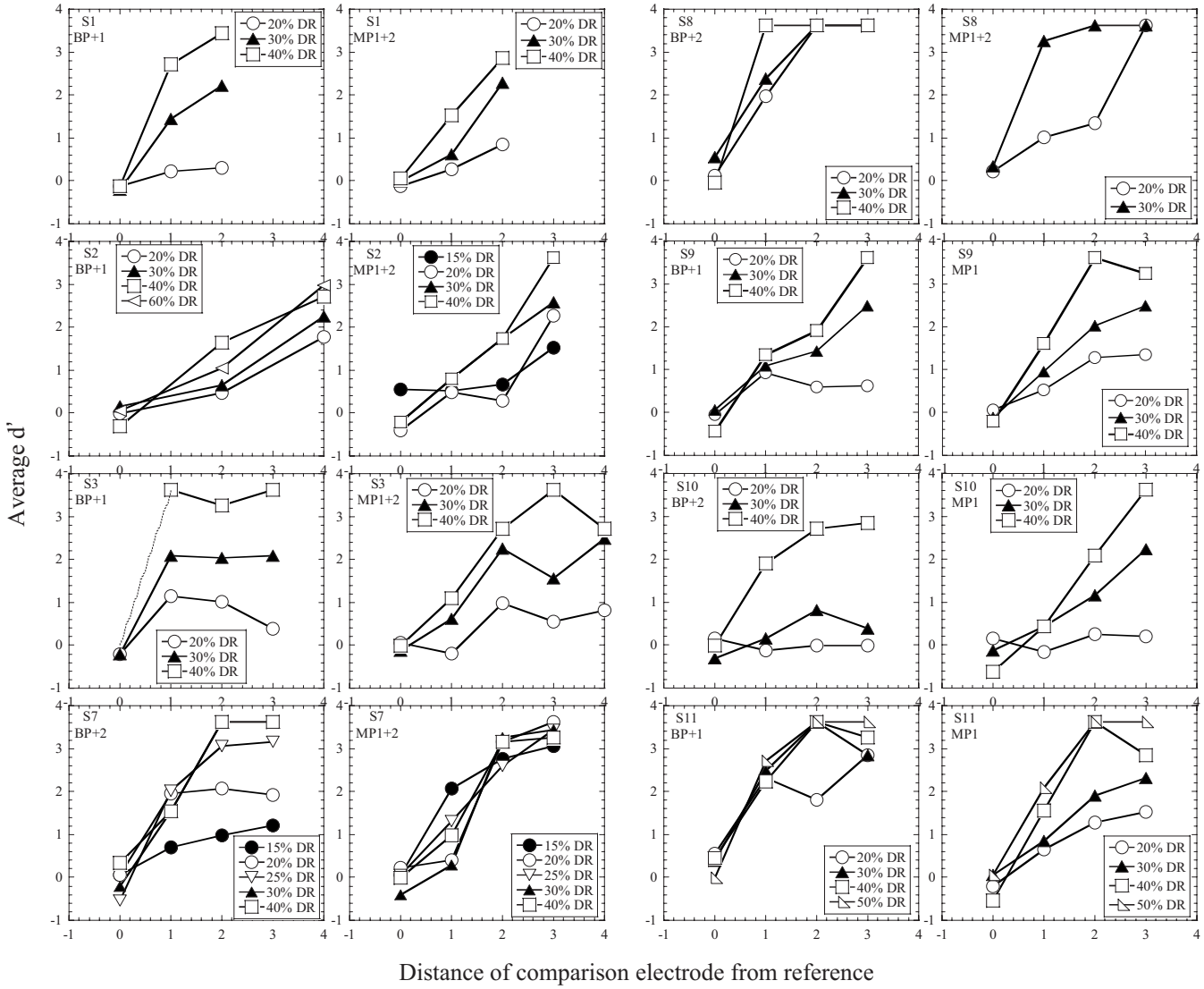


FIG. 5. Performance in the electrode discrimination task as a function of (absolute) distance from the reference electrode (in numbers of electrodes). The ordinate shows the average d' taken across results obtained on equidistant electrodes on the basal and apical sides of the reference. Left-hand panels show results obtained using bipolar stimulation, and right-hand panels show results obtained using monopolar stimulation. Right-hand pairs of panels show results obtained in subjects S1, S2, S3, and S7; left-hand pairs show results obtained in subjects S8, S9, S10, and S11. Within each panel, the parameter is stimulus level.

nificantly correlated with each other ($r=-0.696$, $p=0.025$). However, the corresponding measures were not significantly correlated at the 30% DR level.

The mean of the EDI_1 and EDI_2 measures obtained at 20% DR and 30% DR was computed (EDI_{mean}). The mean of the 100 Hz MDTs measured at 20% and 30% DR (MDT_{mean}) was also computed. These two measures were significantly correlated with each other [$r=-0.689$, $p=0.029$: see Fig 8(c)].

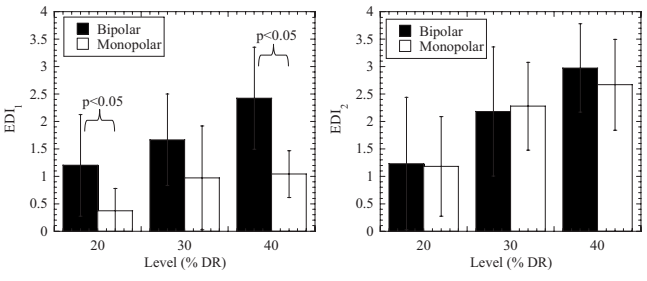
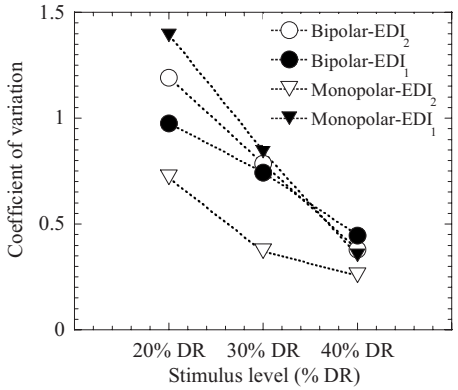


FIG. 6. EDI_1 (left-hand panel) and EDI_2 (right-hand panel) averaged across all eight subjects who provided data in both stimulation modes, plotted against the stimulus level. Dark and light bars represent results obtained in bipolar and monopolar modes, respectively. Error bars show ±1 s.d. Asterisks indicate significant differences between means (paired t-test).

Analyses of data obtained at the 40% DR level showed no significant correlations between any of the measures of electrode discrimination and corresponding measures of modulation sensitivity.

In monopolar mode, significant correlations were observed only in two cases. A significant correlation was observed between the EDI_2 measure at the 20% DR and the 20% DR 100 Hz MDT ($r=-0.678$, $p=0.031$) and also the 20% DR 10 Hz MDT ($r=-0.668$, $p=0.035$). However, no significant correlations were observed between the 20% EDI_1 and either of the 20% MDTs. Similarly, no significant correlations were observed between the EDI_2 index at 30% DR and the 30% DR MDTs, either at 100 Hz or at 10 Hz.



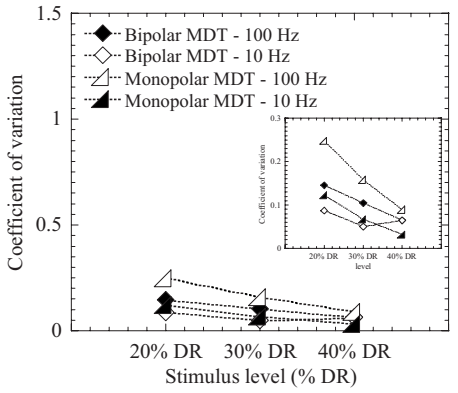
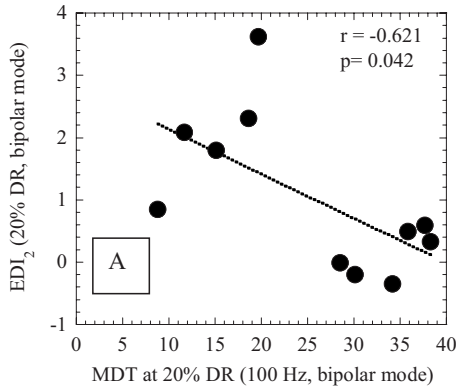
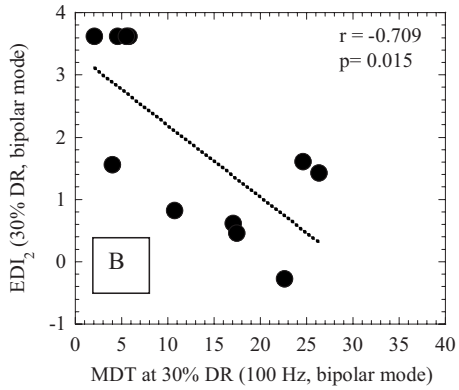


FIG. 7. The coefficient of variation for electrode discrimination measures (upper panel) and modulation detection thresholds (lower panel) plotted as a function of carrier level, for 20%, 30% and 40% DR levels.





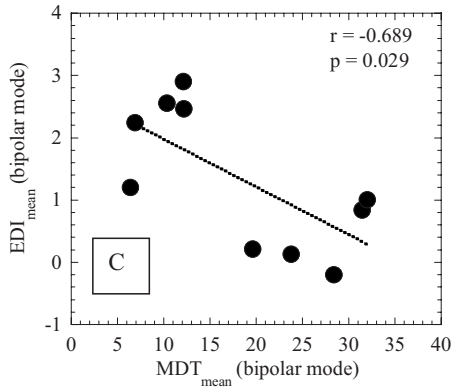


FIG. 8. Scatterplots of EDI_2 obtained at 20% and 30% DRs (A and B, respectively) plotted against the MDT obtained at each level. Panel C shows the EDI_{mean} plotted against the MDT_{mean}.

Consistent with these findings, the EDI_{mean} and MDT_{mean} computed for data in monopolar mode were not significantly correlated with each other.

In summary, the most significant correlations between electrode discrimination and modulation sensitivity were observed at the lower levels, particularly in bipolar stimulation mode and when the modulation rate was 100 Hz. Correlations observed between the EDI measures and modulation sensitivity were generally weaker when the modulation frequency was 10 Hz rather than 100 Hz. At the higher stimulation level of 40% DR, the strength of the correlations weakened. Finally, measures obtained using the monopolar stimulation mode only showed a significant correlation at the lowest level. No other significant correlations were observed in monopolar mode.

Adjustments to the criterion alpha value for significance are sometimes recommended when multiple comparative analyses are conducted on independent variables to investigate several independent hypotheses. Such adjustments should be made cautiously, particularly in studies with small *n*, because as the risk for Type 2 errors declines, the risk for Type 1 errors may increase. In the present case, independent (but related) measures of sensitivity were compared to answer a single question. In this instance, adjustments are not necessary and therefore, no adjustment was made for the criterion significance level (Motulsky, 1995, Chap. 13).

IV. DISCUSSION

The experiments included in this report were designed to examine the relation between modulation sensitivity and electrode discrimination. Effects of stimulation mode, level, and modulation rate on the individual measures and the relations between them, were also investigated. Consistent with previous findings (Shannon, 1992; Fu, 2002; Chatterjee and Oba, 2005; Pfingst *et al.*, 2007; 1999; McKay *et al.*, 1999), results indicate that both modulation sensitivity and electrode discrimination improve with increasing stimulus level. McKay *et al.* (1999) found only small improvements in electrode discrimination between 40% and 70% DRs. This is consistent with our present findings indicating that the largest changes in electrode discrimination occur below 40% DR. The present results are also consistent with the observation of Pfingst *et al.* (1999) showing moderate improvements in electrode discrimination with increasing level in some instances.

Small effects of stimulation mode were observed in the electrode discrimination and modulation detection measures. Significantly improved performance was observed in one of the electrode discrimination measures (EDI_1) in the bipolar mode over the monopolar mode. However, the EDI_2 measure, taken two electrodes away from the signal electrode, did not reveal significant differences between the two modes. When modulation detection thresholds were compared across stimulation mode, a small but significant difference was observed, with mean thresholds in bipolar mode being slightly better than corresponding thresholds in monopolar mode. It is to be noted that Galvin and Fu (2005) did not observe significant effects of stimulation mode on modulation detection thresholds in their study. However, the narrowest bipolar stimulation mode they used was BP+3. It is likely that the effect of stimulation mode becomes even smaller when the bipolar mode is broadened.

At the lower levels of 20% and 30% DRs, significant correlations were observed between 100-Hz modulation thresholds and the EDI_2 measure in the bipolar mode. Observed correlations were weak and generally not significant when the modulation rate was 10 Hz, or in monopolar mode.

At the lower levels and in bipolar stimulation mode, the spread of excitation was likely contained to a small group of neurons: in this case, performance in the electrode discrimination task may have more truly reflected the status/response of local groups of neurons. At higher levels (i.e., 40% DR), excitation possibly spread to other areas, and the psychophysical measures no longer reflected the responses of a small group of neurons. In addition, as stimulation level increased, performance in both tasks approached ceiling. Accordingly, the coefficients of variation in the electrode discrimination and modulation detection threshold data decreased systematically (Fig. 7). This decline in intersubject variation likely contributed to the lack of observed significant correlations among the measures at the 40% DR level.

The fact that more significant correlations were observed in the bipolar mode than in the monopolar mode supports the hypothesis that more focused stimulation modes may target the response properties of local groups of neurons, and possibly contribute to the diagnosis of local regions of damage. In partial support of the assumption that bipolar stimuli provide a more focused field, we observed larger values of d' in the electrode discrimination task in bipolar stimulation than in monopolar stimulation at two of the three stimulation levels analyzed. Note also that the one significant correlation obtained in monopolar mode was observed at the lower level (20% DR), which would be expected to correspond to a narrower stimulating field.

An important caveat to note is that the human data (psychophysical or physiological) do not provide clear support for the idea that the bipolar stimulation mode does, in fact, elicit a narrower excitation pattern than the monopolar mode. Although this has been shown to be the case in animal studies (e.g., Bierer and Middlebrooks, 2002; Snyder *et al.*, 2004), psychophysical results obtained in human CI users are mixed. Some studies suggest that similar widths of the excitation pattern are evoked by bipolar and monopolar stimulation modes (Kwon and van den Honert, 2006; Cohen *et al.*, 2001). However, differences between psychophysical tuning curve shapes obtained in the two modes were noted by Nelson *et al.* (2008). Cohen *et al.* (2001) also noted that the peaks of forward masking patterns obtained in monopolar mode were flatter than those obtained in bipolar mode. The methodology used in some of the human psychophysical experiments makes the results somewhat difficult to interpret. For instance, Kwon and van den Honert (2006) and Nelson *et al.* (2008) used different probe stimulation modes to measure the forward masking pattern/psychophysical tuning curve in monopolar vs bipolar stimulation, making it difficult to compare the masked threshold shifts directly from one to the other mode. Ideally, the probe in a forward masking experiment should be as restricted in spatial/spectral spread as possible, and the same probe should be used to map out the masking pattern or psychophysical tuning curve. Cohen *et al.* (2001) did used a fixed BP+1 probe stimulation mode in their study. However, the masker level was fixed at the MCL: at this relatively high level, differences between the two modes of stimulation may be minimal owing to spread of excitation. Despite the mixed results observed in the psychophysical experiments, the fact remains that the human psychophysical data do not show the dramatic differences in spread of excitation between the two modes that are observed in the animal neurophysiological data. Cohen *et al.* (2003) and Hughes and Stille (2009) demonstrated good correspondence between psychophysical forward masking patterns obtained in humans and parallel measurements obtained using the physiological ECAP response in the same listeners. This suggests that, at least in humans, psychophysical masking patterns provide a good estimate of physiological spread of excitation, and partially rules out the psychophysics-physiology difference as the source of the discrepancy between animal and human data. While methodological differences (electrode design, stimulus parameters, etc.) may be partly to blame, it is difficult to conceive of convincing reasons for the difference in findings. Recently, Schoenecker *et al.* (2009) reported that, in contrast to spatial broad onset responses, steady-state physiological responses recorded in the central nucleus of the inferior colliculus to long-duration, high rate pulse trains are narrower in spatial extent in both monopolar and bipolar stimulations. This indicates that the difference between the two modes depends on the stimulus characteristics. Further research clearly needs to be done to clarify these issues. Overall, the human psychophysical data show—at best—more subtle effects of stimulation mode than those observed in the animal neurophysiological data.

Technical constraints resulted in less-than-ideal conditions in the present experiment. For instance, the bipolar stimulation mode was broader than desired in most cases, bringing it closer to the monopolar mode. Further, there is some evidence that stimuli with shorter pulse phase durations may elicit narrower excitation patterns than those with longer pulse phase durations (Grill and Mortimer, 1996), which would imply that the bipolar stimuli in our experiment were less spatially focused than desired. The computational model used by Grill and Mortimer (1996) produced results

that were supported by experimental data obtained from stimulation of the sciatic nerve and recordings of the torque generated at the ankle joint of adult cats. Stimuli were also different from those used in the present study. Psychophysical results from our own laboratory suggest that excitation patterns do not change greatly with changing pulse phase duration (Chatterjee *et al.*, 2001), but more research is needed to further clarify the issue. If it is the case that shorter pulse durations yield narrower excitation patterns, then in our experiment, the monopolar stimuli produced more spatially focused patterns than desired. Thus, the difference between monopolar and bipolar stimuli would have been reduced further. These limitations may have reduced the strength of the findings: if the experiments had better controlled for pulse phase duration, the bipolar mode would have produced more focused patterns and perhaps resulted in stronger correlations.

It is also true that changing the pulse phase duration and the inter-phase gap can introduce changes in the nature and site of spike initiation, particularly in individuals with surviving peripheral processes (Shepherd *et al.*, 2001; Shepherd and Javel, 1999; van den Honert and Stypulkowski, 1984). It has also been shown that sensitivity to changes in the inter-phase gap and pulse duration is related to spiral ganglion survival in guinea pigs (Prado-Guitierrez *et al.*, 2006). If so, then aspects of temporal coding may be sensitive to the pulse phase duration, as well as the number and health of surviving neurons. For all these reasons, it would be desirable to better control the pulse phase duration in future studies.

A possible explanation for the lack of significant correlations observed when the modulation frequency was 10 Hz, is that encoding the slower modulation frequency does not pose a significant challenge to the auditory system. This is reflected in the greater mean sensitivity observed in the 10 Hz modulation detection task than in the 100 Hz task. Ceiling performance was reached at relatively low stimulus levels in the modulation detection task at 10 Hz modulation rate. At 100 Hz, phase-locking is likely to be less efficient; thus, this task may have been more useful as an indicator of nerve damage; if the correlation between the two measures reflects the status of the nerve, it is reasonable to expect that the 100 Hz task would be more predictive than the 10 Hz task.

Unlike the EDI_2 measure, the EDI_1 measures did not show strong correlations with the MDT measures at the 30% DR level in bipolar mode. We do not have a convincing explanation for this observation at this time.

To summarize, the results presented here suggest that locally specific measures of spatial and temporal resolutions are correlated in CI listeners. The fundamental mechanism underlying the correlation is not known: however, we speculate that auditory-nerve survival at specific locations may dictate performance in both kinds of measures.

The present results provide support for the notion that the correlations reported between measures of modulation sensitivity or spectral resolution and phoneme recognition scores may reflect the contribution of both temporal and spectral cues to speech perception with CIs. It is to be noted that the subjects in Fu's (2002) study were all tested in the bipolar stimulation mode; based on the results of the present study, the correlation between spectral and temporal resolutions is strongest under these conditions.

Although the correlations observed here were significant only at the lower stimulation levels and with relatively simple stimuli, the results have implications for CI listeners' performance in speech perception. The electrode discrimination and modulation detection tasks presented here define the least challenging task listeners may face in the spectral and temporal domains, respectively. In everyday listening, CI users must perform far more challenging tasks involving the discrimination and identification of multi-channel stimuli that are only subtly different in their spectro-temporal patterns.

Difficulties revealed in a simple electrode discrimination task are likely to be amplified when listeners are faced with the need to analyze more complex signals such as speech. Also, multi-channel loudness summation generally results in lower levels of per-channel stimulation; thus, listeners' psychophysical performance at low levels of stimulation may be more predictive of their performance in multi-channel auditory perception.

We would caution against the conclusion that no relationship exists between electrode discrimination and modulation sensitivity in monopolar stimulation: it is possible that a weaker relationship does exist between the two measures, which a study with a large pool of subjects might discover. The present results merely indicate that the relationship between the two is stronger in bipolar stimulation than in monopolar stimulation. The observed relationship between spatial and temporal resolutions in the bipolar mode also does not imply that temporal and spectral cues are processed by the same neural coding mechanisms. All that we can infer from the findings of the present study is that both measures may be limited by some common underlying factor. Further research is needed to discover whether this "factor" is linked to nerve survival, as we speculate, or not. We further speculate that suprathreshold measures of spatial/spectral and/or temporal resolution, such as those used in this study using lower stimulation levels and/or narrow stimulation modes, would serve as useful indicators of the functional health of local populations of neurons in cochlear implant patients.

ACKNOWLEDGMENTS

We are grateful to our research participants for their time and support of our work. We thank Mark E. Robert for software development. We thank the A. E. and an anonymous reviewer for their comments, which resulted in substantial improvements to the manuscript. We are grateful to Kara C. Schvartz for her comments on an earlier version of the manuscript. This work was funded by NIDCD Grant No. R01-DC004786 to M.C.

Bierer, J. A., and Middlebrooks, J. C. (**2002**). "Auditory cortical images of cochlear-implant stimuli: Dependence on electrode configuration," J. Neurophysiol. **87**, 478–492.

Cazals, Y., Pelizzone, M., Saudan, O., and Boex, C. (**1994**). "Low-pass filtering in amplitude modulation detection associated with vowel and consonant identification in subjects with cochlear implants," J. Acoust. Soc. Am. **96**, 2048–2054.

Chatterjee, M., and Robert, M. E. (**2001**). "Noise enhances modulation sensitivity in cochlear implant listeners: Stochastic resonance in a prosthetic sensory system?," J. Assoc. Res. Otolaryngol. **2**, 159–171.
Chatterjee, M., and Oba, S. I. (**2005**). "Noise improves modulation detection by cochlear implant listeners at moderate carrier levels," J. Acoust. Soc. Am. **118**, 993–1002.
Chatterjee, M., Shannon, R. V., Galvin, J. J., and Fu, Q.-J. (**2001**). "Spread of excitation and its effect on auditory perception with cochlear implants," in *Physiological and Psychophysical Bases of Auditory Function*, Proceedings of the 12th International Symposium on Hearing, edited by A. J. M. Houtsma, A. Kohlrausch, V. F. Prijs, and R. Schoonhoven (Shaker, Maastricht).
Cohen, L. T., Saunders, E., and Clark, G. M. (**2001**). "Psychophysics of a prototype perimodiolar cochlear implant electrode array," Hear. Res. **155**, 63–81.
Cohen, L. T., Richardson, L. M., Saunders, E., and Cowan, R. S. (**2003**). "Spatial spread of neural excitation in cochlear implant recipients: Comparison of improved ECAP method and psychophysical forward masking," Hear. Res. **179**, 72–87.
Collins, L. M., Zwolan, T. A., and Wakefield, G. H. (**1997**). "Comparison of electrode discrimination, pitch ranking, and pitch scaling data in postlingually deafened adult cochlear implant subjects," J. Acoust. Soc. Am. **101**, 440–454.
Donaldson, G. S., and Nelson, D. A. (**2000**). "Place-pitch sensitivity and its relation to consonant recognition by cochlear implant listeners using the MPEAK and SPEAK speech processing strategies," J. Acoust. Soc. Am. **107**, 1645–1658.
Fu, Q.-J. (**2002**). "Temporal processing and speech recognition in cochlear implant users," NeuroReport **13**, 1635–1639.
Galvin, J. J., III, and Fu, Q.-J. (**2005**). "Effects of stimulation rate, mode and level on modulation detection by cochlear implant users," J. Assoc. Res. Otolaryngol. **6**, 269–279.
Grill, W. M., Jr., and Mortimer, J. T. (**1996**). "The effect of stimulus pulse duration on selectivity of neural stimulation," IEEE Trans. Biomed. Eng. **43**, 161–166.
Henry, B. A., and Turner, C. W. (**2003**). "The resolution of complex spectral patterns by cochlear implant and normal-hearing listeners," J. Acoust. Soc. Am. **113**, 2861–2873.
Henry, B. A., McKay, C. M., McDermott, H. J., and Clark, G. M. (**2000**). "The relationship between electrode discrimination and speech perception in cochlear implantees," J. Acoust. Soc. Am. **108**, 1269–1280.
Henry, B. A., Turner, C. W., and Behrens, A. (**2005**). "Spectral peak resolution and speech recognition in quiet: normal hearing, hearing impaired, and cochlear implant listeners," J. Acoust. Soc. Am. **118**, 1111–1121.
Hughes, M. L., and Stille, L. J. (**2009**). "Psychophysical and physiological measures of electrical-field interaction in cochlear implants," J. Acoust. Soc. Am. **125**, 247–260.
Kwon, B. J., and van den Honert, C. (**2006**). "Effect of electrode configuration on psychophysical forward masking in cochlear implant listeners," J. Acoust. Soc. Am. **119**, 2994–3002.
Levitt, H. (**1971**). "Transformed up-down methods in psychoacoustics," J. Acoust. Soc. Am. **49**, 467–477.
Litvak, L. M., Spahr, A. J., Saoji, A. A., and Fridman, G. Y. (**2007**). "Relationship between perception of spectral ripple and speech recognition in cochlear implant and vocoder listeners," J. Acoust. Soc. Am. **122**, 982–991.
Macmillan, N. A., and Creelman, C. D. (**2005**). *Detection Theory: A User's Guide*, 2nd ed. (Lawrence Erlbaum Associates, Mahwah, NJ).
McKay, C. M., O'Brien, A., and James, C. J. (**1999**). "Effect of current level on electrode discrimination in electrical stimulation," Hear. Res. **136**, 159–164.
Motulsky, H. (**1995**). *Intuitive Biostatistics* (Oxford University Press, New York).
Nelson, D. A., Donaldson, G. S., and Kreft, H. (**2008**). "Forward-masked spatial tuning curves in cochlear-implant users," J. Acoust. Soc. Am. **123**, 1522–1543.
Nelson, D. A., Van Tassell, D. J., Schroeder, A. C., Soli, S., and Levine, S. (**1995**). "Electrode ranking of "place pitch" and speech recognition in electrical hearing," J. Acoust. Soc. Am. **98**, 1987–1999.
Pfingst, B. E., Xu, L., and Thompson, C. S. (**2007**). "Effects of carrier pulse rate and stimulation site on modulation detection by subjects with cochlear implants," J. Acoust. Soc. Am. **121**, 2236–2246.
Pfingst, B. E., Holloway, L. A., Zwolan, T. A., and Collins, L. M. (**1999**). "Effects of stimulus level on electrode-place discrimination in human subjects with cochlear implants," Hear. Res. **134**, 105–115.
Prado-Guitierrez, P., Fewster, L. M., Heasman, J. M., McKay, C. M., and Shepherd, R. K. (**2006**). "Effect of interphase gap and pulse duration on electrically evoked potentials is correlated with auditory nerve survival," Hear. Res. **215**, 47–55.
Robert, M. E. (**2002**). *House Ear Institute Nucleus Research Interface User's Guide*, House Ear Institute, Los Angeles.
Schoenecker, M., Stakhovskaya, O., Bonham, B., Snyder, R., Leake, P. (**2009**). "Monopolar intracochlear pulse trains can elicit focal central activation," in 2009 Conference on Implantable Auditory Prostheses (A).
Shannon, R. V. (**1992**). "Temporal modulation transfer functions in patients with cochlear implants," J. Acoust. Soc. Am. **91**, 2156–2164.
Shannon, R. V., Adams, D. D., Ferrel, R. L., Palumbo, R. L., and Grandgenett, M. (**1990**). "A computer interface for psychophysical and speech research with the nucleus cochlear implant" J. Acoust. Soc. Am. **87**, 905–907.
Shepherd, R. K., and Javel, E. (**1997**). "Electrical stimulation of the auditory nerve. I. Correlation of physiological responses with cochlear status," Hear. Res. **108**, 112–144.
Shepherd, R. K., and Javel, E. (**1999**). "Electrical stimulation of the auditory nerve: II. Effect of stimulus waveshape on single fibre response properties," Hear. Res. **130**, 171–188.
Shepherd, R. K., Hardie, N. A., and Baxi, J. H. (**2001**). "Electrical stimulation of the auditory nerve: single neuron strength-duration functions in deafened animals," Ann. Biomed. Eng. **29**, 195–201.
Shepherd, R. K., Roberts, L. A., and Paolini, A. G. (**2004**). "Long-term sensorineural hearing loss induces functional changes in the rat auditory nerve," Eur. J. Neurosci. **20**, 3131–3140.
Snyder, R. L., Bierer, J. A., and Middlebrooks, J. C. (**2004**). "Topographic spread of inferior colliculus activation in response to acoustic and intracochlear electric stimulation," J. Assoc. Res. Otolaryngol. **5**, 305–322.
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. (**1984**). "Physiological responses of the electrically stimulated auditory nerve. II. Single fiber recordings," Hear. Res. **14**, 225–243.
Won, J. H., Drennan, W. R., and Rubinstein, J. T. (**2007**). "Spectral-ripple resolution correlates with speech reception in noise in cochlear implant users," J. Assoc. Res. Otolaryngol. **8**, 384–392.
Xu, L., Thompson, C. S., and Pfingst, B. E. (**2005**). "Relative contributions of spectral and temporal cues in phoneme recognition," J. Acoust. Soc. Am. **117**, 3255–3267.
Zwolan, T. A., Collins, L. M., and Wakefield, G. H. (**1997**). "Electrode discrimination and speech recognition in postlingually deafened adult cochlear implant subjects," J. Acoust. Soc. Am. **102**, 3673–3685.

On the importance of preserving the harmonics and neighboring partials prior to vocoder processing: Implications for cochlear implants

Yi Hu[a)]
Department of Electrical Engineering and Computer Science, University of Wisconsin–Milwaukee, Milwaukee, Wisconsin 53201 and Department of Electrical Engineering, The University of Texas–Dallas, Richardson, Texas 75080

Philipos C. Loizou
Department of Electrical Engineering, The University of Texas–Dallas, Richardson, Texas 75080

(Received 22 June 2009; revised 6 October 2009; accepted 15 October 2009)

Pre-processing based noise-reduction algorithms used for cochlear implants (CIs) can sometimes introduce distortions which are carried through the vocoder stages of CI processing. While the background noise may be notably suppressed, the harmonic structure and/or spectral envelope of the signal may be distorted. The present study investigates the potential of preserving the signal's harmonic structure in voiced segments (e.g., vowels) as a means of alleviating the negative effects of pre-processing. The hypothesis tested is that preserving the harmonic structure of the signal is crucial for subsequent vocoder processing. The implications of preserving either the main harmonic components occurring at multiples of F0 or the main harmonics along with adjacent partials are investigated. This is done by first pre-processing noisy speech with a conventional noise-reduction algorithm, regenerating the harmonics, and vocoder processing the stimuli with eight channels of stimulation in steady speech-shaped noise. Results indicated that preserving the main low-frequency harmonics (spanning 1 or 3 kHz) alone was not beneficial. Preserving, however, the harmonic structure of the stimulus, i.e., the main harmonics along with the adjacent partials, was found to be critically important and provided substantial improvements (41 percentage points) in intelligibility. © *2010 Acoustical Society of America.* [DOI: 10.1121/1.3266682]

PACS number(s): 43.66.Ts, 43.71.Ky [MW] Pages: 427–434

I. INTRODUCTION

The performance, in terms of speech understanding, of cochlear implant (CI) users is known to degrade in noisy conditions. Over the years, many researchers have shown that the use of noise-reduction (NR) algorithms as a pre-processing step is an effective approach to improve speech recognition in noisy listening conditions for unilateral cochlear-implant listeners (Hochberg *et al.*, 1992; Weiss, 1993; Yang and Fu, 2005; Loizou *et al.*, 2005; Hu *et al.*, 2007) as well as for bilateral implant users and CI users wearing two microphones (van Hoesel and Clark, 1995; Hamacher *et al.*, 1997; Wouters and Berghe, 2001; Kokkinakis and Loizou, 2008). Hochberg *et al.* (1992) used the INTEL noise-reduction algorithm to pre-process speech and presented the processed speech to ten Nucleus implant users fitted with the F0/F1/F2 and MPEAK feature-extraction strategies. Consonant-vowel-consonant words embedded in speech-shaped noise at signal to noise ratios (SNRs) in the range of −10 to 25 dB were presented to the CI users. Significant improvements in performance were obtained at SNRs as low as 0 dB. The improvement in performance was attributed to more accurate formant extraction, as the INTEL algorithm reduced the errors caused by the feature-extraction algorithm. Yang and Fu (2005) evaluated the performance of a spectral-subtractive algorithm using subjects wearing the Nucleus-22, Med-El, and Clarion devices. Significant benefits in sentence recognition were observed for all subjects with the spectral-subtractive algorithm, particularly for speech embedded in speech-shaped noise. Loizou *et al.* (2005) evaluated a subspace noise-reduction algorithm which was based on the idea that the noisy speech vector can be projected onto "signal" and "noise" subspaces. The clean signal was estimated by retaining only the components in the signal subspace and nulling the components in the noise subspace. The performance of the subspace reduction algorithm was evaluated using 14 subjects wearing the Clarion device. Results indicated that the subspace algorithm produced significant improvements in sentence recognition scores compared to the subjects' daily strategy, at least in continuous (stationary) noise.

All the above methods were based on pre-processing the noisy signal and presenting the "enhanced" signal to the CI users. Pre-processing techniques, however, can introduce distortions which will be subsequently carried out and introduced in the vocoder stages of processing. Pre-processing can notably suppress the background noise, but can distort the harmonic structure and/or spectral envelope of the signal. The present study focuses on the development of techniques aimed at alleviating the negative effects of pre-processing. In particular, it investigates the potential of preserving the signal's harmonic structure present primarily in voiced segments (e.g., vowels). Both the masking noise and noise-reduction algorithm can degrade the harmonics structure, as

a)Author to whom correspondence should be addressed. Electronic mail: huy@uwm.edu

FIG. 1. (Color online) Block diagram of the overall system.

most noise-reduction algorithms are designed to recover the spectral envelope, while paying little attention to the harmonics. The present study maintains the hypothesis that preserving the harmonic structure of the clean signal is crucial for subsequent vocoder processing. The implications of preserving either the main harmonic components or the main harmonics along with adjacent partials are investigated in the present study. This is done by first pre-processing noisy speech with a conventional noise-reduction algorithm, regenerating the harmonics, and vocoder processing the stimuli with eight channels of stimulation in steady speech-shaped noise at 0–6 dB SNR. The experiments in this study were designed to assess the importance of preserving the signal's harmonic structure prior to vocoder (CI) processing.

II. EXPERIMENT: NOISE REDUCTION AND HARMONICS REGENERATION

A. Methods

1. Subjects

Seven normal-hearing native speakers of American English participated in this experiment. All subjects were paid for their participation, and all of them were undergraduate and graduate students at the University of Texas-Dallas.

2. Stimuli

The target speech materials consisted of sentences from the IEEE database (IEEE, 1969) and were obtained from Loizou (2007). The IEEE corpus contains 72 lists of ten phonetically balanced sentences produced by a male speaker and recorded in a double-walled sound-attenuation booth at a sampling rate of 25 kHz. Further details about the speech recordings can be found in Loizou (2007). The estimated F0 values of the male speaker ranged from 75 to 250 Hz with a mean of 127.57 Hz and a standard deviation of 21.16 Hz.

The masker was steady speech-shaped noise and had the same long-term spectrum as the sentences in the IEEE corpus. Speech-shaped noise was selected as its stationarity minimizes the confounding effect of the accuracy of noise estimation algorithms.

3. Signal processing

The experiments were designed to evaluate the benefits of harmonics regeneration when used in a pre-processing stage to vocoder (cochlear-implant) processing. Figure 1 shows the block diagram of the overall system. A total of six processing conditions were used for this purpose. The first condition was designed to simulate the cochlear-implant processing. As the first step, a pre-emphasis filter with 2000 Hz cutoff and 3 dB/octave rolloff was applied to the signal. An eight-channel noise-excited vocoder was utilized (Shannon *et al.*, 1995). The speech signal was bandpassed into eight frequency bands between 80 and 6000 Hz using sixth-order Butterworth filters. For the specified frequency range, the equivalent rectangular bandwidth (ERB) filter spacing (Glasberg and Moore, 1990) was used to allocate the eight frequency channels (the channel allocation is shown in Table I). The envelopes of the bandpassed signals were obtained by full-wave rectification followed by low-pass filtering using a second-order Butterworth filter with a 400 Hz cut-off frequency. This cut-off frequency was chosen to preserve F0 modulations in the envelopes. The extracted temporal envelopes were modulated with white noise, and bandpass filtered through the same analysis bandpass filters. The resulting (narrow-band filtered) waveforms in each channel were finally summed to generate the stimuli. The level of the synthesized speech signal was scaled to have the same root mean square value as the original speech signal.

TABLE I. Low and high cut-off frequencies (at −3 dB) for the eight channels used in the vocoding stage.

Channel	Low (Hz)	High (Hz)
1	80	221
2	221	426
3	426	724
4	724	1158
5	1158	1790
6	1790	2710
7	2710	4050
8	4050	6000

The other five conditions involved two pre-processing steps prior to vocoding processing (see Fig. 1). The first processing condition involved a NR algorithm based on the minimum mean square error log-spectral amplitude estimation (LogMMSE) proposed by Ephraim and Malah (1985). The LogMMSE algorithm was chosen as this noise-reduction method performed well in both speech quality and speech intelligibility studies (Hu and Loizou, 2007a, 2007b). The same noise estimation algorithm as in Hu and Loizou (2007a, 2007b) was used for estimating/updating the masker spectrum. Fast Fourier transform (FFT) based frame processing was used in the implementation of the LogMMSE algorithm. Speech signals were segmented into 50% overlapping frames using a sliding 20-ms Hanning window. Figure 2 shows the block diagram for the processing. A 8192-point FFT (by zero padding) with a frequency bin resolution of 3.05 Hz was utilized.

The second pre-processing step was designed to evaluate the benefits of harmonics regeneration performed after the noise-reduction stage (see Fig. 2). As mentioned earlier, the rationale for this step is to alleviate any distortions introduced by the noise-reduction algorithm. The majority of phonetic segments are voiced segments (Mines *et al.*, 1978) which can be approximately modeled by harmonic spectra. The harmonics appear at integer multiples of F0.

In order to establish an upper bound and evaluate the potential of the harmonics regeneration stage when combined with noise reduction, we assumed an ideal operating environment. That is, we estimated the F0 from the clean speech signal and regenerated the signal's harmonics with prior knowledge of the clean speech spectrum. More specifi-

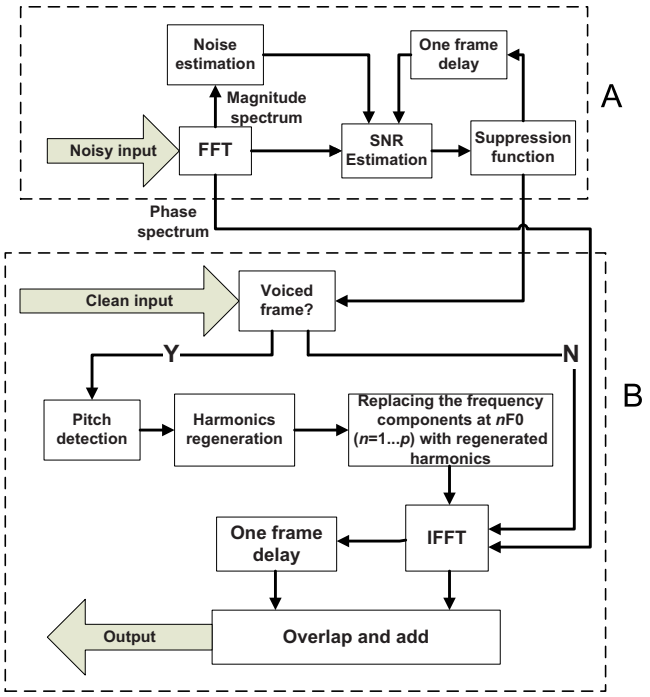


FIG. 2. (Color online) Block diagrams for the combined harmonics-regeneration stage (shown in block B) and noise-reduction stage (shown in block A).

cally, an F0-detection algorithm based on the autocorrelation function (Kondoz, 1999, Chap. 6) was used to obtain the F0 in each frame. The number of regenerated harmonics was then calculated by $p=\lfloor \mathrm{CF}/\mathrm{F0} \rfloor$, where CF is the cut-off frequency below which harmonics are included, and $\lfloor \cdot \rfloor$ is the floor operator. Two cut-off frequency values, 1000 and 3000 Hz, were evaluated. To compensate for the possible inaccuracy of the F0 detector, harmonics were regenerated by extracting the local peaks in a 30-Hz range around nF0, where $n=1,\ldots,p$. The extracted harmonics had a quantization error of roughly 1.53 Hz (half of the FFT frequency resolution). Figure 2 shows the block diagram for the combined noise-reduction (block A) and harmonics-regeneration (block B) stages.

The magnitude spectra of voiced phonetic segments (e.g., vowels) possess a harmonic structure. The harmonics are evident at multiples of F0. In addition, sideband components or partials, falling between the main harmonic components (which occur primarily at multiples of F0), are also present in voiced magnitude spectra. To assess the importance of preserving the harmonic structure of voiced segments, two conditions were created. In the first condition, only the main harmonic amplitudes were included (the partials were left noise-suppressed), while in the second condition, both the main harmonics and neighboring partials were included. We denote the first condition as Main-xkHz, where x denotes the cut-off frequency (1 or 3 kHz) up to which harmonics are included, and the second condition in which both the main harmonics and neighboring partials are included as PartH-xkHz. The main harmonic amplitudes were extracted from the clean magnitude spectrum based on the estimated F0 value. The partials were not extracted from the clean spectra. Instead, a simple approach was taken to generate the neighboring partials. This was done by multiplying the main harmonic amplitudes by a Gaussian-shaped function (see Fig. 3) and sampling the Gaussian function at 16 discreet frequencies to the left and right of the main harmonics. Note that the FFT resolution was 3 Hz; hence the Gaussian function spanned a total bandwidth of 100 Hz. This bandwidth was chosen to accommodate the F0 of the male speaker. The Gaussian function was derived heuristically by inspecting the magnitude spectra of several frames of voiced segments. More complex algorithms could alternatively be used to generate the Gaussian function; however, we chose the function shown in Fig. 3 for its simplicity and practical implications. In a realistic implementation, the partials do not need to be estimated from the noisy signal, only the main harmonics need to be estimated.

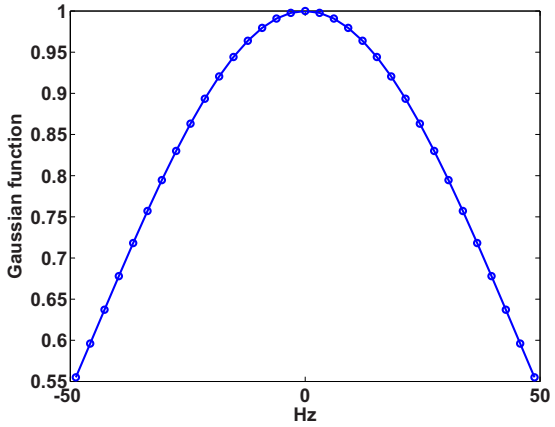


FIG. 3. (Color online) Gaussian-shaped function used for generating partials adjacent to the main harmonics.

Figure 4 shows example plots of the FFT magnitude

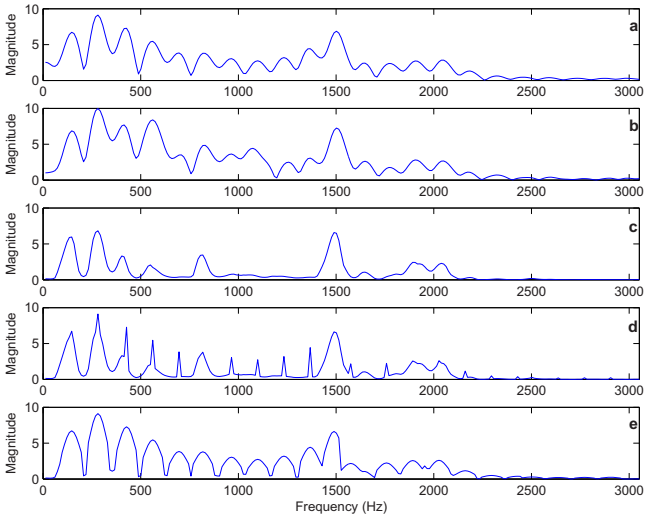


FIG. 4. (Color online) Example FFT magnitude spectra (displayed in linear units) of a voiced segment extracted from a sentence (for better clarity, only the spectrum spanning the frequency range 0–3 kHz is displayed). The top panel shows the clean speech spectrum. The second panel shows the noisy speech spectrum (SNR=0 dB) and the third panel shows the enhanced speech spectrum after applying the logMMSE noise-reduction algorithm. The fourth panel shows the harmonics-regenerated speech spectrum based only on the main harmonics occurring at multiples of F0. The bottom panel shows the harmonics-regenerated speech spectrum based on both the main harmonics and adjacent partials.

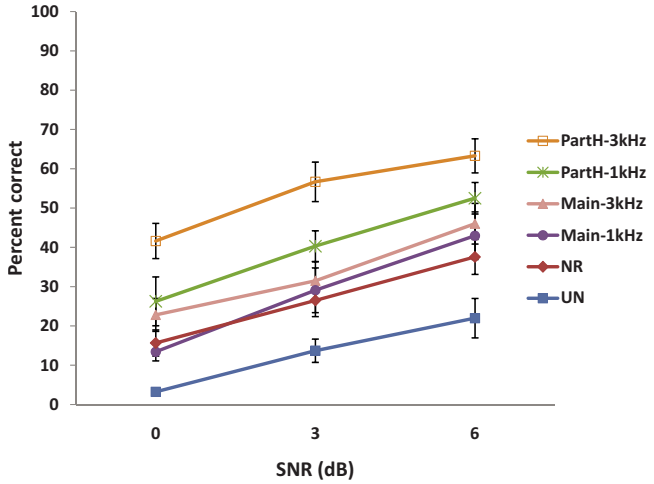


FIG. 5. (Color online) Mean percent correct scores as a function of SNR level. The error bars denote ±1 standard error of the mean.

spectra of the clean, noisy, and enhanced signals, as well as signals with harmonics regenerated. As can be seen from panel (c), although the noise-reduction algorithm suppressed the background noise, the harmonics structure was degraded. Panel (d) in Fig. 4 shows the spectra with only the main harmonics regenerated, and panel (e) shows the spectrum with both the main harmonics and partials regenerated. Clearly, the spectrum shown in panel (e) resembles closer to the clean spectrum (panel a) than the output spectrum (panel c) produced by the noise-reduction algorithm.

4. Procedure

The listening tests were conducted using a personal computer connected to a Tucker-Davis system 3. Stimuli were played monaurally to the subjects through Sennheiser HD 250 Linear II circumaural headphones at a comfortable listening level. The subjects were seated in a double-walled sound-attenuation booth (Acoustic Systems, Inc.). To familiarize each subject with the stimuli, a training session was administrated prior to the formal testing, and each subject listened to vocoded speech stimuli. The training session typically lasted about 15–20 min. During the testing session, the subjects were instructed to write down the words they heard. In total, there were 18 testing conditions (=3 SNR levels ×6 processing methods). For each condition, two lists of sentences were used, and none of the lists was repeated across the testing conditions. The conditions were presented in random order for each subject. The subjects were allowed to take breaks whenever they wanted and no feedback was provided after each testing condition.

B. Results

The mean percent correct scores for all conditions are shown in Fig. 5. Performance was measured in terms of percent of words identified correctly (all words were scored). The scores were first converted to rational arcsine units (RAU) using the rationalized arcsine transform proposed by Studebaker (1985). To examine the effect of cut-off frequency (1 kHz vs 3 kHz) and type of harmonic structure (main harmonics only vs main harmonics plus partials) preserved, we subjected the scores to statistical analysis using the transformed score as the dependent variable, and the SNR levels, cut-off frequency, and type of harmonic structure as the three within-subjects factors. Analysis of variance with repeated measures indicated significant effects of SNR levels [$F(2,12)=20.72$, $p<0.001$], significant effects of cut-off frequency [$F(1,6)=102.80$, $p<0.001$], and significant effects of type of harmonic structure [$F(1,6)=73.14$, $p=0.002$]. There were no significant between-factor interactions except the one between cut-off frequency and type of harmonic structure [$F(1,6)=63.13$, $p=0.039$]. Results indicated that a higher cut-off frequency and inclusion of partials provided additional benefits compared to those obtained using the noise-reduction algorithm alone.

TABLE II. Multiple paired comparisons between the scores obtained in the various conditions.

	0 dB	3 dB	6 dB
NR vs UN	**[a]	**	**
PartH-1 kHz vs NR		**	**
PartH-3 kHz vs NR	**	**	**
PartH-3 kHz vs PartH-1 kHz		**	**

[a]Bonferroni corrected $p<0.0125$, $\alpha=0.05$.

Multiple paired comparisons, with Bonferroni correction, were run between the scores obtained with the corrupted (unprocessed, denoted as UN) and NR algorithm, NR and PartH-1 kHz, NR and PartH-3 kHz, and PartH-3 kHz and PartH-1 kHz at various SNR levels. The Bonferroni corrected statistical significance level was set at $p<0.0125$ ($\alpha=0.05$). The results are shown in Table II. The comparisons indicated statistically significant differences between the UN and NR scores at all three SNR levels, suggesting that the NR algorithm used in this study can provide benefit for vocoded speech in steady-state noise. The scores obtained with the PartH-1 kHz stimuli at lower SNR levels (0 dB) were not significantly higher ($p=0.1$) than those obtained with the NR scores but were significantly better ($p=0.005$) at higher SNR levels (3 and 6 dB), suggesting that maintaining the signal's harmonics structure below 1000 Hz can further improve the benefits with noise-reduction methods at higher SNR levels. The scores obtained with the 3000 Hz cut-off frequency were significantly higher than those obtained with the 1000 Hz cut-off frequency at higher SNR levels ($p<0.01$), but they did not differ at 0 dB ($p=0.02$), indicating additional benefits when using a higher harmonics-regeneration cut-off frequency.

III. GENERAL DISCUSSION AND CONCLUSIONS

The above results and analysis clearly indicate that significant improvement in intelligibility, relative to NR processing alone, can be obtained when the harmonic structure of the input signal is preserved prior to vocoder processing. In particular, the scores obtained in the PartH-1 kHz and PartH-3 kHz conditions yielded the largest improvements. This was not surprising, since clean harmonic amplitudes spanning the range of 0–1 kHz or 0–3 kHz were used dur-

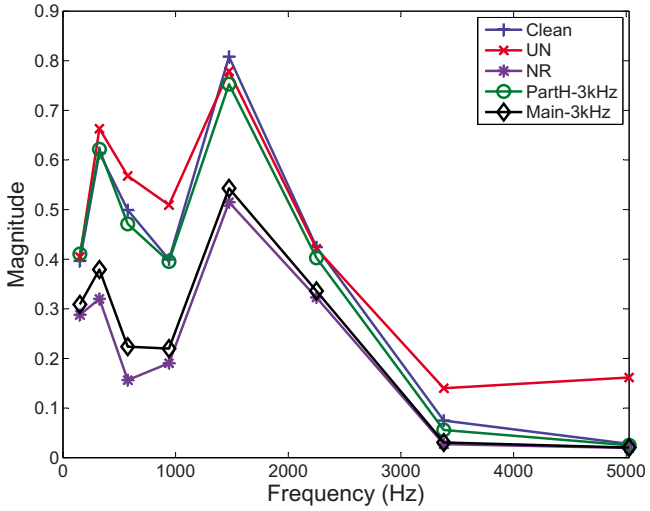


FIG. 6. (Color online) The vocoded spectra of a voiced segment (same as that used in Fig. 4) processed in the various conditions.

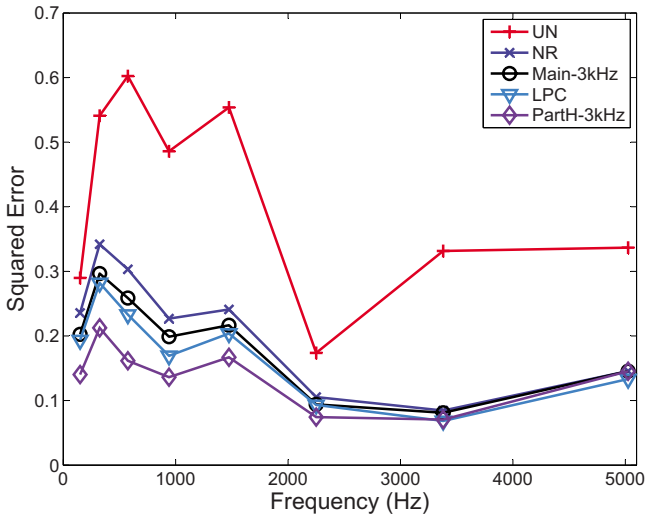


FIG. 7. (Color online) Plots of the squared error between the clean envelope and processed envelopes in the various conditions. The squared error was obtained by averaging, across ten IEEE sentences, the band squared error of processed envelopes.

ing the voiced segments of the corrupted signal. Performance obtained in the PartH-3 kHz condition was significantly higher than that obtained in the PartH-1 kHz condition owing to the fact that the stimuli in the PartH-3 kHz condition preserved to some degree formant frequency (F1 and F2) information. In contrast, the stimuli in the PartH-1 kHz condition preserved primarily F1 information. In effect, the stimuli in PartH- 1 kHz and PartH-3 kHz conditions provided glimpsing[1] of the F1 and F2 information present in the voiced segments (e.g., vowels and semivowels) and thus enabled listeners to identify more words in the otherwise noisy speech stream (unvoiced segments were left corrupted). The PartH-1 kHz outcome is consistent with the outcomes from our prior studies (Li and Loizou, 2007, 2008) that indicated that glimpsing in the low-frequency region (<1000 Hz) can bring substantial benefits to speech intelligibility since listeners had a clear access to the voiced/unvoiced landmarks, which are posited to facilitate syllable/word segmentation (Stevens, 2002).

The most interesting finding from this study is the outcome that performance in the PartH-3 kHz condition was significantly higher than performance in the Main-3 kHz condition. The stimuli in both conditions contained the clean signal's harmonic components spanning the range 0–3 kHz (see example in Fig. 4). The fact that the scores in the Main-3 kHz condition (which only preserved the main harmonic components of the clean signal) did not yield an improvement in intelligibility, relative to the NR condition, suggests that preserving only the main harmonics (i.e., harmonics occurring at multiples of F0) is not sufficient or beneficial, at least in the context of vocoder processing. The introduction of partials [see Fig. 4, panel (e)] adjacent to the main harmonics was found to be necessary to yield substantial improvements in intelligibility. This is because, in the context of vocoder processing, the inclusion of partials (adjacent to the harmonics) yielded channel envelope amplitudes closer to those of the clean signal's envelope amplitudes. This is demonstrated in the example shown in Fig. 6. The accuracy in envelope amplitude estimation is quantified in Fig. 7, in terms of the squared error[2] between the envelope amplitude values of the clean and processed signals. The smallest squared error value (i.e., amplitudes closest to the clean envelope amplitudes) was obtained with the PartH-3 kHz processed signals. Note that the resulting channel envelope amplitudes (following vocoder processing) of the Main-3 kHz stimuli were closer in value to those in the NR stimuli, despite the preservation of the main harmonic components in the Main-3 kHz stimuli. This was consistent with the rather equitable intelligibility scores observed (see Fig. 5) in the Main- 3 kHz and NR conditions. In addition to the use of squared error, we also quantified the fidelity of envelope reconstruction using the metric[3] developed in Sheft *et al.* (2008). The resulting correlation coefficients for each band are shown in Fig. 8. As can be seen, higher correlation coefficient (hence, better envelope reconstruction) is obtained with the PartH-3 kHz processed signals, consistent with the outcome shown in Fig. 7. In summary, preserving

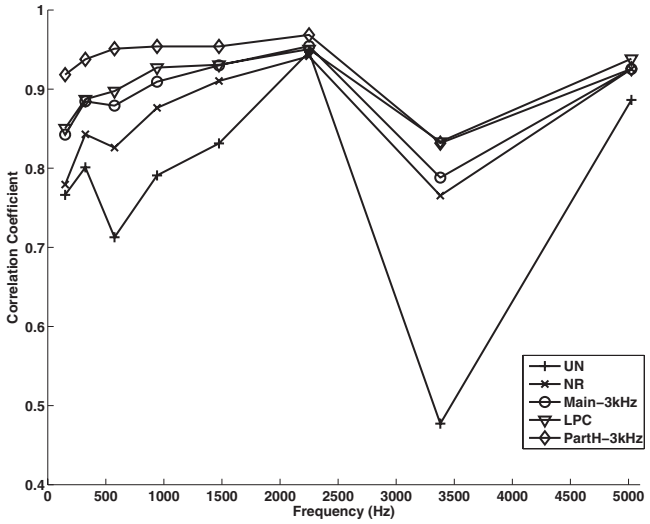


FIG. 8. Plots of the band correlation coefficients of the clean envelopes and processed envelopes.

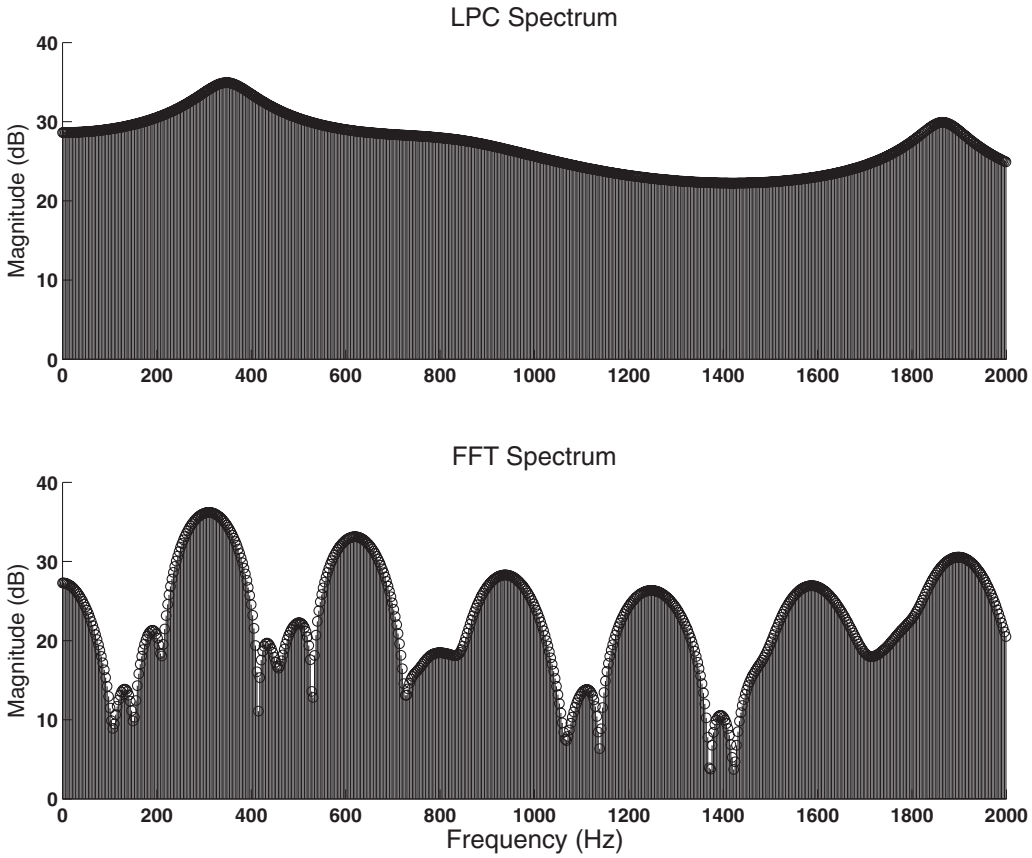


FIG. 9. Top panel shows the LPC spectrum of a voiced segment and bottom panel shows the corresponding FFT spectrum. Segment was taken from a sentence produced by a female speaker. Only the 0–2000 Hz region is shown for better visual clarity.

both the main harmonics and adjacent partials results in envelope amplitudes closer in value to those obtained by vocoding the clean signal.

It is also interesting to note that the spectral envelopes of the stimuli in the Main-3 kHz and PartH-3 kHz conditions were identical (see Fig. 4). Yet, maintaining the spectral envelope of the clean signal alone (during voiced segments) was not found to be sufficient, at least in the context of vocoder processing. A different way of demonstrating this is to consider preserving the linear predictive coding (LPC) spectrum rather than the harmonic spectrum (see example in Fig. 9). The LPC spectrum preserves the spectral envelope of the signal but lacks the harmonic structure (see Fig. 9) present in voiced speech segments such as vowels. As shown in Figs. 7 and 8, preserving the LPC spectrum resulted in poorer envelope reconstruction compared to preserving the harmonic spectrum. In brief, preserving the harmonic structure of the stimulus, and, in particular, preserving the main harmonics along with the adjacent partials, was found to be critically important for vocoder processing. As shown in Fig. 4, the introduction of the partials adjacent to the harmonics provided a better spectral representation of the valleys, and alleviated to some degree spectral distortions introduced by the NR algorithm [see panel (c) vs panel (e)].

Figure 10 shows example plots of the vocoded temporal envelopes of channel 2 (center frequency=324 Hz). It is clear that the noise-reduction algorithm preserved the envelope peaks and deepened the envelope valleys, therefore effectively increasing the envelope dynamic range within each channel. However, as shown in the third panel (from top), the harmonics structure existing in the clean speech voiced segments was severely distorted following the noise-reduction stage. Harmonics-regeneration techniques can partly restore

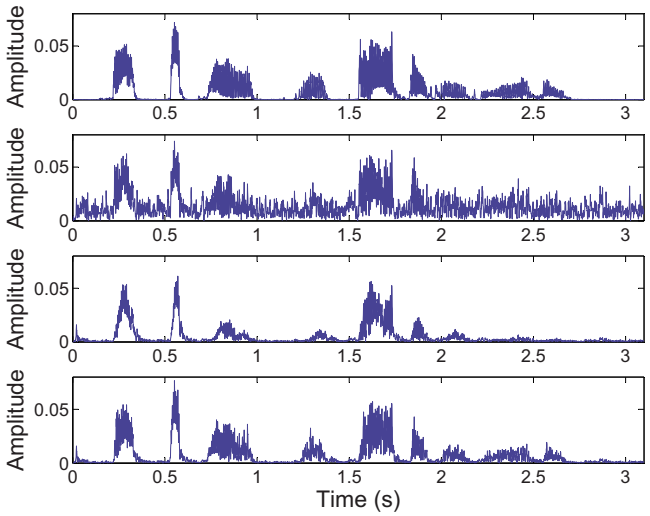


FIG. 10. (Color online) An example plot of the vocoded speech at channel 2 (center frequency=324 Hz). The top panel shows the temporal envelope of the vocoded clean speech. The second panel shows the temporal envelope of the vocoded noisy speech at SNR=0 dB. The third panel shows the temporal envelope of the vocoded speech after noise reduction. The bottom panel shows the temporal envelope of the vocoded speech in the PartH-3 kHz condition which preserved the main harmonics along with the adjacent partials.

the harmonics structure (see bottom panel in Fig. 10) and provide a better envelope representation of the voiced segments.

In summary, regenerating harmonics can improve the spectral representation and temporal envelopes of vocoded speech and can provide substantial intelligibility benefits, relative to NR processing alone.

A. Practical implementation

The present study demonstrated the full potential, in terms of intelligibility improvement, of preserving the signals' harmonic structure prior to vocoder processing. For that, we assumed we had access to the clean harmonic amplitudes and accurate F0 values. In practice, the F0 needs to be estimated from the noisy signal. Accurate F0 detection algorithms exist that can operate at the SNR levels tested in this study. The algorithm in Zavarehei *et al.* (2007), for instance, produced 13% and 7% pitch estimation errors at 5 and 10 dB babble noise, and 11% and 6% errors at 5 and 10 dB train noise. There exist several techniques for estimating, or rather regenerating, the harmonic amplitudes from a noisy signal. Such techniques can potentially be used as a pre-processing step in vocoder processing (see Fig. 1). Harmonics regeneration can be implemented using adaptive comb filtering (Nehorai and Porat, 1986) techniques, nonlinear functions (Plapous *et al.*, 2006), and codebook-based techniques that capitalize on the fact that the harmonic amplitudes are highly correlated (Chu, 2004; Zavarehei *et al.*, 2007). Once the harmonic amplitudes are estimated, it is straightforward to estimate the partials adjacent to the main harmonics using the Gaussian model shown in Fig. 3. The two additional steps (i.e., noise reduction and harmonics generation) will no doubt introduce additional complexity; however, the intelligibility benefits (see Fig. 5) clearly outweigh the additional computational load.

B. Implications for cochlear implants

The results from the present study suggest that the noise-reduction algorithm (Ephraim and Malah, 1985) alone, when used in a pre-processing stage to vocoder processing (see Fig. 1), can bring significant improvements in intelligibility (approximately 10–15 percentage points). This improvement is comparable to that obtained with other pre-processing algorithms (Yang and Fu, 2005) applied to cochlear implants. The implementation of the noise-reduction algorithm used in the present study can be easily integrated with existing speech coding strategies (e.g., ACE strategy) that rely on FFT processing rather than on filterbank processing to derive the channel envelopes. The added complexity is low, as it only involves SNR estimation (see Fig. 2) followed by the multiplication of the noisy FFT magnitude spectrum by a suppression function (which depends on the estimated SNR in each FFT bin). Further, and more substantial, improvements in intelligibility can be realized with the use of the proposed harmonics-regeneration technique, which in turn requires F0 estimation (during voiced segments) and harmonic amplitude estimation. Real-time F0 estimators suitable for CI processing have been demonstrated in Zakis *et al.* (2007); hence the F0 estimation does not pose a problem. The improvement in intelligibility, at least for the type of masker (steady noise) examined in this study, were quite substantial (41 percentage points), making the harmonics-regeneration stage worth incorporating in future speech coding strategies for cochlear implants.

ACKNOWLEDGMENTS

This research was supported by NIH/NIDCD Grant Nos. R03-DC008887 (Y.H.) and R01-DC07527 (P.C.L.). The authors thank Dr. Magdalena Wojtczak and two reviewers for their very helpful comments, and Dr. Felix Chen for his assistance with data collection.

[1]We are referring to the more general form of "glimpsing" occurring at the time-frequency unit level (Cooke, 2006; Li and Loizou, 2007) rather than glimpsing of the wideband signal, as is often assumed in single competing-talker situations. Our explanation is based on a different definition of what constitutes a glimpse: "a time-frequency region which contains a reasonably undistorted "view" of local signal properties" (Cooke, 2006). Glimpses of speech in steady background noise might, for instance, comprise of all time-frequency (T-F) bins or regions having a local SNR exceeding a certain threshold value (e.g., 0 dB). Based on the above definition of glimpsing and the data in Li and Loizou (2007), listeners are able to glimpse the target signal even in steady, continuous, background noise.

[2]The squared error between the clean envelope and processed envelope amplitudes in channel m is computed as follows: $\Sigma_n(x_m(n)-p_m(n))^2$, where $x_m(n)$ and $p_m(n)$ denote the clean and processed envelopes at time n. The squared error was computed across the whole utterance including voiced and unvoiced segments, and averaged across ten sentences.

[3]For each utterance and each channel, Pearson's correlation coefficient was computed between the envelopes of the original stimulus and the processed signals. The correlation estimates were averaged across ten IEEE sentences. A high correlation coefficient indicates a close resemblance between the original and processed envelopes.

Chu, W. C. (**2004**). "Vector quantization of harmonic magnitudes in speech coding applications—A survey and new technique," EURASIP J. Appl. Signal Process. **17**, 2601–2613.

Cooke, M. (**2006**). "A glimpsing model of speech perception in noise," J. Acoust. Soc. Am. **119**, 1562–1573.

Ephraim, Y., and Malah, D. (**1985**). "Speech enhancement using a minimum mean-square error log-spectral amplitude estimator," IEEE Trans. Acoust., Speech, Signal Process. **33**, 443–445.

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

Hamacher, V., Doering, W., Mauer, G., Fleischmann, H., and Hennecke, J. (**1997**). "Evaluation of noise reduction systems for cochlear implant users in different acoustic environments," Am. J. Otol. **18**, S46–S49.

Hochberg, I., Boorthroyd, A., Weiss, M., and Hellman, S. (**1992**). "Effects of noise and noise suppression on speech perception by cochlear implant users," Ear Hear. **13**, 263–271.

Hu, Y., and Loizou, P. (**2007a**). "A comparative intelligibility study of single-microphone noise reduction algorithms," J. Acoust. Soc. Am. **122**, 1777–1786.

Hu, Y., and Loizou, P. (**2007b**). "Subjective comparison and evaluation of speech enhancement algorithms," Speech Commun. **49**, 588–601.

Hu, Y., Loizou, P., Li, N., and Kasturi, K. (**2007**). "Use of a sigmoidal-shaped function for noise attenuation in cochlear implant," J. Acoust. Soc. Am. **122**, EL128–EL134.

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

Kokkinakis, K., and Loizou, P. C. (**2008**). "Using blind source separation techniques to improve speech recognition in bilateral cochlear implant patients," J. Acoust. Soc. Am. **123**, 2379–2390.

Kondoz, A. M. (**1999**). *Digital Speech: Coding for Low Bit Rate Communication Systems* (Wiley, New York).

Li, N., and Loizou, P. C. (**2007**). "Factors influencing glimpsing of speech in noise," J. Acoust. Soc. Am. **122**, 1165–1172.

Li, N., and Loizou, P. C. (**2008**). "The contribution of obstruent consonants

and acoustic landmarks to speech recognition in noise," J. Acoust. Soc. Am. **124**, 3947–3958.
Loizou, P. (**2007**). *Speech Enhancement: Theory and Practice* (CRC, Boca Raton, FL).
Loizou, P., Lobo, A., and Hu, Y. (**2005**). "Subspace algorithms for noise reduction in cochlear implants," J. Acoust. Soc. Am. **118**, 2791–2793.
Mines, M., Hanson, B., and Shoup, J. (**1978**). "Frequency of occurrence of phonemes in conversational English," Lang Speech **21**, 221–241.
Nehorai, A., and Porat, B. (**1986**). "Adaptive comb filtering for harmonic signal enhancement," IEEE Trans. Acoust., Speech, Signal Process. **34**, 1124–1138.
Plapous, C., Marro, C., and Scalart, P. (**2006**). "Improved signal-to-noise ratio estimation for speech enhancement," IEEE Trans. Audio, Speech, Lang. Process. **14**, 2098–2108.
Shannon, R., Zeng, F.-G., Kamath, V., Wygonski, J., and Ekelid, M. (**1995**). "Speech recognition with primarily temporal cues," Science **270**, 303–304.
Sheft, S., Ardoint, M., and Lorenzi, C. (**2008**). "Speech identification based on temporal fine structure cues," J. Acoust. Soc. Am. **124**, 562–575.
Stevens, K. N. (**2002**). "The contribution of obstruent consonants and acoustic landmarks to speech recognition in noise," J. Acoust. Soc. Am. **111**, 1872–1891.
Studebaker, G. A. (**1985**). "A 'rationalized' arcsine transform," J. Speech Hear. Res. **28**, 455–462.
van Hoesel, R., and Clark, G. (**1995**). "Evaluation of a portable two-microphone adaptive beamforming speech processor with cochlear implant patients," J. Acoust. Soc. Am. **97**, 2498–2503.
Weiss, M. (**1993**). "Effects of noise and noise reduction processing on the operation of the Nucleus-22 cochlear implant processor," J. Rehabil. Res. Dev. **30**, 117–128.
Wouters, J., and Berghe, J. V. (**2001**). "Speech recognition in noise for cochlear implantees with a two-microphone monaural adaptive noise reduction system," Ear Hear. **22**, 420–430.
Yang, L., and Fu, Q. (**2005**). "Spectral subtraction-based speech enhancement for cochlear implant patients in background noise," J. Acoust. Soc. Am. **117**, 1001–1003.
Zakis, J. A., McDermott, H. J., and Vandali, A. E. (**2007**). "A fundamental frequency estimator for the real-time processing of musical sounds for cochlear implants," Speech Commun. **49**, 113–122.
Zavarehei, E., Vaseghi, S., and Yan, Q. (**2007**). "Noisy speech enhancement using harmonic-noise model and codebook-based post-processing," IEEE Trans. Audio, Speech, Lang. Process. **15**, 1194–1203.

Unsteady laryngeal airflow simulations of the intra-glottal vortical structures

Mihai Mihaescu[a)]
Department of Aerospace Engineering and Engineering Mechanics, University of Cincinnati, 310 Rhodes Hall, Cincinnati, Ohio 45221-0070

Sid M. Khosla, Shanmugam Murugappan, and Ephraim J. Gutmark[b)]
Department of Otolaryngology-Head and Neck Surgery, University of Cincinnati Medical Center, 231 Albert Sabin Way, Cincinnati, Ohio 4256-0528

(Received 22 October 2008; revised 3 November 2009; accepted 13 November 2009)

The intra-glottal vortical structures developed in a static divergent glottis with continuous flow entering the glottis are characterized. Laryngeal airflow calculations are performed using the Large Eddy Simulation approach. It has been shown that intra-glottal vortices are formed on the divergent wall of the glottis, immediately downstream of the separation point. Even with non-pulsatile flow entering the glottis, the vortices are intermittently shed, producing unsteady flow at the glottal exit. The vortical structures are characterized by significant negative static pressure relative to the ambient pressure. These vortices increase in size and strength as they are convected downstream by the flow due to the entrained air from the supra-glottal region. The negative static pressures associated with the intra-glottal vortical structures suggest that the closing phase during phonation may be accelerated by such vortices. The intra-glottal negative pressures can affect both vocal fold vibration and voice production. © *2010 Acoustical Society of America.* [DOI: 10.1121/1.3271276]

PACS number(s): 43.70.Aj, 43.70.Bk [AL] Pages: 435–444

I. INTRODUCTION

The present research numerically characterizes the unsteady intra-glottal velocity and pressure fields in a three-dimensional, symmetric larynx model of the true vocal folds during the later part of closing. During this phase of phonation, the glottis is divergent. A static model is used here so that the effects of vocal fold movement on the flow are not considered. The primary hypothesis of this work is that intra-glottal vortices form in the superior aspect of the divergent glottis, and that these vortices produce significant negative aerodynamic pressures, relative to both ambient and sub-glottal pressures. These intra-glottal vortices form in a divergent duct due to flow separation and thus will be referred to as flow separation vortices (FSV). The formulated hypothesis is strongly supported by findings in the excised canine larynx model (Khosla *et al.*, 2007, 2008a, 2008b; Murugappan *et al.*, 2009), assuming that the pressures at the glottal exit are similar to the pressures just below the glottal exit, at the same phase. One of the goals of this paper was to test this assumption in a static model since there have been no analytical or computational models that explicitly studied the FSV. In order to test this hypothesis, this study focused on the analysis of the vortical structures developed on the divergent section of the intra-glottal region of the laryngeal airflow, the generation mechanism, and the associated pressures as the vortices travel through and above the glottis.

The hypothesis stated above is clinically important since the findings in the previously mentioned animal models also suggest that these negative pressures play an important role in determining the closing speed of the vocal fold. Gorham-Rowan and Morris (2006) suggested that the vocal fold closing speed is related to the maximum flow declination rate (MFDR). Understanding the mechanisms contributing to the MFDR is clinically important, since the MFDR is highly correlated with vocal intensity (Sundberg and Gauffin, 1979; Holmberg *et al.*, 1988; Gauffin and Sundberg, 1989; Sapienza and Stathopoulos, 1994). Stevens (1998) noted that the rapid reduction in flow is also important for generating acoustic energy over a broad frequency range and showed analytically that increasing the rate of flow shutoff will produce increased energy in the higher harmonics, a theory supported by multiple findings in patients (Klatt and Klatt, 1990; Hanson, 1997; Gobl, 1989).

Fant (1982) showed analytically that a higher MFDR can be obtained by increasing the inertance of the vocal tract, by increasing vocal fold closing speed, or by increasing the maximum lateral vocal fold displacement (MLD). Increasing MLD will increase the elastic recoil forces, which will accelerate vocal fold closing speed and MFDR, and thus augment intensity. However, this mechanism is not consistent with the one study in humans that simultaneously looked at the relationships between intensity and both vocal fold closing speed and MLD. By using image-processing techniques to analyze stroboscopic images in patients under different phonatory conditions, Woo (1996) showed that vocal intensity was highly correlated with maximum vocal fold closing speed but, in contradiction with Fant's theories, not with maximum lateral displacement.

[a)]Author to whom correspondence should be addressed. Electronic mail: mihai.mihaescu@uc.edu
[b)]Also at Department of Aerospace Engineering and Engineering Mechanics, University of Cincinnati, 799 Rhodes Hall, Cincinnati, Ohio 45221-0070.

Woo's (1996) findings are consistent with results in excised canine larynx models (Khosla *et al.*, 2009; Murugappan *et al.*, 2009), which suggest that, under certain conditions, the contribution of the intra-glottal vortices to determining rapid vocal fold closing speed is much higher than the contribution of maximum lateral displacement. Their results also indicate that the FSV are associated with significant intra-glottal negative pressures during closing. It is interesting to note that an earlier study in an excised canine hemilarynx (Alipour and Scherer, 2000) showed negative pressures produced during the later part of closing, when the glottis is divergent. In this case the authors attributed the negative pressures occurring at the upstream glottal sections to Bernoulli effects, flow acceleration, and separation. The negative pressures occurring during closing in the upper portion of the glottis (downstream) were attributed to the curvature of the vocal fold and to rarefaction produced by closing of the folds. These findings in animal models contradict one of the basic assumptions used in many phonatory models: that the pressures downstream of the point of separation are uniform and equal to the ambient pressure (Pelorson *et al.*, 1994; Story and Titze, 1995; Lous *et al.*, 1998). As Krane and Wei (2006) noted, this assumption may not be correct if the glottal flow is significantly unsteady. Therefore, an accurate characterization of the unsteady vortical flow in the intra-glottal region and its interaction with the vocal folds is important.

The use of Computational Fluid Dynamics (CFD) in investigating the non-linear transitional/turbulent glottal airflow has been facilitated by advances in computer technology over the past few decades. Using unsteady CFD solvers, two-dimensional, axisymmetric, and three-dimensional glottal configurations have been analyzed with and without vocal fold motion (Zhao *et al.*, 2001, 2002; Zhang *et al.*, 2002; Hofmans *et al.*, 2003; Alipour and Scherer, 2004; Decker and Thomson, 2007; Mihaescu *et al.*, 2007; Suh and Frankel, 2007). In many of these numerical studies, the sound generated by the laryngeal airflow was considered (Zhao *et al.*, 2001, 2002; Zhang *et al.*, 2002; Mihaescu *et al.*, 2007; Suh and Frankel, 2007). Parametric studies involving the influence of Reynolds number on the flow characteristics and the effect of different divergent glottal angles on the flow separation point have also been performed (Zhang *et al.*, 2002; Alipour and Scherer, 2004; Mihaescu *et al.*, 2007). It is known that for a diverging shape of the vocal folds, when the angle of the glottis exceeds a certain minimal value, the flow separates and vortices are formed immediately downstream of the separation point (Zhao *et al.*, 2001, 2002; Zhang *et al.*, 2002). Although it was recognized that the unsteady vortex shedding results in unsteady forces on the surface of the vocal folds (Zhao *et al.*, 2001, 2002), the features of these vortices have not been investigated in detail. None of these models suggests that the FSV produced significant negative aerodynamic pressures and none of them show the pressures associated with these vortical structures inside the glottis.

There are a few computational larynx models that showed negative wall pressures in the superior aspect of the glottis during closing in addition to the low wall pressures occurring due to Bernoulli effects at the minimum glottal width. However, none of them showed how the pressure field within the glottis changes with the intra-glottal flow evolution. Using Direct Numerical Simulation (DNS), Zhang *et al.* (2002) studied the aeroacoustics of the glottal flow for different phonatory conditions and glottal configurations. During the closing phase, in some of the cases, negative wall pressures and vortices were shown at the superior edge of the glottis exit. These vortices occurred near the glottal exit, therefore having a minimal effect on vocal fold closing speed and MFDR. Zhao *et al.* (2002a) stated that the intra-glottal flow separation vortex "triggers a Kelvin–Helmholtz-type instability of the jet and secondary vortices are formed near the outlet of the glottis."

Using Large Eddy Simulation (LES), Suh and Frankel (2007) showed that the intra-glottal flow attaches to one wall of the divergent duct, which they refer to as the "flow wall." Along that wall, they showed negative intra-glottal pressures in the superior aspect of the glottis. The magnitude of these negative pressures was consistent with the previously mentioned findings in excised canine larynges. However, they described vortices only along the separated flow shear layer, opposite the wall to which the flow was attached. The negative wall pressure was thus related to the adjacent flow rather than to the vortices in the detached shear layer. Therefore, the attached flow described in Suh and Frankel, 2007 does not allow testing of the latter part of our hypothesis—that the flow separation vortices produce significant negative aerodynamic pressures on the wall adjacent to them.

Employing the LES methodology, the present study presents a detailed analysis of the intra-glottal velocity fields, vortical structures, and intra-glottal pressures in a static duct. This analysis will allow us to determine if the flow separation vortices generate significant negative intra-glottal pressures in the superior aspect of the glottis, and thus if the closing phase during phonation may be accelerated by such vortices.

II. METHODS

Computational fluid dynamics methods predict flow characteristics by solving the flow governing Navier–Stokes equations. Within the CFD framework, the steady Reynolds averaged Navier–Stokes (RANS) technique utilizing appropriate two-equation turbulence models (e.g., k-ε or k-ω) cannot accurately describe unsteady flow fields involving flow separation and adverse pressure gradients (Wilcox, 1993). With such models, only information about the local mean flow is computed, and all the flow unsteadiness is filtered out. In flow situations that involve adverse pressure gradients and separation, unsteady RANS based on shear-stress transport k-ω model is superior to other RANS models since it offers a better trade-off between accuracy and efficiency (Suh and Frankel, 2008). In the past few decades, LES has become the major computational tool for studying unsteady turbulent flows (Pope, 2000). LES resolves a large range of turbulent scales (i.e., the energy containing eddies) and thus is able to capture the turbulence dynamics. Only the smallest scales in the turbulent flow need to be approximated by modeling.

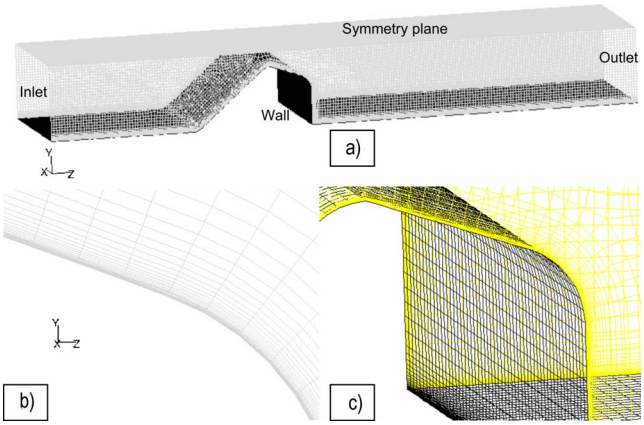


FIG. 1. (Color online) The three-dimensional symmetric larynx model: (a) view of the computational domain showing the inlet and the outlet planes; (b) details of the grid resolution in the y-z mid-plane (mid-coronal plane of the model); the boundary layer region on the 20° divergent side of the glottis is depicted; (c) details of the three-dimensional computational grid in the glottis region.

In the present research, unsteady LES (Fluent Inc.®) is employed to compute the laryngeal airflow through a three-dimensional static and diverging larynx model that assumes symmetry relative to the mid-sagittal plane of the larynx. It is known in both mechanical and computational three-dimensional larynx models that the glottal jet is often asymmetric, attaching to one side or the other of the glottis due to the bistable character of the flow. The phenomenon has been observed and has been analyzed in several studies (Shinwari *et al.*, 2003; Neubauer *et al.*, 2007; Mihaescu *et al.* 2007; Nomura and Funada, 2007; Suh and Frankel, 2007, 2008). On the other hand, experimental studies performed in excised canine larynges using particle imaging velocimetry (Khosla *et al.*, 2008b; Murugappan *et al.*, 2009) showed that flow separation vortices form on both sides of the glottis if the mucosal waves are relatively symmetric. However, if the mucosal waves are significantly asymmetric, the flow does separate on only one side (Khosla *et al.*, 2009). These findings are consistent with results in a self-oscillating physical model with asymmetric vibration reported by Neubauer *et al.* (2007). Since we are interested in the normal situation, where the mucosal waves are relatively symmetric, the present LES study assumes symmetry at the mid-sagittal plane to avoid skewing of the glottal jet relative to the axial direction.

The glottal model is depicted in Fig. 1(a). The vocal folds shape is characterized by a 20° divergent angle corresponding to the closing phase of the phonation cycle. The computational domain consists of a rectangular cross-section of 15.24 mm in the sagittal direction (x), 7.62 mm in the y-direction in the coronal plane (perpendicular to the sagittal plane and parallel to the main flow direction), and 60 mm in the stream-wise direction (z). The minimum distance between the vocal fold and the mid-sagittal plane of the larynx is 1.5 mm resulting in a full glottal width (D_g) of 3 mm. Grid parameters are important for efficient and accurate flow simulation. An unstructured hexahedral body fitted mesh with roughly 7×10^5 computational volumes is used to discretize the computational domain. The Taylor micro-scale (λ_T), which is a measure for the computational cell size, is an important parameter for LES computations. Assuming that the integral length scale (l) is one order of magnitude smaller than the geometrical characteristic length scale defined as the glottal width ($l=0.1D_g$) and a turbulence intensity of 10%, the Taylor micro-scale is (Pope, 2000):

$$\lambda_T=[15\nu_{\text{air}}0.1D_g/(0.1W_{\max})]^{1/2}.$$

Using this definition and considering an air kinematic viscosity of $\nu_{\text{air}}\sim1.79e-5$ m^2/s, a glottal width of $D_g\sim3$ mm, and a maximum axial velocity of $W_{\max}\sim21$ m/s, the Taylor micro-scale is estimated to be $\lambda_T\sim0.2$ mm. Accordingly, the maximum interval size used to discretize the volume of the glottal model is of the same order as the Taylor micro-scale. Detailed LES flow calculations in the near-wall region (i.e., near the vocal folds) are possible due to the high grid resolution used in this region, the computational grid is smoothly stretched in the wall normal direction. Figures 1(b) and 1(c) present details of the computational grid used at the vicinity of the vocal folds. The discretization of the flow governing equations on the computational domain is performed using second order finite-volume schemes. An implicit second order discretization scheme is used for the time integration. The coupling between the velocity field and the pressure field is realized using the semi-implicit method for pressure-linked equation (SIMPLE) algorithm (Van Doormaal and Raithby, 1984). In the present computations, the Smagorinsky–Lilly sub-grid scale model is employed (Lilly, 1966).

The simulations were conducted for a flow rate of 27.5 l/min. The correspondent trans-glottal pressure was approximately 295 Pa (3 cm H_2O). The low flow Mach number (Mach<0.3) allowed using the incompressible flow formulation. A constant axial velocity profile was used as inlet boundary condition (sub-glottal) to match the desired volumetric flux, and 10% turbulence intensity was specified at the inlet surface. The Reynolds number at the minimum glottal width, based on the velocity computed from the bulk flow rate, the glottal width, and the kinematic viscosity of the air was roughly 3500. Symmetry boundary conditions were used at the mid-sagittal plane of the larynx (i.e., upper boundary of the computational domain) and no-slip boundary conditions for velocity were set at the solid surfaces (i.e., side and bottom boundaries of the computational domain). For the outlet plane, a flux conserving zero gradient boundary condition was applied. To reduce the computational time required for the LES unsteady calculations, the flow field was initialized with a previously obtained converged solution based on steady RANS formulation with a k-ε turbulence model.

The time-average flow quantities presented in this study were statistically averaged over a period of 10 000 time-steps (roughly 30 flow-through times), achieving a converged solution at each time-step. The time-step used with the present LES calculations was $\Delta t=0.000\,02$ s. The size of each of the scales developed in the flow field that can be solved by LES is highly dependent on the chosen grid resolution and can be quantified by the energy spectrum. A post-priori spectral analysis was performed in order to check if the grid

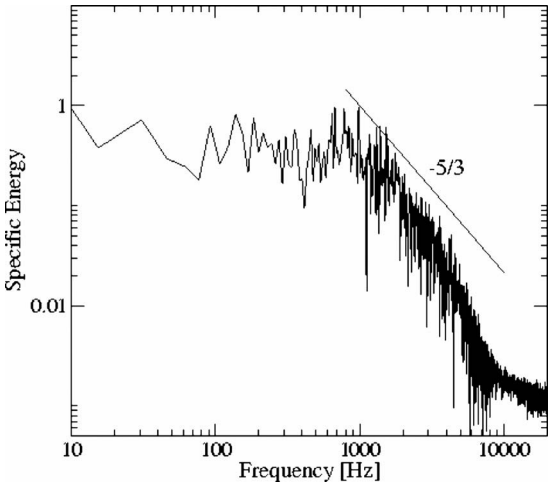


FIG. 2. Turbulent kinetic energy spectrum calculated at a point situated in the supra-glottal region of the larynx. The spectrum follows the −5/3 Kolmogorov power law very well, the spatial resolution of the computational grid being fine enough to resolve a large number of scales within the inertial subrange.

resolution was adequate for solving a large range of scales (i.e., vortical structures) within the inertial subrange (Pope, 2000). The turbulent kinetic energy spectrum in Fig. 2 shows that the energy decay followed the −5/3 Kolmogorov power law, so that the resolution used in the simulations was fine enough to resolve about one order of magnitude of the inertial subrange, a condition required by LES.

III. RESULTS

The time-averaged axial velocity distributions, as predicted by LES, in the mid-coronal plane (y-z mid-plane) of the computational domain are shown in Figs. 3 and 4. The mean flow solution [Figs. 3 and 4(a)] describes the acceleration of the flow due to the convergent geometry of the vocal folds. A maximum velocity was reached at the site of the minimum distance between the glottis and the symmetry plane (i.e., upper boundary of the computational domain). Further downstream, the flow decelerated due to the sudden expansion of the geometry in the lateral direction. The resulting adverse pressure gradient separated the flow from the 20° divergent side of the glottis. A recirculation flow region developed downstream of the separation point that was initiated at the minimum glottal width. Negative axial velocities very close to the wall were associated with this reversed flow region, as shown in Fig. 4(b). The time-averaged axial velocity profiles were extracted along lines situated in the mid-coronal plane of the computational model, at different locations in the flow direction, downstream of the minimum glottal width region as presented in Fig. 4(a). In the intra-glottal region, additional energy that sustains this recirculation bubble was provided by the entrained air from the supra-glottal region.

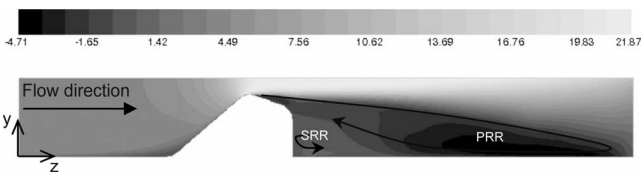


FIG. 3. Time-averaged axial velocity distribution (m/s) as calculated by LES in the mid-coronal plane of the larynx model. Flow separation occurs at the diverging side of the glottis and recirculation regions identified by negative axial velocity values are developed. The maximum velocity occurs at the site of the minimum glottal width (21.87 m/s), while the minimum velocity in the recirculation region is −4.71 m/s. Note: PRR—primary recirculation region; SRR—secondary recirculation region.

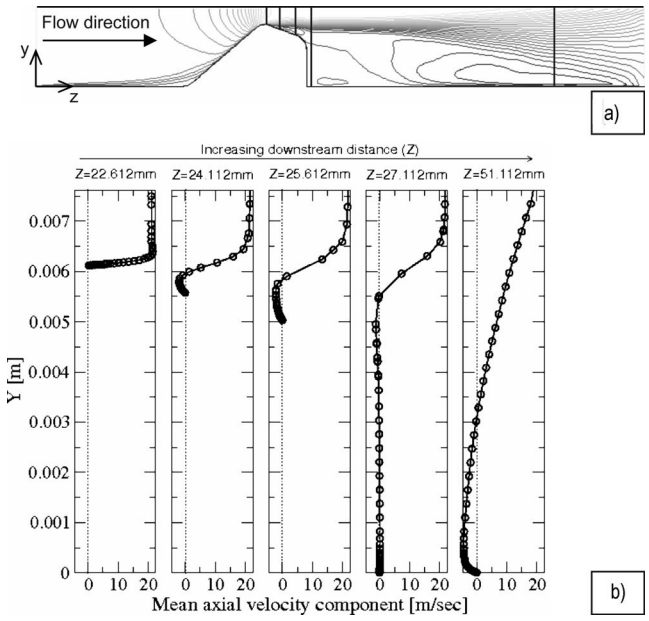


FIG. 4. Time-averaged axial velocity distribution (m/s) as calculated by LES in the mid-coronal plane: (a) location of the lines along which the data was extracted; (b) plots of the time-averaged axial velocity profiles extracted along the lines situated at different locations in the flow direction (z).

The laryngeal airflow was unsteady even with stationary vocal folds. Snapshots of static pressure distributions and velocity vector at different time instances ($t=3.8\times10^{-4}$, $t=5.0\times10^{-4}$, and $t=6.20\times10^{-4}$ s) during the LES calculations are presented in Figs. 5(a)–5(c). These close-ups of the *instantaneous* flow through the glottis at different time stages showed that intra-glottal vortical structures were generated between the divergent slope of the vocal folds and the separated center jet, just downstream of the separation point. The core of these vortical structures was characterized by significant negative static pressure values relative to the surrounding static pressure field. The evolution and convection of these structures resulted in a non-uniform and unsteady static pressure distribution on the vocal folds divergent surface. The convective velocity of these vortices was estimated to be approximately 10 m/s.

The surface static pressure on the glottal wall was affected by the passing of the vortex. Figure 6(a) presents the time history plots of instantaneous static pressure monitored in three different locations (P_1, P_2, and P_3) near the vocal folds surface, along the divergent glottal slope in the mid-coronal plane of the model. Details of the instantaneous isocontours of static pressure spatial distributions at the three time instances ($t=2.0\times10^{-4}$, $t=3.6\times10^{-4}$, and $t=5.0\times10^{-4}$ s) marked on Fig. 6(a), and the location of the three monitoring points (P_1, P_2, and P_3) are depicted in Fig. 6(b). Near the vocal fold, as the vortical structure is moving from

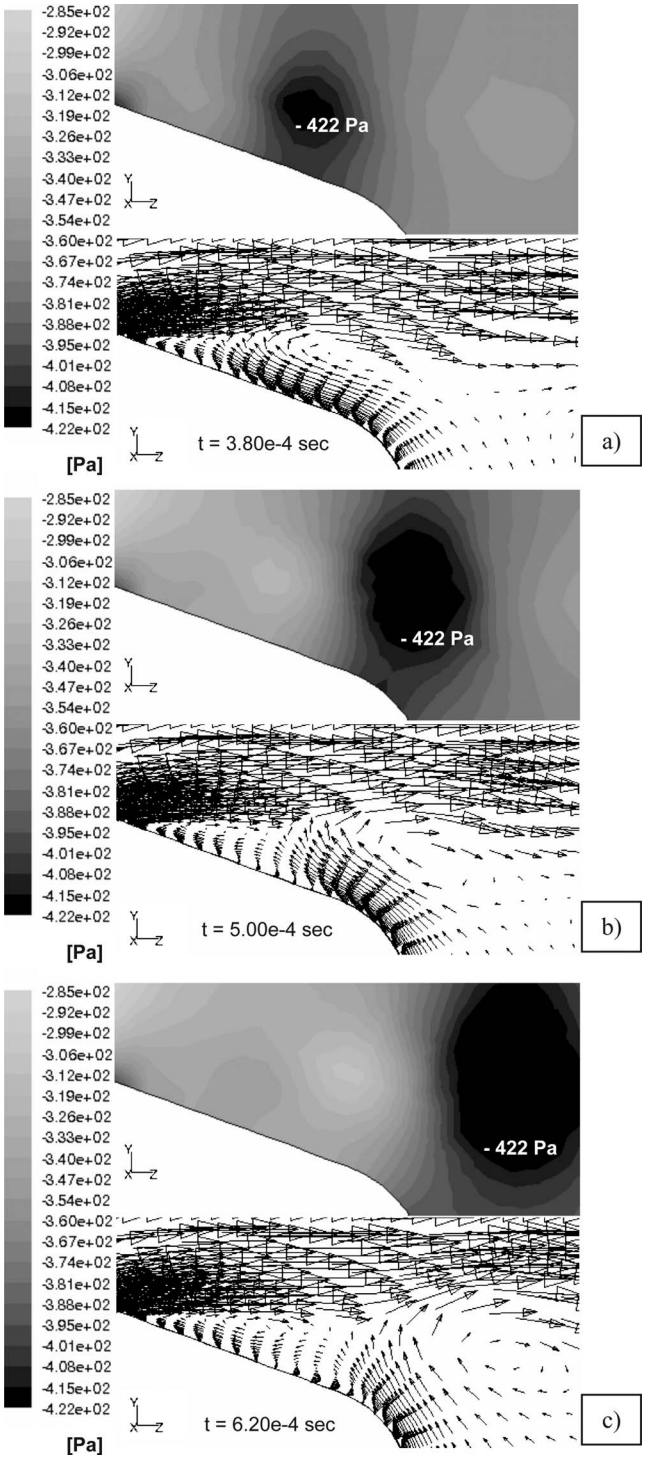


FIG. 5. Flow evolution in the intra-glottal region at three time instances as calculated by LES: (a) 3.8×10^{-4} s, (b) 5.0×10^{-4} s, and (c) 6.2×10^{-4} s. The instantaneous static pressure in Pa (first row) and the corresponding vortical structures represented by velocity vector (second row) are depicted. Note that the legend scale is in Pa: 1 Pa=0.0102 cm H_2O; −422 Pa= −4.30 cm H_2O.

location P_1 to P_3, the static pressure values are decreasing from roughly −320 Pa at P_1 ($t=2.0\times10^{-4}$ s) to −400 Pa at P_3 ($t=5.0\times10^{-4}$ s).

The intra-glottal pressure field was strongly affected by the presence of an intra-glottal vortical structure. Figure 7 presents a comparison in terms of the instantaneous static pressure distributions (pressure drop from sub-glottal pressure) along the centerline of the glottal jet, along a line located in the shear-layer, and along the glottal wall [see Fig. 6(b) for the location of the monitoring lines]. The data are presented for the time instance of $t=3.80\times10^{-4}$ s, when a vortex is formed and is present in the glottis, as shown in Fig. 5(a). As the flow accelerates when it enters the glottis, the pressure decreases. The data extracted along the wall and along the developed shear-layer showed similar profiles, with the pressure values reaching a first minimum at the location of the minimum cross-sectional area (i.e., nearby the glottal entrance) and a second local minimum due to the presence of the intra-glottal vortex. Another observation was that the pressure values on the centerline of the glottal jet have only one minimum that corresponds to the location of the intra-glottal vortex. It should be also noted that the low pressure

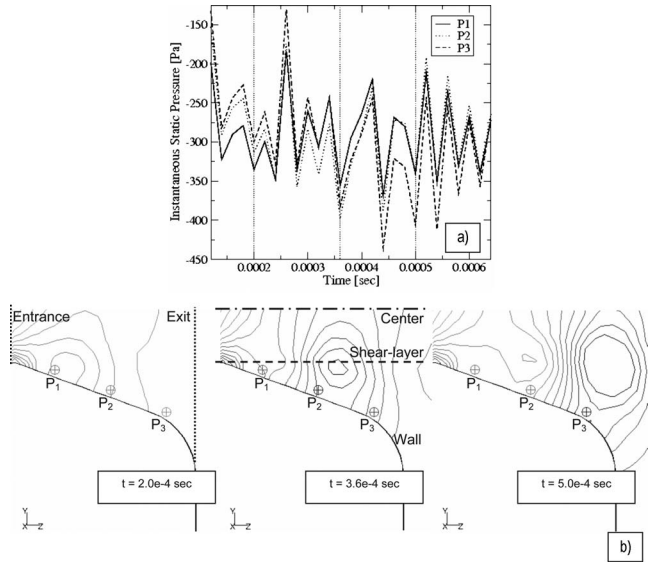


FIG. 6. Intra-glottal static pressure evolution (Pa) in time and space: (a) time history plots of instantaneous static pressure monitored in three different points; (b) locations of the monitoring lines (center, shear-layer, wall) and points (P_1, P_2, and P_3) and the static pressure distributions at three time instances ($t=2.0\times10^{-4}$ s, $t=3.6\times10^{-4}$ s, and $t=5.0\times10^{-4}$ s) indicated in Fig. 6(a).

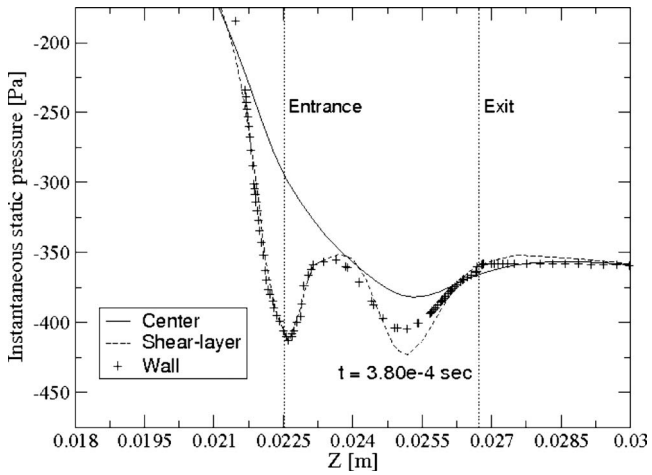


FIG. 7. Instantaneous intra-glottal static pressure distributions (Pa) at the time instance of $t=3.8\times10^{-4}$ s along the centerline of the glottal jet, along a line located in the shear-layer, and along the glottal wall as shown in Fig. 6(b). Note that the glottal entrance and exit are marked on the figure.

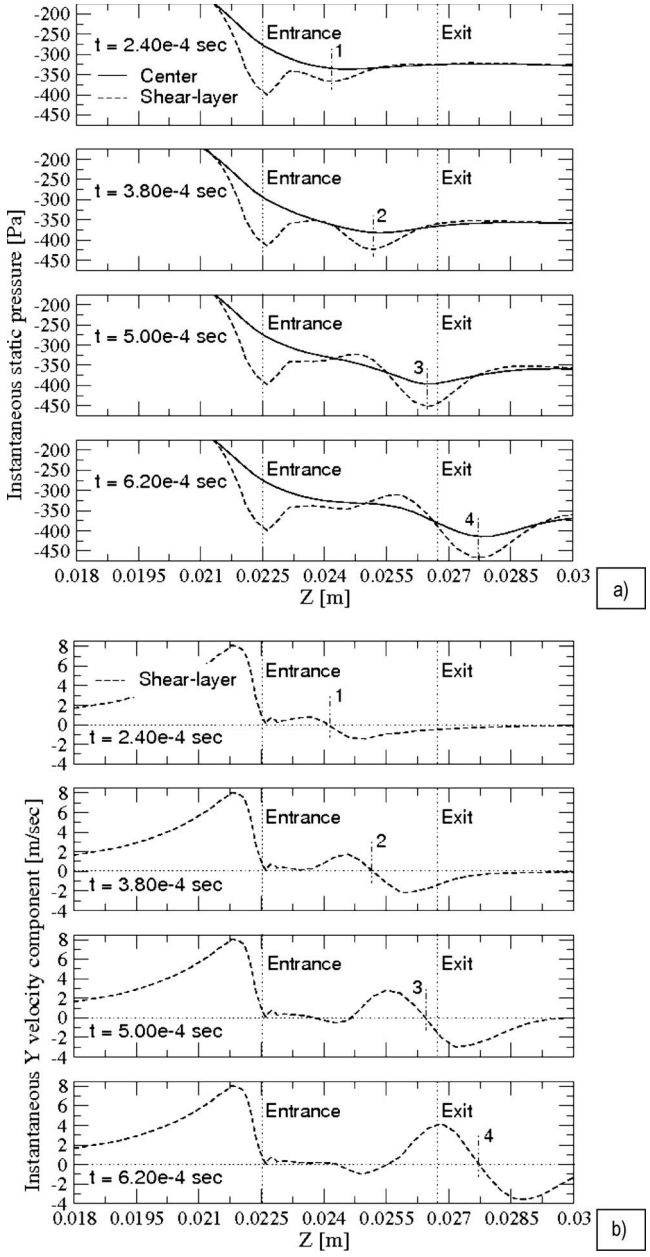


FIG. 8. Time sequences of pressure and Y velocity component profiles associated with intra-glottal vortex evolution: (a) static pressure distributions (Pascal) at four time instances ($t=2.4\times10^{-4}$ s, $t=3.8\times10^{-4}$ s, $t=5.0\times10^{-4}$ s, and $t=6.2\times10^{-4}$ s) along the glottal jet centerline and along the developed shear-layer. (b) Y velocity (m/s) plots at four time instances ($t=2.4\times10^{-4}$ s, $t=3.8\times10^{-4}$ s, $t=5.0\times10^{-4}$ s, and $t=6.2\times10^{-4}$ s) along the developed shear-layer. Note that with 1, 2, 3, and 4 marked on the figure, the locations of the vortex at each of the four time instances are analyzed and Y velocity is the velocity component normal to the flow direction.

values induced by the intra-glottal vortex were lower than the wall pressure values developed at the minimum glottal width.

As the vortex was convected downstream by the flow, it gained strength and increased in size. This resulted in an increase in magnitude of the negative static pressure that characterizes the vortex core. This is presented in Figs. 8(a) and 8(b) using time sequences of static pressure and of Y velocity component plots along the developed shear-layer in the mid-coronal plane (y-z mid-plane). Note that the pressure in Fig. 8(a) is plotted along the glottal jet centerline and the

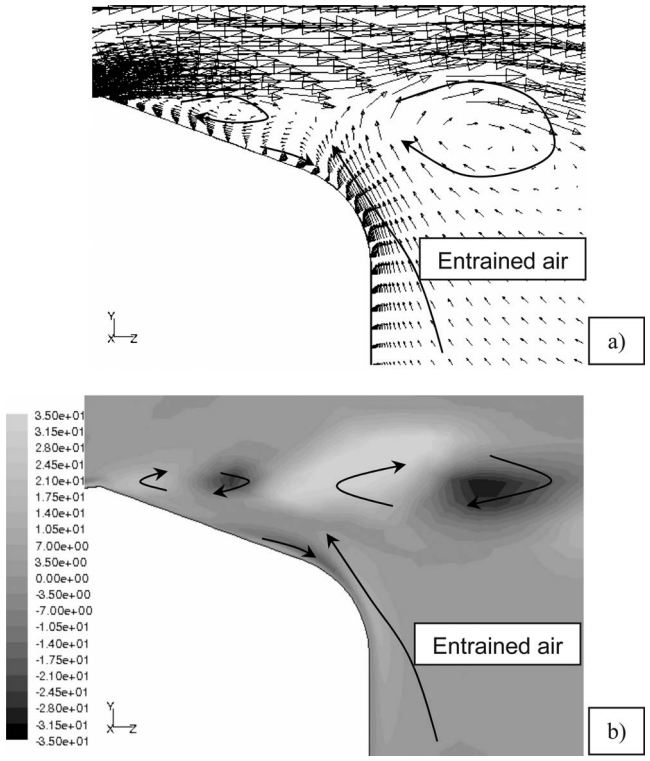


FIG. 9. Instantaneous flow characteristics in the intra-glottal region at the time instance of $t=6.4\times10^{-04}$ s: (a) instantaneous velocity vectors (m/s); (b) the corresponding contours of helicity (m/s^2).

Y velocity is the velocity component in the coronal plane normal to the stream-wise flow direction (z). The four time instances chosen are of $t=2.40\times10^{-4}$, 3.80×10^{-4}, 5.0×10^{-4}, and 6.20×10^{-4} s and relevant data for the last three time successions were presented also in Figs. 5(a)–5(c). While the vortex traveled from position 1 (time instance of $t=2.40\times10^{-4}$ s) to position 4 (time instance of $t=6.20\times10^{-4}$ s), the pressure in the vortex core decreased from roughly −370 to −470 Pa [Fig. 8(a)], whereas the maximum absolute value of the Y velocity component, extracted along the shear-layer that passes through the vortex, increased four times from roughly ±1 to ±4 m/s [Figs. 8(b)].

The strengthening of the intra-glottal vortices as they traveled downstream through the divergent section of the glottis was attributed to the entrained air from the supra-glottal region. This is demonstrated in Fig. 9 that shows a close-up of the instantaneous velocity vectors and the corresponding contours of helicity (m/s^2) at a time instance of $t=6.4\times10^{-4}$ s. Helicity is a scalar quantity defined as the dot product of the velocity and vorticity vectors and it is used as a method to visualize vortical structures since its sign characterizes locally the direction of the swirl with respect to the main flow direction (Degani *et al.*, 1990). A positive sign for helicity means that the rotation is in the same direction as the main flow, and negative means that the rotation is in the opposite direction. The helicity calculated near the glottal wall indicated that there was a strong entrainment of air from the supra-glottal region that interacted with the wall (vorticity generated near by the wall due to the viscous effects) and dominated the near-wall region at the exit of the glottis.

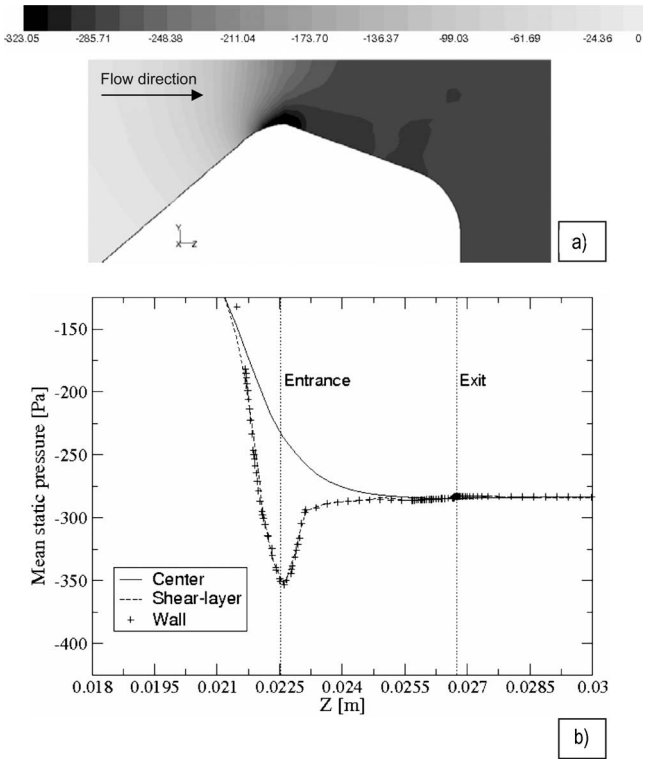


FIG. 10. Time-averaged intra-glottal static pressure distribution (Pa) as calculated by LES in the mid-coronal plane: (a) detail of the static pressure distribution in the mid-coronal plane; (b) comparisons of the mean wall static pressure values with the pressure predicted along the centerline and shear-layer of the glottal jet.

The *time-averaged* static pressure field calculated by LES in the mid-coronal plane of the model is presented in Fig. 10(a). Averaging the pressure data over 10 000 time-steps filtered out the intra-glottal pressure variations due to the presence of the vortices as it was earlier shown in the instantaneous data (Fig. 5). The calculated mean static pressure decreased on the convergent wall of the glottis as the flow accelerated due to glottal channel narrowing [Figs. 10(a) and 10(b)]. The minimum static pressure, on the surface of the vocal folds, was found at the location of minimum cross-sectional glottal area. This was due to the highest flow velocity at the minimal glottal area associated with the effect of the glottis curvature on the pressure field, as explained by Gauffin and Liljencrants (1988). Following the flow direction along the glottal wall, downstream of the minimum cross-sectional area, a slight pressure recovery was found. However, the intra-glottal wall pressure was still lower as compared with the centerline pressure, as shown in Fig. 10(b), in particular, for the region located in the first half of the divergent glottis where the separation induced vortices are produced. The higher pressure found in the center of the intra-glottal flow (higher than the calculated pressure on the wall) is also a condition for the flow to be able to follow the expansion of the divergent glottis (Gauffin and Liljencrants, 1988).

The characteristic frequencies of the intra-glottal and supra-glottal vortical structures were analyzed using velocity spectra. The spectra were obtained by recording the axial velocity time-history at several monitoring points and performing discrete Fourier transforms. Figures 11(b) and 11(c) show the frequency spectra obtained at two locations in the shear layer as marked in Fig. 11(a). The dominant frequencies in the intra-glottal region [Fig. 11(b)] were 92 Hz (St $\sim$0.013) and 153 Hz (St$\sim$0.022), while further downstream, in the supra-glottal region, the dominant frequencies were around 1000 Hz (St$\sim$0.143) [Fig. 11(c)]. The Strouhal number (St=fD_g/W) was based on the minimum glottal width (D_g) and the mean velocity of the glottal jet at this location ($W\sim 21$ m/s). The 92 and 153 Hz frequencies were detected also in the supra-glottal region of the shear-layer, as a result of the convection of the intra-glottal vortices downstream by the laryngeal jet.

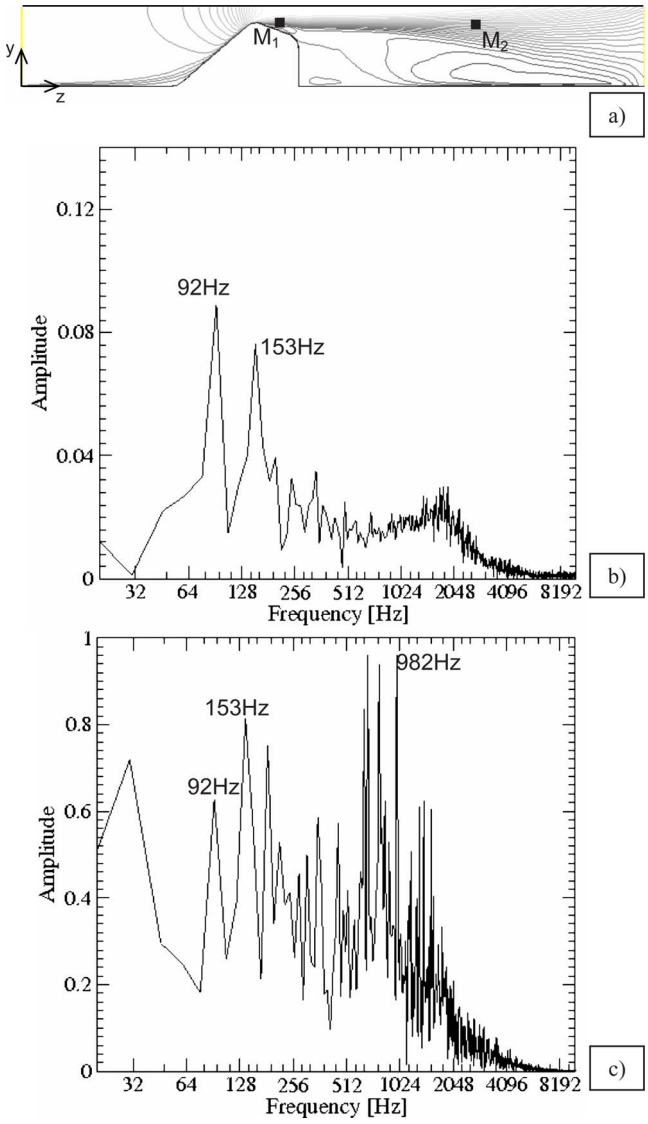


FIG. 11. (Color online) Spectral analysis of the vortical structures: (a) the locations of the monitoring points (M_1 and M_2) in the glottal shear-layer; (b) intra-glottal dominant frequencies at M_1; (c) supra-glottal dominant frequencies *at* M_2.

IV. DISCUSSION AND CONCLUSIONS

Experimental and unsteady numerical studies performed in the last few decades suggest that vortical structures develop in the laryngeal airflow (Kucinschi *et al.*, 2006; Khosla

et al., 2007; Zhao *et al.*, 2001; Triep *et al.*, 2005; Mihaescu *et al.*, 2007; Neubauer *et al.*, 2007). These structures are formed in the supra-glottal region due to the shear-layer instabilities of the laryngeal jet. Additional vortices are formed in the intra-glottal region during the closing phase of the phonation cycle due to flow separation in the divergent glottal duct (Zhao *et al.*, 2001, 2002; Zhang *et al.*, 2002; Khosla *et al.*, 2007). The fact that the unsteady vortex shedding results in unsteady forces on the surface of the vocal folds is documented (Zhao *et al.*, 2001, 2002; Zhang *et al.*, 2002), but the characteristics of these vortices and their effect on the intra-glottal pressures have not been addressed in detail. This is important since the intra-glottal pressures can affect both vocal fold vibration and voice production.

The present research deals with laryngeal airflow simulations based on LES approach in a symmetric rigid larynx model. The glottis has a divergent shape with an angle of 20°. The LES results captured the flow separation that occurs at the divergent slope of the vocal folds and the intra-glottal vortices that develop immediately downstream of the separation point. These flow structures induced significant negative static pressures relative to both the ambient pressure field and the maximum subglottal pressure. The vortices were also convected by flow, while increasing in strength, near the divergent section of the glottal wall.

Numerical predictions of the *instantaneous* static pressure distribution on the glottal wall as a function of stream-wise distance were previously reported (Zhang *et al.*, 2002). The calculations were performed using DNS of the laryngeal airflow. The motion of the vocal folds in several idealized axisymmetric models of the human vocal tract was studied with or without the presence of the ventricular folds. It was shown that during the closing phase of the phonation cycle (i.e., divergent glottal shape), lower pressures may appear on the divergent slope of the glottis downstream of the expected low at the minimum glottal width. The negative pressures at the inferior aspect were predominantly due to Bernoulli effects, since the glottis is the narrowest at this point. The negative pressures in the wider superior aspect of the glottis were consistent with the hypothesis that the vortices are characterized by negative pressures. However, no correlations were made between these low pressure values and the vortical structures passing near the glottal wall.

The time-averaged intra-glottal pressure distribution predicted by the present LES calculation was shown to be influenced by the low-negative static pressure values induced by the flow structures developed due to the separation of the flow. Thus, in the first half of the divergent wall of the folds, the pressure was lower as compared to its values toward the symmetry plane of the model. The minimum pressure value found on the glottal wall, at the site of the narrowest glottal width, was followed by a slight pressure recovery in the stream-wise direction. LES data concerning the intra-glottal pressure distribution in a rigid convergent-divergent glottal model that considered both left and right vocal folds were recently reported (Suh and Frankel, 2007). It was shown that the glottal jet skews and attaches to one wall, a severe flow separation occurring on the other/opposite wall of the glottis just downstream of the minimum glottal area. The pressure data, on the side where the flow separates, were found to be slightly higher than the pressures on the opposite wall, where flow attaches. As in the present simulations, the minimum pressure was at the location of the narrowest glottal width on the side where separation occurs. Downstream of the separation point, their results indicated a uniform pressure distribution on the vocal fold; the interaction between the glottal jet and the vocal fold wall through eventually generated structures in the shear-layer not being intense due to the severe separation. In the present LES, due to the assumed flow symmetry, the glottal jet skewing is avoided and there is an important interaction between the intra-glottal structures and the divergent wall of the glottis.

The spectral analysis of the flow data revealed that the vortices developed in the intra-glottal region are characterized by lower frequencies ($St \sim 0.02$) as compared with those developed in the supra-glottal region ($St \sim 0.143$). This suggests that the intra-glottal flow structures are induced at lower frequency as compared with the vortices generated in the supra-glottal region in the glottal jet shear-layer. Experimental studies concerning the flow structures generated downstream of divergent larynx models, and their characteristic frequencies, indicated that the Strouhal number of vortex shedding is approximately 0.145 for glottal flows with the Reynolds number of roughly 2000 (Kucinschi *et al.*, 2006; Zhang *et al.*, 2004). This is in good agreement with our findings ($St \sim 0.143$). The same experiments shown that for glottal flows at different Reynolds numbers (roughly range of $1500 < Re < 3300$), the vortex shedding frequency corresponds to a relatively constant value of the Strouhal number. This behavior was not observed for the low Reynolds number cases ($Re < 1500$).

As shown in both experimental and numerical studies, vortical structures affect the voice production through contributions as monopole, dipole, and quadrupole acoustic sources (McGowan, 1988; Barney *et al.*, 1999; Titze, 2000; Zhao *et al.*, 2001; Zhang *et al.*, 2004; Suh and Frankel, 2007). The monopole acoustic source that is associated with fluid displacement due to acceleration of a moving surface was found to dominate the other acoustic sources at very high vocal fold vibration frequencies (Zhao *et al.*, 2002), while the dipole acoustic source generated by the pressure distribution (fluctuating loading) on a surface was found to dominate for convergent-divergent glottal models (Zhao *et al.*, 2001, 2002; Zhang *et al.*, 2004; Suh and Frankel, 2007). The quadrupole acoustic source is generated by turbulence and it was found to be weak when the vocal folds are vibrating or for a convergent-divergent glottal shape. However, it was found to dominate the sound field for straight and converging glottal shaped models in the absence of vocal folds vibration (Zhang *et al.*, 2004). The intra-glottal separation vortices and their correspondent negative pressures may influence all these acoustic sources.

The negative static pressure on the vocal folds walls that is induced by the intra-glottal vortical structures exerts lateral closing force on the divergent section of the glottis. These forces may result in the acceleration of the closing phase of the phonation cycle relative to the opening phase affecting the monopole acoustic source. This asymmetry of

the cycle may have an important implication on voice generation mechanisms. It is known that rapid glottal closing correlates well with voice intensity and loudness (Sundberg and Gauffin, 1979; Woo, 1996; Stevens, 1998).

In summary, the present LES study of a symmetric larynx model with a divergent glottis shows that the pressure loads on the divergent glottal slope are not uniform and unsteady, and are affected by the vortical structures that are convected by the flow near the vocal folds. It is believed that in the case of a pulsatile flow situation or in the case when the vocal folds are moving, the intra-glottal vortices characteristic to the closing phase will be even more important since they are regulated by the flow cycle. Their strength may be also increased for higher trans-glottal pressures or decreased for a less divergent glottal angle. The process by which these separation vortices are generated implies that in order to characterize them, unsteady computational methods that determine the instantaneous flow field must be used, the steady RANS models not being suitable for such calculations.

The results obtained in the present unsteady computational study prove the hypothesis that intra-glottal vortices form in the superior aspect of the divergent glottis and that these vortices generate significant negative aerodynamic pressures, relative to both ambient pressure and maximum sub-glottal pressure.

ACKNOWLEDGMENT

This project was supported by Award No. R01 DC009435 from the National Institute on Deafness and Other Communication Disorders.

Alipour, F., and Scherer, R. C. (**2000**). "Dynamic glottal pressures in an excised hemilarynx model," J. Voice **14**, 443–454.

Alipour, F., and Scherer, R. C. (**2004**). "Flow separation in a computational oscillating vocal fold model," J. Acoust. Soc. Am. **116**, 1710–1719.

Barney, A., Shadle, C. H., and Davies, P. O. A. L. (**1999**). "Fluid flow in a dynamic mechanical model of the vocal folds and tract. I. Measurements and theory," J. Acoust. Soc. Am. **105**, 444–455.

Decker, G. Z., and Thomson, S. L. (**2007**). "Computational simulations of vocal fold vibration: Bernoulli versus Navier-Stokes," J. Voice **21**, 273–284.

Degani, D., Seginer, A., and Levy, Y. (**1990**). "Graphical visualization of vortical flows by means of helicity," AIAA J. **28**, 1347–1352.

Fant, G. (**1982**). "Preliminaries to analysis of the human voice source," Department of Speech, Music, and Hearing, Quarterly Progress and Status Report (KTH), STL-QPSR Report No. 23(4)/1982, http://www.speech.kth.se/prod/publications/files/qpsr/1982/1982_23_4_001-027.pdf (Last viewed 11/2/2009).

Gauffin, J., and Liljencrants, J. (**1988**). "Modeling the air flow in the glottis," Ann. Bull. RILP **22**, 41–52.

Gauffin, J., and Sundberg, J. (**1989**). "Spectral correlates of glottal voice source waveform characteristics," J. Speech Hear. Res. **32**, 556–565.

Gobl, C. (**1989**). "A preliminary study of acoustic voice quality correlates," Department of Speech, Music, and Hearing—Quarterly Progress and Status Report (KTH), STL-QPSR Report No. 30(4)/1989, http://www.speech.kth.se/prod/publications/files/qpsr/1989/1989_30_4_009-022.pdf (Last viewed 11/2/2009).

Gorham-Rowan, M., and Morris, R. (**2006**). "Aerodynamic analysis of male-to-female transgender voice," J. Voice **20**, 251–262.

Hanson, H. M. (**1997**). "Glottal characteristics of female speakers: Acoustic correlates," J. Acoust. Soc. Am. **101**, 466–481.

Hofmans, G. C. J., Groot, G., Ranucci, M., Graziani, G., and Hirschberg, A. (**2003**). "Unsteady flow through in-vitro models of the glottis," J. Acoust. Soc. Am. **113**, 1658–1675.

Holmberg, E., Hillman, R. E., and Perkell, J. S. (**1988**). "Glottal air flow and transglottal pressure measurements for male and female speakers in soft, normal and loud voice," J. Acoust. Soc. Am. **84**, 511–529.

Khosla, S. M., Murugappan, S., and Gutmark, E. J. (**2008b**). "What can vortices tell us about vocal fold vibration and voice production," Curr. Opin. Otolaryngol. Head Neck Surg. **16**, 183–187.

Khosla, S. M., Murugappan, S., Gutmark, E. J., and Scherer, R. C. (**2007**). "Vortical flow field during phonation in an excised canine larynx model," Ann. Otol. Rhinol. Laryngol. **116**, 217–228.

Khosla, S. M., Murugappan, S., Lakhamraju, R., and Gutmark, E. J. (**2008a**). "Using particle image velocimetry to measure anterior-posterior velocity gradients in the excised canine larynx model," Ann. Otol. Rhinol. Laryngol. **117**, 134–144.

Khosla, S. M., Murugappan, S., Paniello, R., Ying, J., and Gutmark, E. J. (**2009**). "Role of vortices in voice production: Normal versus asymmetric tension," Laryngoscope **119**, 216–221.

Klatt, D. H., and Klatt, L. C. (**1990**). "Analysis, synthesis, and perception of voice quality variations among female and male talkers," J. Acoust. Soc. Am. **87**, 820–857.

Krane, M. H., and Wei, T. (**2006**). "Theoretical assessment of unsteady aerodynamic effects in phonation," J. Acoust. Soc. Am. **120**, 1578–1588.

Kucinschi, B. R., Scherer, R. C., DeWitt, K. J., and Ng, T. T. M. (**2006**). "Flow visualization and acoustic consequences of the air moving through a static model of the human larynx," J. Biomech. Eng. **128**, 380–390.

Lilly, D. K. (**1966**). "On the application of the eddy viscosity concept in the inertial subrange of turbulence," NCAR Manuscript No. 123.

Lous, N. J. C., Hofmans, G. C. J., Veldhuis, R. N. J., and Hirschberg, A. (**1998**). "A symmetrical two-mass vocal-fold model coupled to vocal tract and trachea, with application to prosthesis design," Acta. Acust. Acust. **84**, 1135–1150.

McGowan, R. (**1988**). "An aeroacoustic approach to phonation," J. Acoust. Soc. Am. **83**, 696–704.

Mihaescu, M., Gutmark, E. J., Khosla S. M., Scherer, R. C., and Fuchs, L. (**2007**). "Flow and acoustics simulations based on LES and an acoustic analogy; an application to laryngeal airflow," AIAA Paper No. AIAA 2007-919.

Murugappan, S., Khosla, S. M., Casper, K., Oren, L., and Gutmark, E. J. (**2009**). "Flow fields and acoustics in a unilateral scarred vocal fold model," Ann. Otol. Rhinol. Laryngol. **118**, 44–50.

Neubauer, J., Zhang, Z., Miraghaie, R., and Berry, D. A. (**2007**). "Coherent structures of the near field flow in a self-oscillating physical model of the vocal folds," J. Acoust. Soc. Am. **121**, 1102–1118.

Nomura, H., and Funada, T. (**2007**). "Effects of the false vocal folds on sound generation by an unsteady glottal jet through rigid wall model of the larynx," Acoust. Sci. Technol. **28**, 403–412.

Pelorson, X., Hirschberg, A., van Hassel, R. R., Wijnands, A. P. J., and Auregan, Y. (**1994**). "Theoretical and experimental study of quasisteady—Flow separation within the glottis during phonation. Application to a modified two-mass model," J. Acoust. Soc. Am. **96**, 3416–3431.

Pope, S. B. (**2000**). *Turbulent Flows* (Cambridge University Press, Cambridge, England).

Sapienza, C. M., and Stathopoulos, E. T. (**1994**). "Comparison of maximum flow declination rate: Children versus adults," J. Voice **8**, 240–247.

Shinwari, D., Scherer, R. C., DeWitt, K. J., and Afjeh, A. A. (**2003**). "Flow visualization and pressure distributions in a model of the glottis with a symmetric and oblique divergent angle of 10 degrees," J. Acoust. Soc. Am. **113**, 487–497.

Stevens, K. N. (**1998**). *Acoustic Phonetics* (MIT, Cambridge, MA).

Story, B. H., and Titze, I. R. (**1995**). "Voice simulation with a body-cover model of the vocal folds," J. Acoust. Soc. Am. **97**, 1249–1260.

Suh, J., and Frankel, S. H. (**2007**). "Numerical simulation of turbulence transition and sound radiation for flow through a rigid glottal model," J. Acoust. Soc. Am. **121**, 3728–3739.

Suh, J., and Frankel, S. H. (**2008**). "Comparing turbulence models for flow through a rigid glottal model (L)," J. Acoust. Soc. Am. **123**, 1237–1240.

Sundberg, J., and Gauffin, J. (**1979**). in "Waveform and spectrum of the glottal voice source," *Frontiers of Speech Communication Research*, edited by B. Lindholm and S. Ohman (Academic, New York), pp. 301–320.

Titze, I. R. (**2000**). *Principles of Voice Production* (National Center for Voice and Speech, Iowa City, IA).

Triep, M., Brucker, C., and Schroder, W. (**2005**). "High-speed PIV measurements of the flow downstream of a dynamic mechanical model of the human vocal folds," Exp. Fluids **39**, 232–245.

Van Doormaal, J. P., and Raithby, G. D. (**1984**). "Enhancements of the

SIMPLE method for predicting incompressible fluid flows," Numer. Heat Transfer **7**, 147–163.
Wilcox, D. C. (**1993**). *Turbulence modeling for CFD* (DCW Industries, La Canada, CA).
Woo, P. (**1996**). "Quantification of videostrobolaryngoscopic findings-measurements of the normal glottal cycle," Laryngoscope **106**, 1–27.
Zhang, C., Zhao, W., Frankel, S. H., and Mongeau, L. (**2002a**). "Computational aeroacoustics of phonation, Part II: Effects of flow parameters and ventricular folds," J. Acoust. Soc. Am. **112**, 2147–2154.
Zhang, Z., Mongeau, L., Frankel, S. H., Thomson, S., and Park, J. B. (**2004**). "Sound generation by steady flow through glottis-shaped orifices," J. Acoust. Soc. Am. **116**, 1720–1728.
Zhao, W., Frankel, S. H., and Mongeau, L. (**2001**). "Numerical simulations of sound from confined pulsating axisymmetric jets," AIAA J. **39**, 1868–1874.
Zhao, W., Zhang, C., Frankel, S. H., and Mongeau, L. (**2002b**). "Computational aeroacoustics of phonation, Part I: Computational methods and sound generation mechanisms," J. Acoust. Soc. Am. **112**, 2134–2146.

Acoustic consequences of articulatory variability during productions of /t/ and /k/ and its implications for speech error research

Stefania Marin[a)]
Institute of Phonetics and Speech Processing, Ludwig-Maximilians-University Munich, 80799 Munich, Germany

Marianne Pouplier
Institute of Phonetics and Speech Processing, Ludwig-Maximilians-University Munich, 80799 Munich, Germany and Haskins Laboratories, New Haven, Connecticut 06511

Jonathan Harrington
Institute of Phonetics and Speech Processing, Ludwig-Maximilians-University Munich, 80799 Munich, Germany

(Received 30 March 2009; revised 13 August 2009; accepted 27 October 2009)

An increasing number of studies has linked certain types of articulatory or acoustic variability with speech errors, but no study has yet examined the relationship between such articulatory variability and acoustics. The present study aims to evaluate the acoustic properties of articulatorily errorful /k/ and /t/ stimuli to determine whether these errors are consistently reflected in the acoustics. The most frequent error observed in the articulatory data is the production of /k/ and /t/ with simultaneous tongue tip and tongue dorsum constrictions. Spectral analysis of these stimuli's bursts shows that /k/ and /t/ are differently affected by such co-production errors: co-production of tongue tip and tongue dorsum during intended /k/ results in typical /k/ spectra (and hence in tokens robustly classified as /k/), while co-productions during intended /t/ result in spectra with roughly equal prominence at both the mid-frequency (/k/-like) and high-frequency (/t/-like) ranges (and hence in tokens ambiguous between /k/ and /t/). This outcome is not due to an articulatory timing difference, but to tongue dorsum constriction having an overall greater effect on the acoustic than a tongue tip constriction when the two are co-produced.
 [DOI: 10.1121/1.3268600]

PACS number(s): 43.70.Fq [DAB] Pages: 445–461

I. INTRODUCTION

The current study investigates the relationship between articulation and acoustics in utterances with alternating initial consonants in order to increase our understanding of how articulatory variability during production of coronal and dorsal stops, potentially resulting in speech errors, is reflected in the acoustic signal. Several studies have found that the alternating consonants in such utterances display an increased amount of articulatory and acoustic variability compared to environments in which these consonants do not alternate. Under certain circumstances, this variability may be perceived as errorful by listeners (Boucher, 1994; Goldstein *et al.*, 2007; McMillan *et al.* 2009; Mowrey and MacKay, 1990; Pouplier, 2003, 2007, 2008; Pouplier and Goldstein, 2005). However, there has been no systematic study to examine the relationship between such articulatory variability and its acoustic consequences. That is, previous studies either examined articulatory variability without a systematic analysis of its impact on acoustics or they examined acoustic variability without the availability of information on the articulations that give rise to the acoustics. Due to the complex, nonlinear relationship between articulation and acoustics, it is, however, important to gain a principled understanding of how errorful articulatory and acoustic variability relate to each other. The present study aims to address this issue by evaluating the acoustic properties of /k/ and /t/ stimuli with known articulatory configurations in order to determine how the increased articulatory variability may be reflected acoustically and also to determine whether it may be reflected acoustically equally for both intended /k/ and /t/.

An examination of the acoustic properties of errors is especially opportune given the great importance of error data for models of speech production (Dell, 1986; Levelt, 1989; Levelt *et al.*, 1999; Rapp and Goldrick, 2000) and in view of the divergent findings between studies that rely on auditory perception for error detection on the one hand, and studies that investigate errors articulatorily on the other. Thus based on the former, speech errors have traditionally been described in categorical terms as substitution of one symbolic phonological unit with another (Fromkin, 1971, 1973; Shattuck-Hufnagel, 1979, 1983). However, physiological studies have shown first that such perceived categorical substitutions may, in fact, involve the simultaneous production of two constrictions (one intended and one errorful), and

[a)]Author to whom correspondence should be addressed. Electronic mail: marin@phonetik.uni-muenchen.de

second that these constrictions may be produced along a gradient continuum of magnitudes, exhibiting patterns intermediate between those typical for a given category (Boucher, 1994; Goldstein *et al.*, 2007; McMillan *et al.*, 2009; Mowrey and MacKay, 1990; Pouplier, 2003, 2007, 2008).

Different results have also emerged from auditory and articulatory analyses as far as the extent to which coronals and non-coronals are prone to errors. The auditory analysis of Stemberger (1991) showed that coronals are disproportionally susceptible to be replaced by non-coronals in errors. He interpreted these results as providing strong evidence for the phonological underspecification of coronals since an un-(der)specified place of articulation can easily be "replaced" by any other, phonologically specified place of articulation. Articulatory studies on the other hand (Goldstein *et al.*, 2007; Pouplier, 2003, 2008) did not observe this asymmetry in error-proneness between /k/ and /t/: the articulatory errors documented in these studies occurred with equal frequency during productions of both /k/ and /t/. We are left, then, with the following discrepancy: auditory analyses suggest a place asymmetry in speech errors, but physiological analyses have shown that there is none.

An explanation for the discrepancy between these results is that errors involving coronals may be auditorily more salient. That is, the articulatory deviation due to a production error may result in a greater acoustic deviation from the corresponding error-free production for coronals than for dorsals. Under this interpretation, the bias reported in auditorily-based analyses may come about not because coronals are inherently more error-prone, but instead because the hypothesized greater acoustic deviation makes the error much more noticeable than in the case of dorsals. It is precisely this issue that we test in the present study by examining the acoustic properties of speech errors with known articulatory configurations.

Our study is therefore concerned with how articulatory variability is reflected in the acoustic domain: we investigate "partial" errors that involve simultaneous production (henceforth co-production) of two constrictions—one intended, one un-intended (or intruding)—of the type reported by Pouplier (2003) (cf. also Goldstein *et al.*, 2007). Specifically, that study used a repetition task to elicit productions of /k/ and /t/ in utterances with alternating initial consonants (e.g., *cop top*), and the most frequent error observed in their articulatory data was an intrusion of a second articulatory gesture (or constriction) without the intended gesture being reduced. This resulted in the simultaneous presence of both tongue tip (TT) and tongue dorsum (TD) constrictions during the production of errorful /k/ or /t/: while /k/ is normally produced with only a tongue dorsum constriction, in an error, a tongue tip constriction is produced simultaneously with the dorsal constriction, without the intended dorsal constriction being reduced. The intruding tongue tip gesture can display a range of gestural magnitudes, ranging from minimal variability to an articulator position typical for the production of a coronal stop. Likewise during the production of an intended /t/, they observed that the intended tongue tip constriction can, in errorful tokens, be co-produced with an intruding tongue dorsum constriction. In terms of their frequency of occurrence, these co-productions were the most frequent type of error observed; reductions in intended constrictions (e.g., a tongue tip constriction for intended /t/ produced with a reduced constriction) and substitutions (e.g., *only* a /k/-like tongue dorsum constriction during intended /t/) were observed only rarely.

Furthermore, co-productions occurred with equal frequency during intended /k/ and /t/. However, although this type of error was as likely to occur during production of either /k/ or /t/, there has been a reported asymmetry in the perceptual consequences of co-productions during /k/ vs /t/. In their perception experiment, using a subset of the data collected by Pouplier (2003), Pouplier and Goldstein (2005) found that intruding tongue dorsum gestures during /t/ had a systematic perceptual effect, resulting in significantly lower correct identification scores and longer reaction times than for error-free /t/. However, intruding tongue tip gestures during intended /k/ did not significantly affect identification or reaction times. It seems therefore that while /k/ and /t/ are equally prone to error at the articulatory level, the perceptual effects of co-productions are different depending on whether /k/ or /t/ is the intended consonant. Since the acoustic properties of co-productions have not been investigated so far, it is not known whether this difference is due to perceptual biases (cf. for instance, Hume *et al.*, 1999; Smits *et al.*, 1996; Winters, 2000 on the lower perceptual salience of singleton coronals compared to labials and velars) or whether the influence of the additional constriction exerted on the acoustics varies as a function of the constricting articulator/place of articulation of the intruding gesture. By examining the acoustic properties of tokens produced with an intruding gesture, the current study can determine, as described below, whether the previously reported perceptual asymmetry between coronals and dorsals produced with an intruding gesture is due to a difference at the acoustic level or to a perceptual/auditory bias.

It is well established that articulatory changes in the vocal tract will affect the acoustic output to different degrees, depending on the exact location and size of a given constriction formation (cf. Stevens, 1972, 1989). Since in co-productions the intended constriction is stronger both spatially and temporally compared to the intruding constriction (Pouplier, 2003; Pouplier and Waltl, 2008; Pouplier and Goldstein, in press), "adding" a coronal constriction to an intended dorsal constriction may affect the vocal tract area function differently compared to adding a dorsal constriction to an intended coronal constriction. Specifically, an intruding constriction formed behind the main constriction (as is the case for an intended /t/ with an intruding dorsal constriction) could affect the pressure build-up and hence the acoustic properties of the consonant release more than an intruding constriction formed in front of the main constriction (as is the case for an intended /k/ with an intruding coronal constriction). As a result, the acoustics of intended coronals could be affected more by an intruding tongue dorsum gesture than the acoustics of dorsals by a tongue tip intrusion. If an intruding constriction (or gesture) affects the acoustic properties of coronals but not the acoustic properties of dorsals, then the observed perceptual asymmetry could be based

on differences at the acoustic level. Under this hypothesis, we predict that an intruding tongue dorsum gesture during /t/ will result in tokens acoustically different from error-free /t/, while an intruding tongue tip gesture during /k/ will result in tokens acoustically similar to error-free /k/.

On the other hand, Winitz *et al.* (1972) remarked that "it is difficult to smear a high-energy English /t/ into a /p/ or a /k/, but not difficult to smear /p/ into /k/ or the reverse" (pp. 1972–1973), suggesting that acoustically, /t/ bursts should be at least as robust as /k/ bursts, and if anything dorsals should be acoustically more sensitive to articulatory variability compared to coronals. If so, the perceptual results reported by Pouplier and Goldstein (2005) could be due not to an asymmetry at the acoustic level but to a bias exclusively at the auditory/perceptual level. It has been observed that perception of coronal stops produced by native speakers of several languages is generally worse than that of non-coronals either in singleton contexts (Hume *et al.*, 1999; Smits *et al.*, 1996; Winters, 2000)[1] or in clusters (Kochetov and So, 2007; Surprenant and Goldstein, 1998), and also that their perception is degraded more readily than perception of non-coronals by articulatory variability. Byrd (1992) and Chen (2003) observed, using articulatory synthesized stimuli, that perception of coronal stops is more vulnerable to increased articulatory overlap than perception of labial stops. It is not implausible therefore to expect that an intruding gesture may affect the acoustic properties of both coronals and dorsals, but that perception is more sensitive to the changes in the acoustic patterns of coronals compared to dorsals (or in other words, that more acoustic variability is tolerated in dorsals than in coronals before their perception is degraded). Under this hypothesis, we predict that an intruding tongue dorsum gesture during either /t/ or /k/ will result in tokens acoustically different from error-free /t/ and /k/, and that the perceptual patterns reported by Pouplier and Goldstein (2005) may be due to a bias in how robustly these acoustic changes can be auditorily detected. To decide between these two possibilities, we analyzed the relationship between articulation and acoustics for /t/ and /k/ tokens that were produced with an intended as well as an intruding gesture, the latter covering a range of different gestural magnitudes.

A further question we address is whether the acoustic consequences of co-produced intended and intruding gestures are influenced by the timing relation between these two gestures. It is conceivable that different timing relations between the two gestures have different acoustic consequences, depending on which gesture is released last. Pouplier and Goldstein (2005) did not control for articulatory timing in their experiment. Thus, the perceptual pattern they reported may not be due to an asymmetry in how acoustics and/or perception are affected by tongue dorsum vs tongue tip intrusion gestures but rather to different timing patterns between intended and intruding articulations. Pouplier and Goldstein (in press) performed a detailed articulatory timing analysis of the Pouplier (2003) data and found that generally the intended gesture was released last in the case of both intended /k/ and /t/, with relatively few tokens showing the reverse pattern. In conjunction with the results of the perception experiment, this suggests that timing patterns are not a strong predictor of the acoustic consequences: the perception experiment had reported an asymmetry between coronal and dorsal stimuli, yet the timing revealed that for both coronals and dorsals, it was usually the intended gesture that was released last. However, since timing was not explicitly controlled for in the perception experiment, it is possible that, by chance, the perceptual experiment included errorful /t/ tokens with the intruding gesture released last, and no errorful /k/ tokens with the intruding gesture released last, resulting in the systematic lower identification scores for errorful /t/ than /k/. We examined therefore the extent to which acoustics were determined not only by the presence/absence of an intruding gesture but also by the intruding gesture's relative timing to the intended gesture.

II. METHOD

A. Gestural classification

To determine the relationship between the acoustic and articulatory properties of speech errors, we used data for which both acoustic and articulatory information was available. The dataset consisted of simultaneously recorded articulatory and acoustic data collected by Pouplier (2003). Articulatory data were recorded using the Perkell-system articulograph at Haskins Laboratories (Perkell *et al.*, 1992). For the recordings, four sensors were placed on the tongue: TT (attached about 1 cm behind the actual tongue tip), anterior tongue body, posterior tongue body, and TD. Additional sensors were placed one each on the upper and lower lips, the lower teeth to track jaw movement, and, to be able to correct for head movement, the nose ridge and the upper incisors. Standard calibration and postprocessing techniques were performed for each experiment. The articulatory data were sampled at 500 Hz and low-pass filtered at 15 Hz during postprocessing. For the simultaneous recording of acoustic data, a Sennheiser shotgun microphone was positioned about 1 m in front of the subject; acoustic data were sampled at 20 kHz and for one subject (JP) at 48 kHz.

Data from seven speakers of American English were recorded. Subjects were instructed to produce utterances with alternating onset consonants (*cop top, top cop, kip tip, tip kip*) in synchrony to a metronome beat. For the duration of each trial, the subjects saw the utterance they were instructed to pronounce on a computer screen in front of them, and stress placement was indicated in capital letters (e.g., COP top). Experimental variables included two stress conditions (iambic vs trochaic), two vowel contexts (*top cop* vs *tip kip*), phrase position (*top cop* vs *cop top*), and three speaking rates (fast, at 120 beats per minute, medium at 104 beats per minute, and slow at 80 beats per minute, allowing for speaker-specific adjustments within a ± 4 beats per minute range of the target rates). Productions of utterances with nonalternating initial consonants (*cop cop, top top, kip kip, tip tip*) were also included in the dataset and served as controls for the analysis of articulator kinematics, but only stops produced in alternating trials were included in the acoustic analyses. The experimental variables were fully crossed. The data recording and processing procedures are detailed in Pouplier (2003).

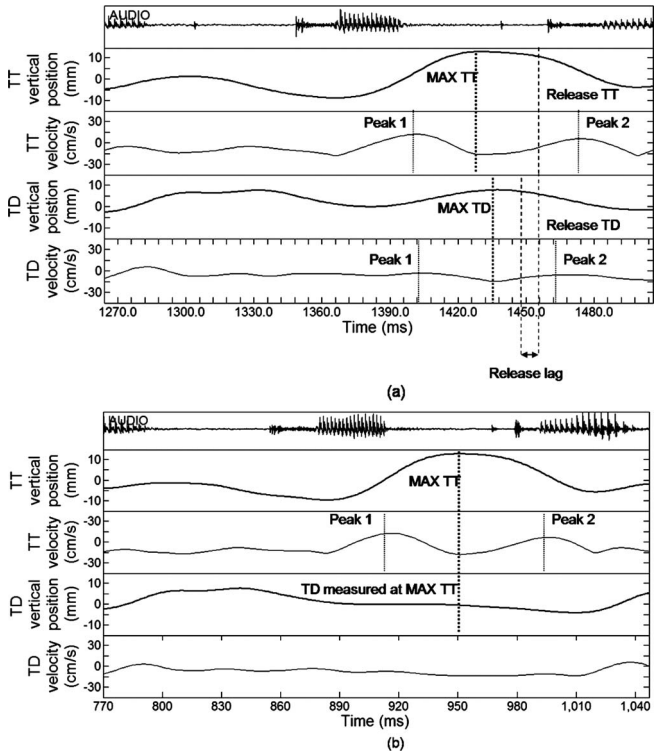


FIG. 1. Example articulatory measurements for /t/ in "top," illustrated on the basis of two "cop top" repetitions produced by subject JX at the fast speaking rate. The panels show vertical position (mm) and vertical velocity (cm/s) profiles of the tongue tip and tongue dorsum sensors. Identified on the figures are maximum velocity points during movement toward the constriction (Peak 1) and away from the constriction (Peak 2) as well as the constriction maxima (MAX) and releases. (a) Production of "top" for which both tongue tip and tongue dorsum movement showed a constriction. Temporal lag between the release of the tongue tip and the release of the tongue dorsum constriction is indicated by the arrow. (b) Production of "top" showing a tongue tip constriction only. For such tokens, tongue dorsum height was extracted at the time of tongue tip maximum. No temporal lag between release of tongue tip and tongue dorsum could be calculated for such tokens.

The movement time functions obtained through the electromagnetic midsagittal articulometer system were analyzed using software algorithms developed at Haskins Laboratories. For consonants /t/ and /k/, vertical position maxima of the tongue tip and tongue dorsum transducer coils were automatically determined on the basis of changes in their velocity profiles. As illustrated in Fig. 1, a vertical position maximum (MAX) was defined as the kinematic event where velocity was at its minimum between two velocity peaks corresponding to the articulator moving toward and away from the constriction. If the labeling algorithm did not find a maximum in one of the signals (e.g., in tongue dorsum during /t/ and tongue tip during /k/), its vertical amplitude value was measured at the time of a maximum in the other signal which the algorithm had identified. For instance, if there was no vertical position maximum for tongue dorsum during /t/ (since the tongue dorsum is not expected to rise during /t/), tongue dorsum was measured at the time of the tongue tip maximum [cf. Fig. 1(b)].

The vertical position maxima in alternating utterances span a whole range of values, from minimally variable to values typical for a canonical controlled stop constriction. Production of /k/ and /t/ was determined as "errorful" or "error-free" on the basis of articulator height (cf. Pouplier, 2003). Typical vertical articulatory positions for TT and TD during production of /k/ and /t/ were determined on the basis of matching conditions from the non-alternating utterances (cf. Goldstein *et al.*, 2007; Pouplier, 2003). Based on a working criterion, tokens were considered errorful if TT height during /k/ and TD height during /t/ were 2 standard deviations away from their mean in non-alternating productions. Errorful tokens were further classified into two constriction magnitude categories—"gradient" and "categorical:" if the maximal vertical position of TT/TD was within 2 standard deviations of the other category mean, the error was considered categorical (C); otherwise it was considered gradient (G). In order to evaluate how the acoustics are affected by different vertical articulator positions (indicative of constriction degree, at least for stop consonants), we followed the classification of Pouplier (2003) procedure here. Several studies have shown that the increased variability in articulator height can, in the right circumstances, be perceived as a speech error (Goldrick and Blumstein, 2006; Pouplier and Goldstein, 2005; Wood, 1997). Under which conditions any given production can be deemed errorful has been subject to vigorous debate. The current paper focuses on the acoustic consequences of variability of articulator height which is observed in utterances that typically elicit speech errors; the question how to negotiate the relationship of "variability" and "error" for any given token is not the focus of the paper. The terms error-free and errorful are labels of convenience to denote ranges of expected values of articulator positions.

Co-production errors were defined as the simultaneous production of an un-intended (or intruding) TT or TD gesture with the intended one. An error-free /t/ is produced with an active TT closure gesture and no actively controlled TD; a co-production error during intended /t/ occurs when a TD gesture is produced along with TT. An error-free /k/ is produced with an active TD closure gesture and no actively controlled TT; a co-production error during intended /k/ occurs when TT is produced along with TD. To summarize, on the basis of the articulatory metric, the following error-free and errorful /k/ and /t/ categories were defined:

- K: error-free /k/—produced with tongue dorsum and tongue tip heights typical for /k/ as determined on the basis of the non-alternating utterances;
- T: error-free /t/—produced with tongue dorsum and tongue tip heights determined typical on the basis of /t/ production in non-alternating utterances;
- KC: /k/ produced with a categorical error—tongue dorsum height is typical for /k/ in non-alternating utterances, tongue tip height is within 2 standard deviations of the mean of non-alternating /t/;
- KG: /k/ produced with a gradient error—tongue dorsum height is typical for /k/ in non-alternating utterances, tongue tip height is 2 standard deviations away from the mean of non-alternating /k/, but not within 2 standard deviations of the mean of non-alternating /t/;
- TC: /t/ produced with a categorical error—tongue tip height is typical for /t/ in non-alternating utterances, tongue dorsum height is within 2 standard deviations of a typical /k/; and

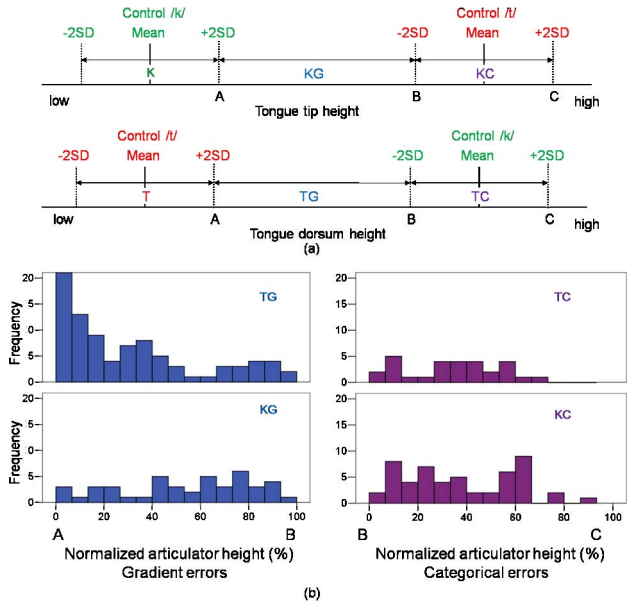


FIG. 2. (Color online) (a) Schematized cut-off points for gradient and categorical errors determined on the basis of the distributions of articulator height during /k/ and /t/ productions in control (non-alternating) utterances. The gradient error interval is defined by the cut-off points labeled with A and B; the categorical error interval is defined by the cut-off points labeled with B and C. (b) Histograms showing the distribution of normalized articulator heights for errorful tokens within the respective category ranges (between cut-off points A and B for gradient errors and between B and C for categorical errors). Plotted are tokens from speaker JX.

- TG: /t/ produced with a gradient error—tongue tip height is typical for /t/ in non-alternating utterances, tongue dorsum height is 2 standard deviations away from a typical /t/, but not within 2 standard deviations of a typical /k/.

Figure 2(a) illustrates the error metric schematically; Fig. 2(b) gives the normalized distributions of tokens within the gradient and categorical error categories for speaker JX (the speaker whose data are used in the classification analysis below).[2] The histograms in Fig. 2(b) show that categorically errorful /k/ and /t/ are similarly distributed within their respective category. For gradient errors, the TG distribution is skewed leftward, indicating that there are more tokens with tongue dorsum heights closer to the values for error-free /t/ than for error-free /k/.

Our data, summarized in Table I, consisted therefore of /k/ and /t/ produced in alternating utterances (*cop top* and similar) and defined articulatorily as errorful (KC, KG, TC, and TG) or error-free (K and T). Subsets of the data were used for various analyses, as will be detailed in Secs. II B and II E.

B. Acoustic analysis procedure

For each articulatorily classified initial consonant (K, T, KC, KG, TC, and TG), the acoustic signal was labeled in PRAAT 4.6.38 (Boersma and Weenink) from the release burst to the beginning of periodic vibration for the following vowel. There is evidence that the place of articulation in an oral stop can be related to the shape of the spectrum in the burst; dorsals are characterized by a mid-frequency spectral peak, resulting in a "compact" spectrum, while coronals are characterized by a high-frequency peak of greater energy than at lower frequencies, resulting in a rising spectrum (Blumstein and Stevens, 1979; Fant, 1960; Halle *et al.*, 1957; Jakobson *et al.*, 1963; Stevens and Blumstein, 1978; Stevens, 1989). It has also been established that temporal information in the spectra is important for the identification of place of articulation of stop consonants (Kewley-Port, 1983; Kewley-Port and Luce, 1984; Kewley-Port *et al.*, 1983), and that it contains cues that are especially important for the identification of dorsals, that is, their compact spectrum persists as a function of time in the burst (Kewley-Port, 1983). Furthermore, previous studies (Forrest *et al.*, 1988; Nossair and Zahorian, 1991) have shown that voiceless stops can be successfully classified on the basis of statistical properties such as mean, slope, and curvature of their burst over time. We therefore analyzed /k/ and /t/ on the basis of the time-varying spectral information available in their burst (from acoustic release to vowel onset) expressed in terms of the spectrum mean, slope, and curvature; the analysis was performed using the EMU speech data analysis system (Cassidy and Harrington, 2001). The spectral data were calculated with a 256 point discrete Fourier transform with a 40 Hz frequency resolution and a 5 ms Blackman window shifting in 5 ms increments over the entire length of the burst-to-vowel interval sampled at 20 kHz. The frequency axis was warped to the auditory bark scale in the frequency range 0–8500 Hz (0–21.26 bark), using the formula in Traunmüller (1990).

The parametrization of the spectral data to include time-varying information is illustrated with an example in Fig. 3. After converting the spectra between the burst onset and vowel onset to bark, each stop consists of a running spectral display, as shown in Fig. 1(a), with bark spectra at intervals of 5 ms. The next step was to reduce each individual spectrum from such a running display to just three values and more specifically to the first three coefficients (C_0, C_1, C_2) that are obtained after applying the discrete cosine transformation (DCT). For an N-point bark spectrum, $x(n)$, extending in frequency from $n=0$ to $N-1$ points, the mth DCT-coefficient C_m $(m=0,1,2)$ was calculated with

$$C_m = \frac{2k_m}{N}\sum_{n=0}^{N-1} x(n)\cos\left(\frac{(2n+1)m\pi}{2N}\right),$$

$$k_m = \frac{1}{\sqrt{2}}, \quad m=0, \quad k_m = 1, \quad m \neq 0. \tag{1}$$

These three coefficients, which are the amplitudes of the first three half-cycle cosine waves derived from the DCT, are

TABLE I. Number of tokens from alternating trials produced by seven subjects.

	Articulatory category						
	K	T	KC	KG	TC	TG	Total
Number of tokens	873	733	125	186	105	282	2304

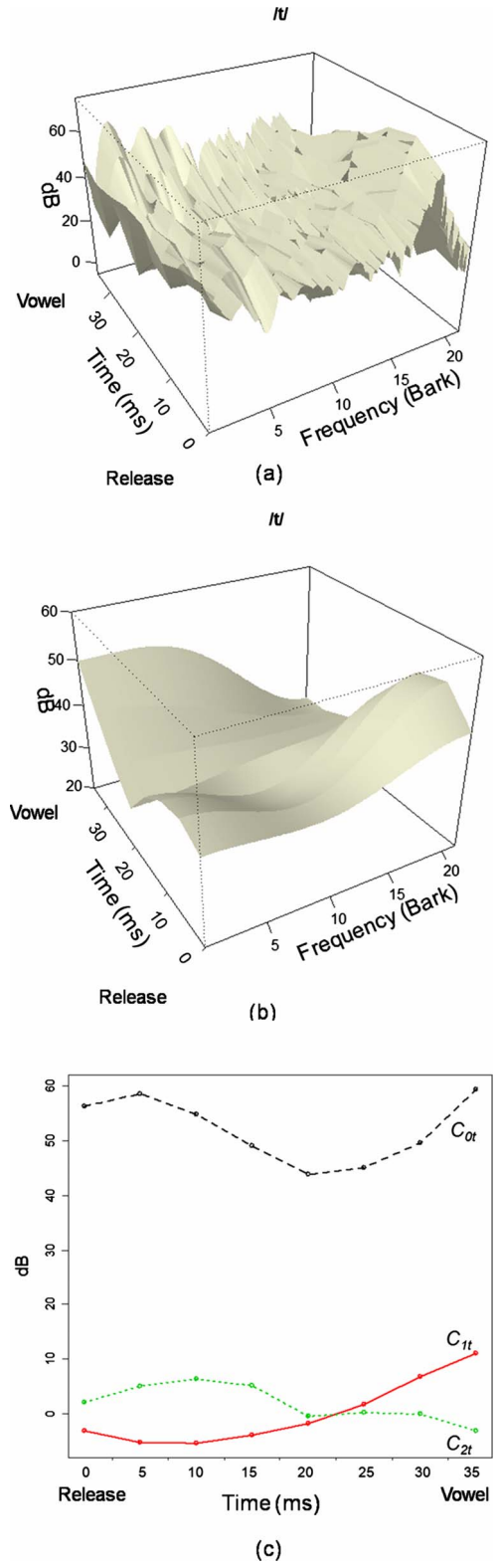


FIG. 3. (Color online) Analysis of a /t/ stop burst over time produced by subject JX in the phrase "cop *T*OP." (a) Bark-scale spectra at 5 ms intervals of the burst between the stop's release (burst onset) and vowel onset. (b) The same spectra as in (a) but smoothed using three DCT-coefficients (C_0, C_1, and C_2), capturing the individual spectra's mean, slope, and curvature. (c) The values of the DCT-coefficients (C_0, C_1, and C_2) obtained for each spectral slice from burst onset to vowel onset in 5 ms increments; each temporal curve (C_{0t}, C_{1t}, and C_{2t}) is fitted using the first three DCT-coefficients capturing its mean, slope, and curvature, resulting in nine coefficients that encode the spectral time-varying properties of the burst.

proportional to the spectrum's mean, slope, and curvature, respectively, and are also essentially equivalent to bark-scaled cepstral coefficients (see Harrington *et al.*, 2008; Nossair and Zahorian, 1991; Watson and Harrington, 1999 for formulas and further details). Consequently, after applying Eq. (1), the stop burst [Fig. 3(a)] was reduced to a triplet of values (the first three DCT-coefficients), with one triplet every 5 ms: these triplets of DCT-coefficients as a function of time can be used to derive a (cepstrally) smoothed spectrum [Fig. 3(b)]. We denote the triplets of DCT-coefficients across time by C_{0t}, C_{1t}, and C_{2t}, as shown in Fig. 3(c).

We now needed to find a way to compress C_{0t}, C_{1t}, and C_{2t} prior to classification. The approach that we used was to apply Eq. (1) again separately to each of the time-varying C_{0t}, C_{1t}, and C_{2t}. This is exactly equivalent to the methodology used in Harrington *et al.* (2008) for compressing time-varying formant trajectories. Thus, whereas each time-varying formant of a vowel was reduced with the DCT to three values in Harrington *et al.* (2008), here we have reduced separately the stop burst's time-varying spectral mean (C_{0t}), time-varying spectral slope (C_{1t}), and time-varying spectral curvature (C_{2t}) each to three values using Eq. (1), a procedure equivalent to a second order polynomial fitting. The end result, then, is that we have compressed the bark scaled spectral information of each stop burst between the release and vowel onset [Fig. 3(a)] to a single point in a nine-dimensional space and moreover in such a way that this compression encodes time-varying information.

C. Classification procedure

We classified the stops in the nine-dimensional space described in Sec. II B using a support vector machine (SVM) algorithm (Baayen, 2008; Bennett and Campbell, 2000; Cortes and Vapnik, 1995; Duda *et al.*, 2000). Non-technically, a SVM separates classes not on the basis of their center or mean but instead on the basis of their margins; to accomplish this, points are projected into a high dimensional space and a separating hyperplane is determined in this space. Among the advantages of this classification method is the fact that SVM makes no assumptions about normality of the data, and that it can handle cases where a class is broken into non-contiguous regions. This is especially important with speech data where phonetic classes can contain non-contiguous members due to context. This consideration is also important for our current dataset since we collapsed over several experimental variables (rate, vowel, stress, and phrase position).[3] Previous work on phonetic classification of vowels and consonants in running sentences (Clarkson and Moreno, 2000) has shown that SVM performs at least as well as Gaussian classifiers. SVM was implemented using the "e1071" library for R (Mayer, 2001) using a Gaussian radial basis function kernel. For classification of speech classes (both vowels and consonants), the choice of kernel is reported not to have a major impact on accuracy (Clarkson and Moreno, 2000).

For classification, the data were split into separate training and testing parts. Our training data were correctly produced stops from six speakers, and our test data were stops

TABLE II. Number of tokens used for training and testing: Training data were produced by six subjects; test data were produced by a new subject (JX) not included in the training.

	Articulatory category						
	K	T	KC	KG	TC	TG	Total
Training data	691	577					1268
Test data	182	156	52	44	29	86	549
Total	873	733	52	44	29	86	1817

produced by a new speaker (JX) not included in training (Table II). Speaker JX was selected for testing as this was the speaker whose production was also used in the perceptual experiment reported by Pouplier and Goldstein (2005), and we reasoned that selecting data from the same speaker for the acoustic analysis would allow for an investigation of the relationship between articulation, acoustics, and perception. Furthermore, this speaker's dataset was relatively balanced for number of error-free /t/ and /k/ tokens as well as for number of errorful tokens. Splitting this subject's data by the conditions stress, rate, phrase position, and vowel context resulted in a low number of errorful tokens for some of the conditions (cf. Table VII in the Appendix). For this reason, the acoustic analyses reported in the main body of the paper were carried out on data collapsed across the conditions stress, rate, phrase position, and vowel context. Analyses by conditions are, as far as they were feasible in terms of number of tokens, reported in the Appendix.

D. Acoustic proximity calculations

We further quantified the acoustic properties of each token in the test data by calculating the Mahalanobis (M) distance of each token to the centroids of both error-free /k/ (M_k) and /t/ (M_t) in the nine-dimensional space to which each token's burst properties were compressed. We used the Mahalanobis distance for its capability to factor in the categories' distributions (in our case the distributional shapes of error-free /k/ and /t/). Relative proximity P of each token to either error-free /k/ or /t/ was calculated as the difference between the two Mahalanobis distances (on the logarithmic scale), using the formula in Eq. (2):

$$P = \log(M_k) - \log(M_t), \tag{2}$$

where M_k is the Mahalanobis distance of a token to the centroid of error-free /k/ and M_t is the Mahalanobis distance of a token to the centroid of error-free /t/.

When P is 0, the token is equidistant between /k/ and /t/; when it is positive, the token is closer to /t/ than to /k/; and when it is negative, the token is closer to /k/ than to /t/.

E. Measures used from previous studies: *d′* perceptual score and articulatory timing

To assess the relationship between acoustics and perception, we used the perceptual measure employed by Pouplier and Goldstein (2005) in their experiment. To examine the role of relative timing of the intended and intruding gestures in shaping the tokens' acoustics, we use the timing measure employed by Pouplier and Goldstein (in press). In what follows we summarize these measures, and the data subsets for which these measures are available.

Pouplier and Goldstein (2005) used in their perceptual experiment a total of 60 tokens, selected from each articulatory classification category (no error, gradient error, and categorical error) and covering within their category a range of different articulator heights for the intruding gesture. Ten tokens were selected per category (except for KC that included 11 tokens and KG that included 9 tokens), all produced by the same speaker (JX). The perceptual responses were obtained for individual tokens from 11 native speakers of English in a go-no go perceptual identification task (for details, see Pouplier and Goldstein, 2005). For each token, we used the perceptual scores (d') (MacMillan and Creelman, 1991) obtained in their experiment.[4] These d' perceptual scores represent the difference between correct and incorrect identifications and were calculated using Eq. (3).

$$d' = z(H) - z(F), \tag{3}$$

where H is the proportion of correct identification responses relative to number of trials, F is the proportion of incorrect identification responses relative to number of trials, and z is the standardized score.

A d' score of 4.65 represents perfect accuracy ($H=99\%$, $F=1\%$; MacMillan and Creelman, 1991), while a negative score means that the proportion of incorrect responses is greater than that of correct ones. In this way, the consistent correct identification of any given token as either /k/ or /t/ has a maximum score of +4.65, while consistent incorrect identification has a negative score to a minimum of −4.65.

For relative timing between intended and intruding articulatory gestures, we used the timing measure employed by Pouplier and Goldstein (in press). For determining articulatory timing, the vertical movement time series of both the intended and intruding gestures were labeled at constriction release, as shown in Fig. 1(a) (cf. Pouplier and Goldstein, in press for details on the segmentation procedure employed).[5] Relative timing at release was calculated as the lag between intended and intruding gestures by subtracting the release timestamp of tongue tip from the release timestamp of tongue dorsum ($\text{lag}=\text{release}_{TD}-\text{release}_{TT}$). A negative lag means that the tongue tip constriction was released last, and a positive lag means that the tongue dorsum was released last. For subject JX, the release lag measure was available for 54 KC tokens, 41 KG tokens, 28 TC tokens, and 43 TG tokens. Because for most error-free tokens the uncontrolled articulator could not be measured [cf. Fig. 1(b)], relative timing for these tokens could not be computed and hence they were not included in the analysis.

III. RESULTS

A. SVM classification

For the classification analysis, training was performed on correctly produced stops from six speakers and testing on

TABLE III. Classification (%) from a SVM closed test in which training and testing were carried out on error-free K and T produced by all six subjects. The main diagonal shows the hit-rate per category. All percentages have been rounded.

	Articulatory category	
Closed test	K	T
Classified as /k/	95	5
Classified as /t/	11	89

TABLE V. Classification (%) in an open SVM test. The training data included the error-free T and K bursts from six subjects, and testing was carried out on all of subject JX's stops. All percentages have been rounded.

	Articulatory category					
	K	T	KC	KG	TC	TG
Classified as /k/	90	15	83	89	52	27
Classified as /t/	10	85	17	11	48	73

stops produced by speaker JX (cf. Table II). Results from both a closed test (training and testing on the training data) and a semi-open test (training and testing on the training data using a four-fold cross-validation) show that error-free /k/ and /t/ are correctly classified by the SVM for the vast majority of cases (Tables III and IV). It must be noted that no speaker normalization was performed for the training data, and that the data were undifferentiated for stress conditions, phrase positions, speech rates, and vowel contexts (cf. Sec. II A).

The test data from subject JX were classified on the basis of the training results obtained from error-free data from six subjects. The confusion matrix (Table V) for the test data shows that error-free /k/ and /t/ were accurately classified at 90% and 85%, respectively. While productions of /k/ with intruding tongue tip gestures were predominantly classified as /k/ (83% for categorical intrusions and 89% for gradient intrusions), there was greater ambiguity in whether errorful /t/ was classified as /k/ or /t/. Thus, about half of /t/ tokens produced with an intruding tongue dorsum of categorical magnitude (TC) were classified as /k/ and about half as /t/. The overall classification pattern, with ambiguity in the classification of TC, remained the same when the data were split for conditions stress, rate, phrase position, and vowel context (Table VIII in the Appendix). TC tokens were ambiguously classified as /k/ or /t/ regardless of stress, rate, phrase position, or vowel context, while no such ambiguity was observed for errorful /k/ tokens.

The distribution of posterior probabilities (Fig. 4) shows that most TC tokens occupied a space between unambiguous /k/ and /t/ (rather than being evenly assigned to the center of the /k/ and /t/ spaces). Thus, while /k/ tokens (errorful or not) and error-free /t/ tokens show a skewed distribution, with half of the tokens having a probability of 90% or more of being correctly classified, errorful /t/ tokens show a much flatter distribution along the probability continuum: for TC, only $\frac{1}{4}$ of the tokens have a probability over 90% of being classified as /t/, another $\frac{1}{4}$ have a probability under 20% of being correctly classified, with half of the tokens falling in the middle of the continuum. This asymmetry cannot be explained by the distributions of articulator heights between errorful /k/ and /t/. Articulator height histograms [Fig. 2(b)] showed that on the basis of vertical position of the intruding articulator, KC and TC were similarly distributed within their defined intervals. For the gradient categories, there were more TG tokens close to the boundary between error-free and errorful /t/ than were KG tokens to the boundary between error-free and errorful /k/ [boundary A in Fig. 2(a)]. Nevertheless, there were more ambiguous TG tokens than KG tokens, although articulator height distributions alone would have predicted the opposite pattern.

TABLE IV. Classification (%) in a semi-open SVM test using four-fold cross-validation. The training data were randomly split into four subsets, and each time three subsets were used for training and the fourth was used for testing. All percentages have been rounded.

Semi-open test (four-fold cross-validation)				
Total accuracy	89			
Single accuracies	91	84	91	91

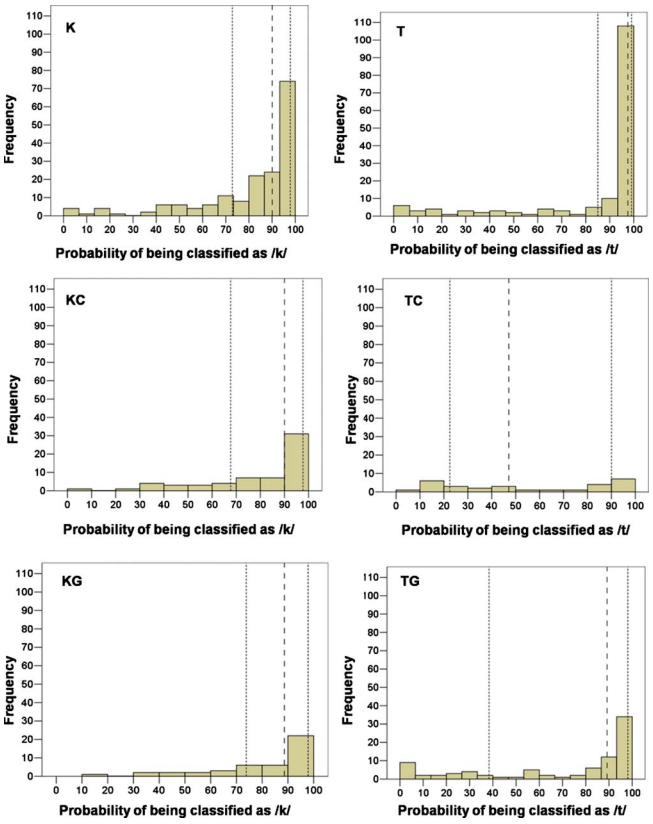


FIG. 4. (Color online) Histograms of posterior probabilities from a SVM classification showing the probability of being classified as /k/ for categories K, KC, and KG and of being classified as /t/ for categories T, TC, and TG. Reference marks (dotted lines) are placed at the 25th, 50th, and 75th percentiles, respectively, ordered from left to right.

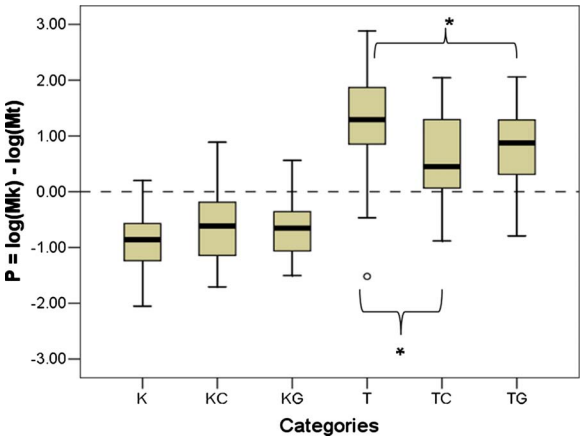


FIG. 5. (Color online) Boxplots showing the median (thick horizontal bar), interquartile range (boxes), error bars (vertical lines), and outliers (circles) for the acoustic measure *P*, the token proximity to /k/ (negative values), or /t/ (positive values). All /k/ categories were significantly different from all /t/ categories. Significant differences within /k/ or /t/ are indicated by *.

B. Proximity measure *P*

To quantify the acoustic similarity of errorful /k/ and /t/ to either typical velars or alveolars, we measured each token's proximity to the centroids of error-free /k/ and /t/. For this measure, described in Sec. II D, we expected that error-free /k/ and /t/ should each be close to their own centroids, resulting in negative *P* values for /k/ and positive values for /t/, which is indeed the pattern observed (Fig. 5). In terms of acoustic proximity *P* of the error tokens to either error-free /k/ or /t/, we observe similar median values for all three /k/ categories (error-free, categorical, and gradient error), but more divergent median values for the /t/ categories, with (some) error /t/ tokens having intermediate values between /k/ and /t/. A Welch's variance-weighted ANOVA (to correct for unequal variances) with dependent variable relative proximity index *P* and factor category (K, KC, KG, T, TC, and TG) was significant [$F(5,145.765)=267.183$, $p<0.001$]. A follow-up *posthoc* test (Games-Howell, appropriate for unequal variances and group sizes, cf. Toothaker, 1993) showed that none of the /k/ groups differed significantly from each other ($p>0.05$), while TC and TG differed from T ($p<0.001$), but not from each other ($p>0.05$); all /t/ groups also differed from the /k/ groups ($p<0.001$). These results show that while errorful /k/ tokens are acoustically close to error-free /k/, errorful /t/ tokens as a group are acoustically close to neither error-free /t/ nor /k/, a result that corroborates the distribution of probabilities from the classification analysis. This shows that an intruding tongue tip during /k/ results in tokens acoustically similar to error-free /k/, while an intruding tongue dorsum during /t/ results in tokens acoustically different from error-free /t/. Relative proximity patterns for the data split according to stress, rate, phrase position, and vowel context remained similar to the overall pattern with no diverging pattern in any of the conditions (Fig. 11 in the Appendix).

C. Relationship between acoustics and perception

Looking exclusively at the data subset used by Pouplier and Goldstein (2005) in their perceptual experiment (cf. Sec. II E), we found that errorful TC tokens differed from error-free /t/ tokens both in terms of their acoustics and in terms of how they were perceived, while errorful /k/ tokens did not differ either acoustically or perceptually from error-free /k/ (Fig. 6). For acoustics, the proximity *P* measure was used, which quantifies relative proximity of a given token to either /k/ or /t/ (described in Sec. II D). For perception, the d' score from Pouplier and Goldstein (2005) was used (described in Sec. II E). Independent sample *t*-tests showed that errorful TC tokens were significantly different from error-free /t/ tokens on both the d' score measure [Welch's $t(9.363)=3.145$, $p=0.011$] and on the acoustic measure [$t(18)=4.053$, $p=0.001$]. (TG tokens were not significantly different from error-free /t/, neither on the perceptual measure [$t(18)=1.42$, $p=0.173$] nor on the acoustic measure, at alpha level 0.012 corrected for multiple comparisons [$t(18)=2.284$, $p=0.035$].[6] Thus, the acoustic results match the previously reported perceptual results when performed on the same dataset (note that the same acoustic pattern is observed

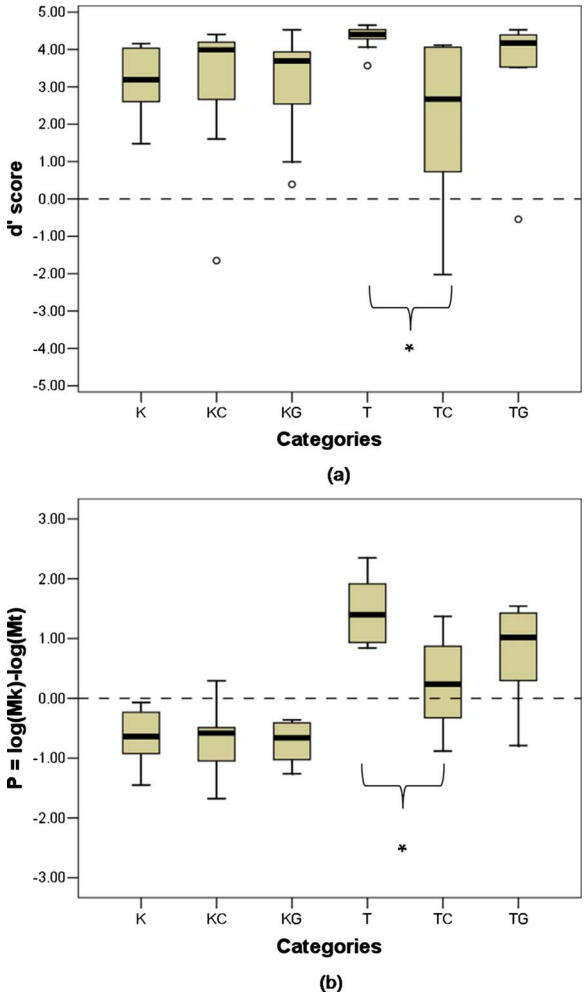


FIG. 6. (Color online) Boxplots showing the median (horizontal bar), interquartile range (boxes), error bars (vertical lines), and outliers (circles) for perceptual d' score (a) and acoustic measure P (b). Significant differences are indicated by *. Total number of tokens across categories for the data subset is $N=60$. Perceptual d' scores range from +4.65 representing maximum consistent correct identification to −4.65 representing consistent incorrect identification. Acoustic proximity indices range from close to /t/ (positive values) to close to /k/ (negative values).

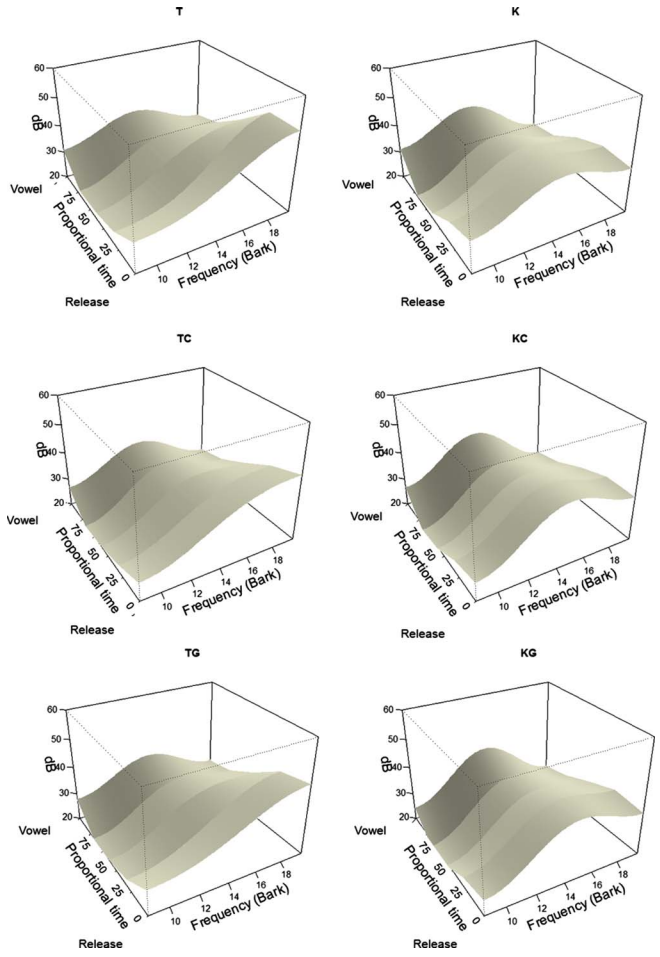


FIG. 7. (Color online) Ensemble-averaged, linearly time-normalized spectra of error-free and errorful /k/ and /t/. Time points are proportional, proceeding in 25% increments from burst onset (release) to vowel onset. The spectra were smoothed using the first three DCT-coefficients. The figures show information equivalent to the coefficients used for classification and acoustic distance calculation.

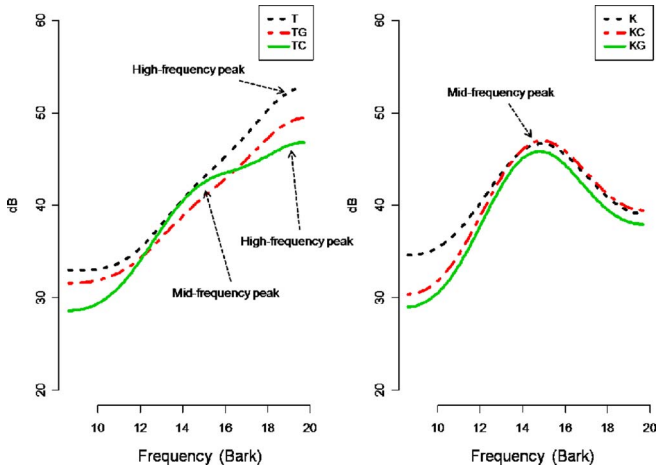


FIG. 8. (Color online) Ensemble-averaged spectra of error-free and errorful /k/ and /t/ at burst onset smoothed using three DCT-coefficients. Arrows indicate spectral energy peaks for spectra of error-free /t/ and /k/ and for /t/ produced with an intruding tongue dorsum of categorical magnitude (TC).

for the larger dataset as well). The asymmetry observed in perception is matched by an acoustic asymmetry, suggesting that the basis for the asymmetry observed in perception was not due to a bias at perceptual level (a bias of the perceptual system) but rather that the bias originated in the acoustic signal and hence the underlying articulatory pattern.

D. Spectral characteristics of errorful /k/ and /t/

The results so far show that while co-production of an intruding tongue tip had a negligible acoustic and perceptual influence on /k/, co-production of an intruding tongue dorsum with an intended tongue tip for /t/ made these tokens more ambiguous acoustically and perceptually. The question that arises in light of these results is how the characteristics of /k/ and /t/ spectra are affected by a second intruding constriction so that errorful intended /t/ tokens become acoustically ambiguous, while intended /k/ tokens remain acoustically unaffected. To address this question, we looked at the acoustics of error-free and errorful /k/ and /t/ between the burst onset and vowel onset in cepstrally smoothed running spectra (Fig. 7) derived from the same triplets of DCT-coefficients that were used to compress the burst to a point in a nine-dimensional space which had formed part of the acoustic classification and distance analyses discussed above.

The main acoustic characteristic of a typical /k/ is a spectral peak at mid-frequency (maintained over time) due to the long front cavity in front of the dorsal constriction, while a typical /t/ is mainly characterized by high spectral energy at high-frequency due to the short cavity in front of the constriction (cf. Blumstein and Stevens, 1979; Stevens, 1989; Stevens and Blumstein, 1978). Figure 7 shows that while the mid-frequency energy for /k/ is not altered by an intruding coronal constriction (that is, errorful /k/ spectra do not differ from error-free /k/), the high-frequency energy characteristic for /t/ is affected by an intruding dorsal constriction. More specifically, TC (and to a lesser extent TG) spectra have less energy at high-frequencies than error-free /t/.

However, since TC spectra have greater high-frequency energy than /k/ spectra, they are spectrally neither quite like a /k/ nor a /t/: that is, they are acoustically ambiguous between the error-free /k/ and /t/ stops. For a more detailed observation, we examined ensemble-averaged spectra at the burst onset (Fig. 8), i.e., the spectra at proportional time point zero in the running spectral displays in Fig. 7. The spectral shape at the burst onset was the same for both errorful and error-free /k/ tokens, with a spectral energy peak at mid-frequency typical for /k/. On the other hand, the spectral shape at the burst onset for errorful TC tokens differed from a typical /t/ spectrum.[7] The spectrum for TC (the continuous line in the left panel in Fig. 8) showed energy peaks both at mid-frequency (/k/-like) and at high-frequency (/t/-like), albeit both peaks were reduced in absolute amplitude compared to either typical /k/ or /t/. Crucially, the energy of the high-frequency peak was considerably lowered so that the mid- and high-frequency peaks were of roughly similar magnitudes, in contrast to the spectral shape for a typical /t/ which has an energy peak at high frequencies, resulting in /t/'s characteristic rising spectrum.

Qualitatively, the same general pattern was observed if the spectral shapes were analyzed taking into account vowel context (Figure 12 in the Appendix). Error-free /k/ and /t/

were, not unexpectedly, different as a function of vowel context: when followed by vowel /ɪ/, the amplitude at mid-frequency for /k/ and at high-frequency for /t/ was higher than in the /ɑ/-vowel context. However, the spectral shape for the errorful tokens was qualitatively similar to the overall observed pattern: errorful /k/ tokens were not different from error-free /k/ in either vowel context, while errorful /t/ tokens (TC, in particular) differed from error-free /k/ and /t/ tokens in both vowel contexts. In both /ɪ/- and /ɑ/-vowel contexts, TC tokens showed energy peaks of comparable amplitudes at both mid- and high-frequencies, with high-frequency amplitudes lower than those of error-free /t/ tokens.

Previous research (Blumstein and Stevens, 1980; Ohde and Stevens, 1983) has shown that lowering the spectral energy at high-frequency results in tokens less likely to be identified as /t/. It is therefore plausible to suggest that the lower energy at high-frequency observed for our TC tokens, combined with an almost equal energy peak at mid-frequency, is the factor responsible for these tokens' acoustic ambiguity both in our classification data and in Pouplier and Goldstein's (2005) perceptual results. The spectral shape for TG was closer to that of T than was TC (Fig. 8): TG's closer proximity to T also matched the classification and perceptual analyses.

Therefore, intrusion of a second constriction affects the spectral properties of the burst asymmetrically: whereas a tongue tip intrusion during the production of an intended /k/ has a marginal effect on the spectrum, a tongue dorsum intrusion during an intended /t/ is accompanied by a change in spectral shape. Recall that the intruding constrictions are weaker than the intended ones both spatially and temporally (cf. Pouplier, 2003; Pouplier and Goldstein, in press), so although both errorful /t/ and /k/ involve the same articulatory gestures, they differ in terms of which gesture is stronger. As shown in Figs. 7 and 8, the weaker coronal constriction during intended /k/ does not affect the spectral shape of the burst. The weaker dorsal constriction during intended /t/ on the other hand affects the spectrum by lowering the spectral energy at high-frequency enough to result in a qualitative change in spectral shape (from a rising spectrum typical for /t/ to a flatter one, with comparable mid- and high-frequency peaks), thereby making TC spectra ambiguous between a /k/ and a /t/. In this sense, it can be stated that the intruding tongue dorsum during an intended /t/ has a qualitative effect on the spectrum, whereas the effect of the intruding tongue tip is marginal.

E. Relationship between acoustics and articulatory timing

Recall that the articulatory measure used for classifying tokens into errorful or error-free was maximal vertical position of the intended and intruding gestures (cf. Sec. II A). This measure provides no insight into how the two gestures are timed relative to each other. Thus, while both error /k/ and /t/ were produced with an intended and intruding gesture, these co-produced gestures might have been timed differently for /t/ and /k/ tokens, and this timing difference may

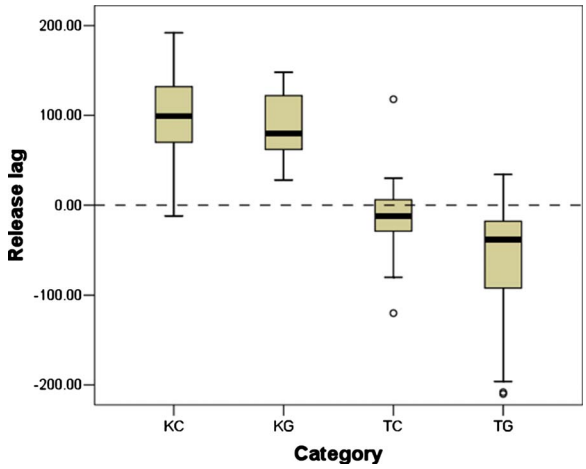


FIG. 9. (Color online) Boxplots showing the median (horizontal bar), interquartile range (boxes), error bars (vertical lines), and outliers (circles) for release lag, measuring relative timing between intended and intruding gestures during errorful /k/ and /t/. Positive values indicate that TD is released later than TT.

be the cause of the observed acoustic asymmetry between error /k/ and /t/. If, for instance, TD were always released last whether intended or intruding, this would lead to an asymmetry in acoustics between /t/ and /k/, assuming that the gesture released last contributes relatively more to the overall acoustic shape of the burst. For the timing measure, release lag from Pouplier and Goldstein's (in press) analysis was used, as described in Sec. II E. For the classification pattern, we used the results from the general classification analysis, as described in Sec. III A, pertaining to the tokens for which the timing measure was available (50 KC, 41 KG, 28 TC, and 34 TG).

Lag values between intended and intruding gestures at release are shown in Fig. 9. In general, median values indicate that most /k/-tokens had positive release lags, and most /t/-tokens negative lags. This means that for both intended /k/ and /t/, constriction of the intended gesture (TD and TT, respectively) was released last. However, the interquartile range for TC spans from negative to positive values, indicating a range from tokens with intended (TT) gesture released last to tokens with intruding (TD) gesture released last. Given this distribution, one may suspect that the asymmetry observed acoustically does not arise from the mere presence of an intruding gesture during /t/ but rather from how these intruding gestures are timed, in a way that is similar to the timing pattern observed for the opposite category. If many of the errorful /t/ tokens show a timing pattern specific to /k/ (i.e., TD released last, possibly as a result of TD being a slower articulator compared to TT), and these are specifically the tokens classified as /k/, then the relative timing between intruding and intended gestures may have been responsible for the acoustic asymmetry between intruded /k/ and /t/. To address this question, we look at the tokens' distribution as a function of both classification and timing pattern (Table VI).

Overall, as shown in Table VI, intended /k/ and /t/ exhibit distinct timing patterns, with the intended gesture (TD and TT, respectively) being released last in a majority of the tokens, confirming the result of Pouplier and Goldstein (in

TABLE VI. Distribution of token types as a function of classification results and relative timing patterns. Percentages in parentheses have been rounded.

	Articulatory timing pattern			
	TT released last		TD released last	
Articulatory category	Classified as /k/	Classified as /t/	Classified as /k/	Classified as /t/
KC	0	1 (2%)	41 (82%)	8 (16%)
KG	0	0	37 (90%)	4 (10%)
TC	9 (32%)	9 (32%)	5 (18%)	5 (18%)
TG	11 (32%)	17 (50%)	2 (6%)	4 (12%)

press) for our subset of the data. For errorful /k/, there were only a couple of tokens for which the intruding gesture was released last, amounting to 2% of the data. Taking the classification pattern into account, with the exception of one token, /k/ tokens classified acoustically as /t/ exhibited the same timing pattern as the majority of /k/ tokens, that is, a timing pattern with TD constriction released last. For /t/ tokens, the majority of both TC (64%) and TG (82%) show a pattern with TT released last, and not the timing pattern common for /k/ tokens (with TD released last). The classification pattern further shows that more errorful /t/ tokens were classified as /k/ with a /t/-like timing pattern (TT released last) than with a /k/-like timing pattern (TD released last), indicating that the classification of /t/ tokens as /k/ is not due to a /k/-like timing pattern. Furthermore, in the case of TC, for tokens with the same timing pattern (either TT or TD released last), half were classified as /k/ and half as /t/, indicating that the timing pattern is not a good predictor of the classification pattern.

Additionally, if timing were responsible for the observed acoustic asymmetry between the effect of an intruding gesture during /t/ vs during /k/, we would expect different timing patterns to result in different spectral shapes for the ambiguous category TC. Specifically, we have shown that the spectral shapes of /t/ tokens with categorical intrusion error exhibit spectral properties of both /t/ and /k/, while errorful /k/ spectra are entirely /k/-like in shape (cf. Fig. 8). We have also seen that of all errorful categories, TC has the most tokens in which intruding gesture is released last (36%). If the gesture released last shapes the overall acoustic pattern, it may be that the shape observed for TC is an average between very /k/-like tokens, that is, tokens where TD is released last (hence the mid-frequency peak energy), and very /t/-like tokens, that is, tokens with TT released last (hence the high-frequency peak energy). If this is the case, we expect the spectral shape for TC tokens with TD released last to look /k/-like, and the spectral shape for TC tokens with TT released last to look /t/-like. Figure 10 shows the spectra of TC tokens with either release pattern. While the timing pattern shows some effect on the spectral shape of the TC tokens, overall the spectral shape for TC tokens is ambiguous between /t/ and /k/ for either timing pattern. The spectra for tokens with TT released last differ from both /k/ and /t/ spectral shapes by having both a mid-frequency and a high-frequency energy peak. The spectra of tokens with TD released last look a bit more /k/-like in that the mid-frequency

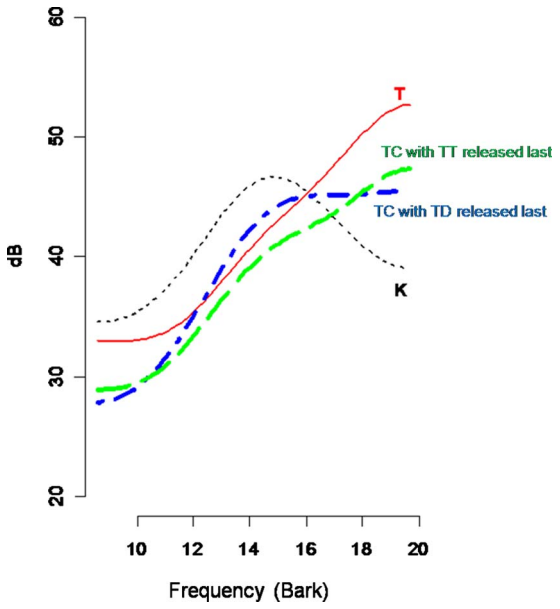


FIG. 10. (Color online) Ensemble-averaged spectra at burst onset smoothed using three DCT-coefficients of TC tokens with TD released last ($N=10$), TC tokens with TT released last ($N=18$), error-free /k/, and error-free /t/.

peak is more prominent; however, unlike /k/, this spectrum still includes an energy peak at high-frequency (roughly equal to the mid-frequency peak), absent from /k/ tokens. For both timing patterns, the energy peak at high-frequency is lower for TC tokens than for error-free /t/, and higher than for error-free /k/ tokens. The overall characteristic spectral shape of TC tokens can therefore not be (exclusively) due to the relative timing between the two gestures, but is instead due to the overall effect that a tongue dorsum constriction has on the energy of the spectrum at both mid- and high-frequencies.

IV. DISCUSSION AND CONCLUSION

Overall, the classification analysis has shown that an intruding tongue dorsum constriction with a /k/-like amplitude (i.e., TC tokens) results in /k/-like acoustics in about half of the intended /t/ tokens. On the other hand, intruding tongue tip constrictions during /k/, even when in the range of intended /t/ in terms of articulator height (KC tokens), result in /t/-like acoustics in only about 11% of the tokens. It has also been shown that the relative timing between intended and intruding gestures cannot explain this asymmetry: instead, the difference is due to different acoustic consequences of an intruding TT vs TD constriction. Overall, the presence of a TD constriction, whether intended or intruding, is likely to result in a /k/ classification for any given token due to the mid-frequency spectral energy peak and the lower high-frequency energies caused by a dorsal constriction. This implies that TD will affect the acoustics even when it is co-produced with a TT constriction with amplitudes in the range for an intended /t/, as is the case for KC and TC tokens. While KC was predominantly classified as /k/ by the SVM algorithm, the TC tokens were classified at chance level as either /t/ or /k/. The probability distribution of the classification algorithm (cf. Fig. 6) revealed that many of the TC

tokens were ambiguous between /k/ and /t/. In line with these results, the spectral analysis suggested that the ambiguity in the case of TC tokens was due to the characteristic shape of these tokens, showing both a mid-frequency energy peak (/k/-like) and a high-energy peak (/t/-like), although both these peaks were notably lower than those of error-free /k/ and /t/. For an intruding TT constriction, a different result became apparent: very few tokens with an intruding TT constriction were ambiguous between /t/ and /k/ in terms of their classification probabilities. The spectral shape of these tokens was indistinguishable from that of tokens produced without a tongue tip intrusion.

The acoustic cues for /k/ and /t/ are differently affected by the type of articulatory variability analyzed here. On the one hand, we observed relatively robust mid-frequency spectral energy in dorsal stops, demonstrated by the lack of difference in spectral shape between errorful and error-free /k/ tokens. By contrast the high-frequency spectral energy that is a major cue for coronal stops was much more variable, as shown by the quantitative and qualitative differences in the spectrum of errorful vs error-free /t/ tokens. Our finding shows, *contra* the prediction by Winitz *et al.* (1972), that articulatory variability of the type discussed here "smears" particularly the high-frequency energy of /t/ bursts, while it leaves mid-frequency energy, and hence /k/ bursts, unaffected. However, the effect is not uniform across our data: recall that the asymmetry between /t/ and /k/ only became evident for the categorical errors, not for the gradient ones. This means that articulator height is one of the main factors conditioning the results: this is because the intruding gesture had to be of a certain magnitude in order to affect sufficiently the acoustic classification results. However, for categorical errors, the asymmetric acoustic pattern for /t/ and /k/ showed that articulator height alone is insufficient to predict the results: unlike for the gradient errors, the location of the constriction becomes relevant. That is, if two categorical constrictions (one intended and one errorful) are present in the vocal tract during the same time interval, it is the constriction more posterior in the vocal tract that will dominate the acoustics. Surprisingly, this effect is observed irrespective of the timing of the release of the constrictions and irrespective of which one of the two constrictions was intended or intruding.

Our results suggest that the perceptual asymmetry observed by Pouplier and Goldstein (2005) is matched by an acoustic asymmetry between the robustness of dorsal stops but vulnerability of coronal stops to articulatory variability. This finding also suggests that coronals' perceptual vulnerability in the context of articulatory variability observed elsewhere (in consonant clusters, for example, cf. Byrd, 1992; Chen, 2003; Kochetov and So, 2007) is likely the result of their acoustic sensitivity to articulatory variability, rather than the result of a bias exclusively at the perceptual/auditory level. The results make predictions about when articulatory variability of the type described here is more likely to be perceived, possibly as a speech error. If, as suggested by the analyses reported in this paper, co-productions of an intruding TD constriction on an intended TT constriction for /t/ have more robust acoustic consequences than co-productions of intruding TT with an intended TD for /k/, then the expectation is that on the basis of acoustic information alone more variability, and hence more potential errors, during /t/ can be detected than errors during /k/. This may explain why, for instance, /t/ substitution by /k/ is reported more frequently than /k/ substitution by /t/ in studies that rely on auditory perception for error detection (Stemberger, 1991) [note that at least some of these transcribed "substitutions" may actually be co-production errors, cf. Boucher (1994) and Wood (1997)]. Our acoustic results suggest that rather than reflecting an error distribution pattern, the observed asymmetries between /t/ and /k/ in transcribed speech corpora (Stemberger, 1991) and in error perception (Pouplier and Goldstein, 2005) have their basis at the acoustic level. Both errorful /t/ and /k/ are equally distributed at the articulatory level—intruding gestures are as likely during intended /t/ and /k/. However, since errors during intended /t/ are more reliably cued acoustically, they are bound to be more readily detected auditorily. On the other hand, if errors during intended /k/ are not reflected in the tokens' acoustics, errorful /k/ productions are much more likely to be missed and therefore to be underrepresented in transcriptions. The theoretical implication is that if the transcription asymmetry does not reflect a production asymmetry, but is instead explained by asymmetric acoustic consequences of intruding gestures at different constriction locations, then no theoretical apparatus (such as coronal underspecification, cf. Stemberger, 1991) is needed to account for the /k/-/t/ asymmetry in speech errors.

To summarize, the results presented in this paper showed that not all intruding gestures have detectable acoustic consequences. Thus, while an intruding tongue dorsum during /t/ results in tokens acoustically different from error-free /t/, an intruding tongue tip during /k/ often has no acoustic consequences, that is, these tokens overall remain similar to error-free /k/. The observed acoustic pattern is also similar to results from a previous perceptual experiment, and it explains those results as being grounded in acoustics. While /k/ tokens are more robust acoustically to co-production of both intended and intruding gestures, /t/ tokens are sensitive acoustically and perceptually to the presence of a TD intruding gesture. The examination of the timing pattern of the intruding gesture relative to the intended gesture showed that the timing pattern could not explain on its own the acoustic properties of the tokens. Rather the amplitude of the intruding gesture seems to determine the acoustic pattern, and specifically, the amplitude of TD seems to have an influence on acoustics regardless of whether it is an intended or intruding gesture.

The study also contributes to the general knowledge on the acoustics of /k/ and /t/. There is ample evidence that /k/ and /t/ are robustly discriminable on the basis of their acoustic properties (Blumstein and Stevens, 1979; Fant, 1960; Halle *et al.*, 1957; Jakobson *et al.*, 1963; Kewley-Port, 1983; Stevens, 1989; Stevens and Blumstein, 1978), and also that the acoustic properties of /k/ and /t/ show different degrees of sensitivity to vowel context: /t/ is more robust than /k/ to vowel coarticulation, as shown by the well-known context-dependent variation in velars (cf. Halle *et al.*, 1957) and by

TABLE VII. Number of tokens for speaker JX split by conditions vowel context, rate, stress, and phrase position.

		Articulatory category						
Condition	Level	K	T	KC	KG	TC	TG	Total
Vowel context	/ɪ/	57	46	23	27	12	41	206
	/ɑ/	125	110	29	17	17	45	343
Rate	Fast	43	48	36	15	21	27	190
	Medium	75	50	14	27	7	46	219
	Slow	64	58	2	2	1	13	140
Stress	Stressed	80	90	28	14	14	34	260
	Unstressed	102	66	24	30	15	52	289
Phrase position	Initial	90	68	30	20	15	43	266
	Final	92	88	22	24	14	43	283

the greater convergence of alveolars to a common locus frequency, even when the following vowel context varies (cf. for example, Kewley-Port, 1982; Lehiste and Peterson, 1961; Lindblom, 1963; Sussman *et al.*, 1991). The results presented here likewise suggest that the acoustic properties of /k/ and /t/ are differently affected by the type of articulatory variation documented here, namely, co-production of an intruding constriction along with the intended one (in a very general sense, this co-production could be viewed as coarticulation with another consonant). However, in contrast to vowel coarticulation, in the case of co-productions the acoustics of /k/ are more robust to this type of articulatory variation compared to the acoustics of /t/. These two patterns of results may seem surprising at first sight but are actually quite consistent with each other when considering that vowel coarticulation influences low-frequency energy, leaving the salient high-frequency energy for /t/ unaffected, but shifting the mid-frequency peak for /k/. In the case of co-productions, an intruding tongue dorsum constriction has a noticeable impact on the high-frequency energy for /t/, while an intruding tongue tip leaves the mid-frequency peak for /k/ unaffected. The results of our study have more general implications for the reliability of acoustic information in detecting articulatory variations during production of /k/ and /t/ and contribute to our knowledge about the relationship between different types of articulatory variability and acoustics.

ACKNOWLEDGMENTS

This work was funded by the German Research Council (Contract No. PO 1269/1-1) and by NIH (Contract No. RO1DC008780-01). The authors thank Aleksandra Bigaji, Sophie Sedlmaier, and Susanne Waltl for help with data analysis and Louis Goldstein, Tine Mooshammer, Elliot Saltzman, Mark Tiede, and two anonymous reviewers for their helpful comments.

APPENDIX

A description of the data for speaker JX split by conditions vowel context, rate, stress, and phrase position is shown in Table VII, and classification scores (%) for speaker JX, split by the same conditions are shown in Table VIII. Relative proximity patterns for the data split according to vowel context, rate, stress, and phrase position are illustrated in Fig. 11. Ensemble-averaged spectra of error-free and errorful /k/ and /t/ by vowel context are shown in Fig. 12. The results in the analyses by condition do not diverge qualitatively from the general pattern of results discussed in the main text.

TABLE VIII. Classification scores (%) for speaker JX, split by conditions vowel context, rate, stress, and phrase position. The numbers represent rounded percentages of tokens classified as /k/. Because for the slow rate there were very few tokens for some of the errorful categories, the classification pattern for the slow rate was not included.

		Articulatory category					
Condition	Level	K	T	KC	KG	TC	TG
Vowel context	/ɪ/	86	13	87	89	58	15
	/ɑ/	91	15	79	88	47	38
Rate	Fast	93	15	83	93	57	41
	Medium	89	18	79	89	43	22
Stress	Stressed	86	11	89	93	57	41
	Unstressed	92	20	75	87	47	17
Phrase position	Initial	89	21	90	95	47	26
	Final	89	10	77	83	57	28

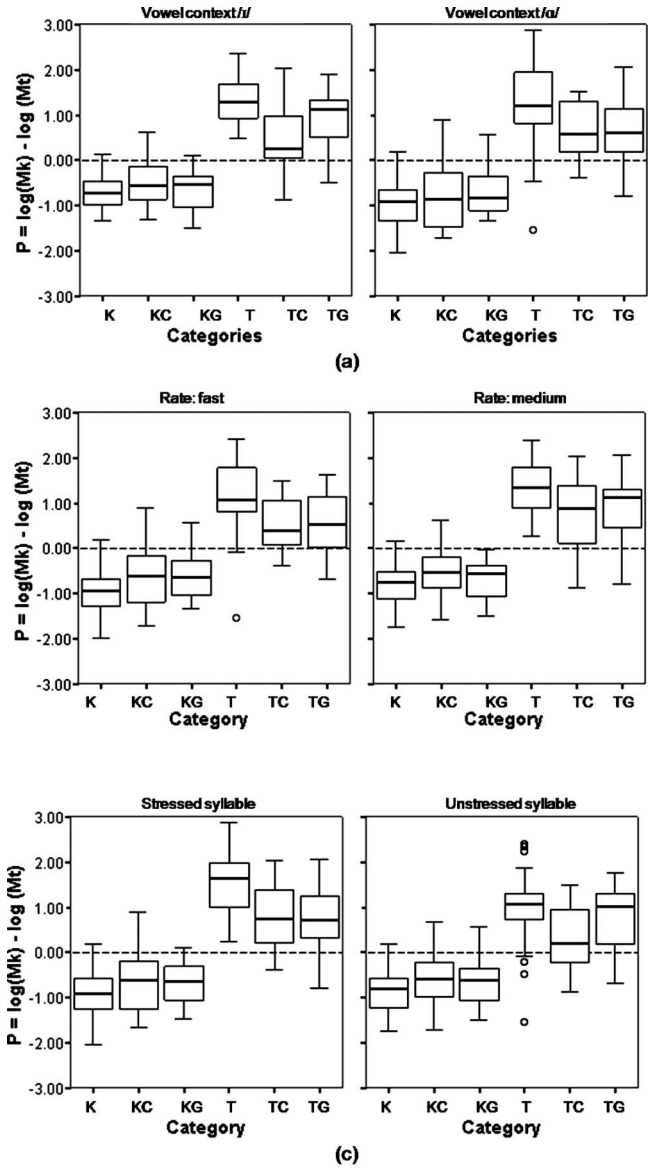


FIG. 11. Boxplots showing the median (thick horizontal bar), interquartile range (boxes), error bars (vertical lines), and outliers (circles) for the acoustic measure *P*, the token proximity to /k/ (negative values), or /t/ (positive values), split by conditions vowel context (a), rate (b), stress (c), and phrase position (d). Because for the slow rate there were very few tokens for some of the errorful categories, the slow rate was not included in this by-rate analysis.

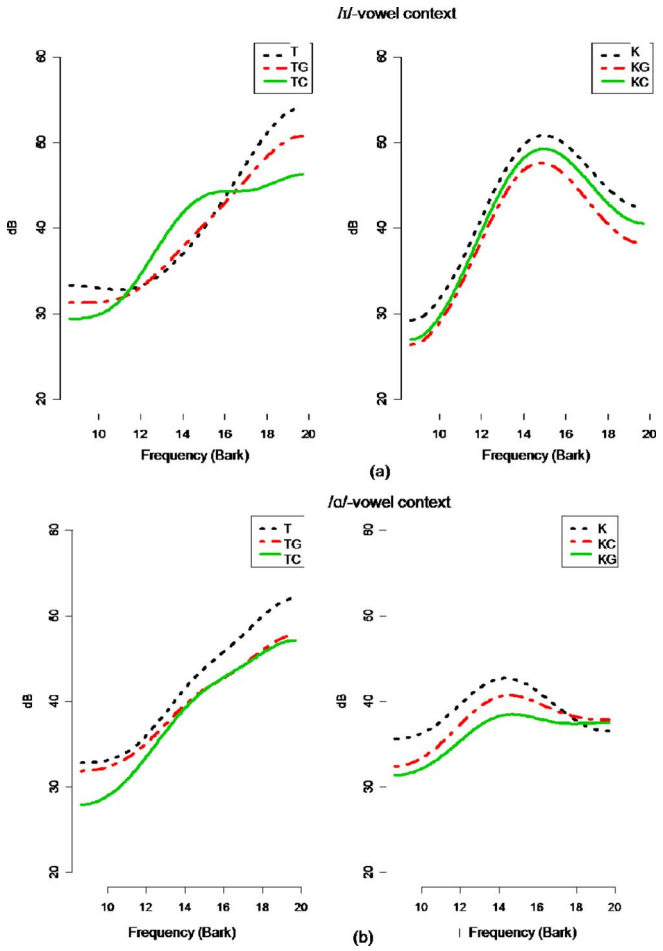


FIG. 12. (Color online) Ensemble-averaged spectra of error-free and errorful /k/ and /t/ by vowel context, smoothed using three DCT-coefficients. The figures show spectral shape at burst onset. Pictured in (a) are spectra of tokens produced in the /ɪ/-vowel context and in (b) are spectra of tokens produced in the /ɑ/-vowel context. At all time-points, TC spectral shape is different from T spectral shape, except at vowel onset, where /t/ and /k/ spectra are similar in shape, reflecting the properties of the vowel.

[1]Conflicting results were obtained in older studies (e.g., Kewley-Port *et al.*, 1983; Winitz *et al.*, 1972). Winitz *et al.* (1972) observed that on the basis of information available in the burst only, /t/ was better identified than /k/, and that including vowel transition information improved identification for /k/, but without resulting in a bias against correct identification of coronals (in this condition, identification of /k/ was equal or slightly better for /k/ in the context of vowels /a/ and /u/, but worse in the context of vowel /i/, similar to the pattern obtained by Kewley-Port *et al.*, 1983). Note, however, that even for the more recent studies (Hume *et al.*, 1999; Winters, 2000), the difference in perceptual salience between dorsals and coronals, while significant, was relatively small (only about 5% in favor of the dorsals).

[2]To be able to display tokens from all conditions in a single figure, articulator heights were normalized, in this figure only, relative to the range of the "gradient" and "categorical" error intervals since the ranges of the intervals differ as a function of the statistical properties of the respective control condition (see above). The range for the categorical error interval was defined symmetrically as 2 standard deviations in both directions from the relevant control means [cf. points B and C in Fig. 2(a)].

[3]Because errorful utterances are relatively few, splitting the data over these variables was not an option.

[4]For the purposes of the current paper, we differ from Pouplier and Goldstein (2005) in the calculation of perceptual score averages. Their focus was a between-subject analysis; hence they calculated category means for each of the 11 participants so that they obtained one perceptual score per participant per category. Our interest here is in perceptual scores for individual utterances, so we averaged for each token across participants' perceptual response, obtaining one d' perceptual score per individual token.

[5]Pouplier and Goldstein (in press) included additional gestural landmarks in their analysis. However, we only refer to relative timing at release since

the articulatory configuration at this point is presumably the one most relevant in shaping the acoustic signal.

[6]The perceptual responses on /t/ tokens with gradient intrusion error (TG) were significantly different from error-free /t/ in the between-subject analysis reported by Pouplier and Goldstein (2005); in our between-token analysis, however, the difference between T and TG was at trend level.

[7]The difference in spectrum shape between /t/ tokens with categorical errors and error-free /t/ is maintained for 75% of the burst.

Baayen, R. (**2008**). *Analyzing Linguistic Data: A Practical Introduction to Statistics Using R* (Cambridge University Press, New York).

Bennett, K., and Campbell, C. (**2000**). "Support vector machines: Hype or hallelujah?," SIGKDD Explorations **2**, 1–13.

Blumstein, S., and Stevens, K. N. (**1979**). "Acoustic invariance in speech production: Evidence from measurements of the spectral characteristics of stops," J. Acoust. Soc. Am. **66**, 1001–1017.

Blumstein, S., and Stevens, K. N. (**1980**). "Perceptual invariance and onset spectra for stop consonants in different vowel environments," J. Acoust. Soc. Am. **67**, 648–662.

Boersma, P., and Weenink, D. **2009**. *PRAAT: Doing phonetics by computer*, retrieved from http://www.praat.org/ (Last viewed March 30, 2009).

Boucher, V. J. (**1994**). "Alphabet-related biases in psycholinguistic enquiries: Considerations for direct theories of speech production and perception," J. Phonetics **22**, 1–18.

Byrd, D. (**1992**). "Perception of assimilation in consonant clusters: A gestural model," Phonetica **49**, 1–24.

Cassidy, S., and Harrington, J. (**2001**). "Multi-level annotation in the Emu speech database management system," Speech Commun. **33**, 61–77.

Chen, L. (**2003**). "The origins in overlap of place assimilation," in *Proceedings of the XXIIth West Coast Conference on Formal Linguistics (WCCFL 22)* (Cascadilla Press, Somerville, MA).

Clarkson, P., and Moreno, P. J. (**2000**). "On the use of support vector machines for phonetic classification," in *Proceedings of the International Conference on Acoustics, Speech and Signal Processing*, Vol. **2**, pp. 585–588.

Cortes, C., and Vapnik, V. (**1995**). "Support-vector network," Mach. Learn. **20**, 1–15.

Dell, G. (**1986**). "A spreading-activation theory of retrieval in sentence production," Psychol. Rev. **93**, 283–321.

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

Fant, G. (**1960**). *Acoustic Theory of Speech Production* (Mouton De Gruyter, The Hague).

Forrest, K., Weismer, G., Milenkovic, P., and Dougall, R. N. (**1988**). "Statistical analysis of word-initial voiceless obstruents: Preliminary data," J. Acoust. Soc. Am. **84**, 115–123.

Fromkin, V. A. (**1971**). "The non-anomalous nature of anomalous utterances," Language **47**, 27–52.

Fromkin, V. A. (**1973**). *Speech Errors as Linguistic Evidence* (Mouton De Gruyter, The Hague).

Goldrick, M., and Blumstein, S. (**2006**). "Cascading activation from phonological planning to articulatory processes: Evidence from tongue twisters," Lang. Cognit. Processes **21**, 649–683.

Goldstein, L., Pouplier, M., Chen, L., and Byrd, D. (**2007**). "Gestural action units slip in speech production errors," Cognition **103**, 386–412.

Halle, M., Hughes, W., and Radley, J. (**1957**). "Acoustic properties of stop consonants," J. Acoust. Soc. Am. **29**, 107–116.

Harrington, J., Kleber, F., and Reubold, U. (**2008**). "Compensation for coarticulation, /u/-fronting, and sound change in standard southern British: An acoustic and perceptual study," J. Acoust. Soc. Am. **123**, 2825–2835.

Hume, E., Johnson, K., Seo, M., Tserdanelis, G., and Winters, S. (**1999**). "A cross-linguistic study of stop place perception," in *Proceedings of the XIVth International Congress of Phonetic Sciences, San Francisco*, CA, pp. 2069–2072.

Jakobson, R., Fant, G., and Halle, M. (**1963**). *Preliminaries to Speech Analysis* (MIT Press, Cambridge, MA).

Kewley-Port, D. (**1982**). "Measurement of formant transitions in naturally produced stop consonant-vowel syllables," J. Acoust. Soc. Am. **72**, 379–389.

Kewley-Port, D. (**1983**). "Time-varying features as correlates of place of articulation in stop consonants," J. Acoust. Soc. Am. **73**, 322–335.

Kewley-Port, D., and Luce, P. A. (**1984**). "Time-varying features of initial stop consonants in auditory running spectra: A first report," Percept. Psychophys. **35**, 353–360.

Kewley-Port, D., Pisoni, D. B., and Studdert-Kennedy, M. (**1983**). "Perception of static and dynamic acoustic cues to place of articulation in initial stop consonants," J. Acoust. Soc. Am. **73**, 1779–1793.

Kochetov, A., and So, C. K. (**2007**). "Place assimilation and phonetic grounding: A cross-linguistic perceptual study," Phonology **24**, 397–432.

Lehiste, I., and Peterson, G. E. (**1961**). "Transitions, glides and diphthongs," J. Acoust. Soc. Am. **33**, 268–277.

Levelt, W. J. M. (**1989**). *Speaking: From Intention to Articulation* (MIT Press, Cambridge, MA).

Levelt, W. J. M., Roelofs, A., and Meyer, A. S. (**1999**). "A theory of lexical access in speech production," Behav. Brain Sci. **22**, 1–75.

Lindblom, B. (**1963**). "On vowel reduction," Report No. 29, The Royal Institute of Technology, Speech Transmission Laboratory, Stockholm, Sweden.

MacMillan, N. A., and Creelman, C. D. (**1991**). *Detection Theory: A User's guide* (Cambridge University Press, Cambridge, MA).

Mayer, D. (**2001**). "Support vector machines," R News **1**, 23–26.

McMillan, C., Corley, M., and Lickley, R. (**2009**). "Articulatory evidence for feedback and competition in speech production," Lang. Cognit. Processes **24**, 44–66.

Mowrey, R., and MacKay, I. (**1990**). "Phonological primitives: Electromyographic speech error evidence," J. Acoust. Soc. Am. **88**, 1299–1312.

Nossair, Z., and Zahorian, S. (**1991**). "Dynamical spectral features as acoustic correlates for initial stop consonants," J. Acoust. Soc. Am. **89**, 2978–2991.

Ohde, R. N., and Stevens, K. N. (**1983**). "Effect of burst amplitude on the perception of stop consonant place of articulation," J. Acoust. Soc. Am. **74**, 706–714.

Perkell, J., Cohen, M., Svirsky, M., Matthies, M., Garabieta, I., and Jackson, M. (**1992**). "Electromagnetic midsagittalarticulometer (EMMA) systems for transducing speech articulatory movements," J. Acoust. Soc. Am. **92**, 3078–3096.

Pouplier, M. (**2003**). "Units of phonological encoding: Empirical evidence," Ph.D. thesis, Yale University, New Haven, CT.

Pouplier, M. (**2007**). "Tongue kinematics during utterances elicited with the SLIP technique," Health Serv Res. **50**, 311–341.

Pouplier, M. (**2008**). "The role of a coda consonant as error trigger in repetition tasks," J. Phonetics **36**, 114–140.

Pouplier, M., and Goldstein, L. (**2005**). "Asymmetries in the perception of speech production errors," J. Phonetics **33**, 47–75.

Pouplier, M., and Goldstein, L. "Intention in articulation: Articulatory timing in alternating consonant sequences and its implications for models of speech production," Language and Cognitive Processes (in press).

Pouplier, M., and Waltl, S. (**2008**). "Articulatory timing of coproduced gestures and its implications for models of speech production," in *Proceedings of the 8th International Seminar on Speech Production*, pp. 19–22.

Rapp, B., and Goldrick, M. (**2000**). "Discreteness and interactivity in spoken word production," Psychol. Rev. **107**, 460–499.

Shattuck-Hufnagel, S. (**1979**). "Speech errors as evidence for a serial-ordering mechanism in sentence production," *Sentence Processing: Psycholinguistic Studies Presented to Merrill Garrett* (Lawrence Erlbaum, Hillsdate, NJ), pp. 295–342.

Shattuck-Hufnagel, S. (**1983**). "Sublexical units and suprasegmental structure in speech production planning," *The Production of Speech* (Springer, New York), pp. 109–136.

Smits, R., ten Bosch, L., and Collier, R. (**1996**). "Evaluation of various sets of acoustic cues for the perception of prevocalic stop consonants. I. Perception experiment," J. Acoust. Soc. Am. **100**, 3852–3864.

Stemberger, J. P. (**1991**). "Apparent anti-frequency effects in language production: The addition bias and phonological underspecification," J. Mem. Lang. **30**, 161–185.

Stevens, K. N. (**1972**). "The quantal nature of speech: Evidence from articulatory-acoustic data," *Human Communication: A Unified View* (McGraw-Hill, New York), pp. 51–66.

Stevens, K. N. (**1989**). "On the quantal nature of speech," J. Phonetics **17**, 3–45.

Stevens, K. N., and Blumstein, S. (**1978**). "Invariant cues for place of articulation in stop consonants," J. Acoust. Soc. Am. **64**, 1358–1368.

Surprenant, A. M., and Goldstein, L. (**1998**). "The perception of speech gestures," J. Acoust. Soc. Am. **104**, 518–529.

Sussman, H. M., McCaffrey, H. A., and Matthews, S. A. (**1991**). "An investigation of locus equations as a source of relational invariance for stop place categorization," J. Acoust. Soc. Am. **90**, 1309–1325.

Toothaker, L. E. (**1993**). *Multiple Comparison Procedures* (Sage, Thousand Oaks, CA).
Traunmüller, H. (**1990**). "Analytical expressions for the tonotopic sensory scale," J. Acoust. Soc. Am. **88**, 97–100.
Watson, C. I., and Harrington, J. (**1999**). "Acoustic evidence for dynamic formant trajectories in Australian English vowels," J. Acoust. Soc. Am. **106**, 458–468.
Winitz, H., Scheib, M. E., and Reeds, J. A. (**1972**). "Identification of stops and vowels for the burst portion of /p, t, k/ isolated from conversational speech," J. Acoust. Soc. Am. **51**, 1309–1317.
Winters, S. (**2000**). "Turning phonology inside out: Testing the relevant salience of audio and visual cues for place of articulation," *Ohio State Working Papers in Linguistics* (Ohio State University, Columbus, OH), pp. 168–199.
Wood, S. (**1997**). "Electropalatographic study of speech sound errors in adults with acquired aphasia," Ph.D. thesis, Queen Margaret University College, Edinburgh, UK.

Catalan speakers' perception of word stress in unaccented contexts

Marta Ortega-Llebaria
Department of Spanish and Portuguese, University of Texas-Austin (UT), 1 University Station, Austin, Texas 78712

Maria del Mar Vanrell
Department of Translation and Language Sciences, University Pompeu Fabra, C/Roc Boronat, 138, 08018 Barcelona, Spain

Pilar Prieto
Department of Translation and Language Sciences, University Pompeu Fabra, C/Roc Boronat, 138, 08018 Barcelona, Spain and Catalan Institute of Research and Advanced Studies (ICREA), Passeig Lluís Companys, 23, 08010 Barcelona, Spain

(Received 26 October 2008; revised 5 October 2009; accepted 19 October 2009)

In unaccented contexts, formant frequency differences related to vowel reduction constitute a consistent cue to word stress in English, whereas in languages such as Spanish that have no systematic vowel reduction, stress perception is based on duration and intensity cues. This article examines the perception of word stress by speakers of Central Catalan, in which, due to its vowel reduction patterns, words either alternate stressed open vowels with unstressed mid-central vowels as in English or contain no vowel quality cues to stress, as in Spanish. Results show that Catalan listeners perceive stress based mainly on duration cues in both word types. Other cues pattern together with duration to make stress perception more robust. However, no single cue is absolutely necessary and trading effects compensate for a lack of differentiation in one dimension by changes in another dimension. In particular, speakers identify longer mid-central vowels as more stressed than shorter open vowels. These results and those obtained in other stress-accent languages provide cumulative evidence that word stress is perceived independently of pitch accents by relying on a set of cues with trading effects so that no single cue, including formant frequency differences related to vowel reduction, is absolutely necessary for stress perception.
 [DOI: 10.1121/1.3268506]

PACS number(s): 43.71.Es [AJ] Pages: 462–471

I. INTRODUCTION

A historically long line of studies on the acoustic correlates of stress has often led to seemingly contradictory results that prevented a consensus on the common factor that differentiates stressed from unstressed syllables, reflecting the complexity underlying the realization of word stress. Beckman and Edwards (1994) stated that the apparently conflicting results may be because stress cues were examined without controlling for sentence intonation, i.e., pitch accents. A second variable to control is the use that different languages make of vowel reduction in relation to stress (Campbell and Beckman, 1997). Thus, the aim of the present study is to contribute to the body of research on the acoustic realization of word stress by examining the use that speakers of Central Catalan make of duration, overall intensity, and spectral tilt cues to perceive word stress while controlling for the two confounding factors of pitch accents and the formant frequency differences related to vowel reduction.

Following studies of English stress (Bolinger, 1958; Pierrehumbert, 1980; Beckman, 1986; Beckman and Edwards, 1994; Campbell and Beckman, 1997; de Jong, 1991), we assume that stress is used to convey prominence at all levels of the prominence hierarchy, i.e., word stress, phrase stress, and sentence stress, and that this prominence is cumulative across levels. For example, when the sentence "I saw Laura" is spoken as an answer to the question "Who did you see?" (see Fig. 1), the syllable [ˈlɔ] in [ˈlɔɹə] is perceived as stressed because it contains the only open vowel in the word and because, at the sentence level, it bears a pitch accent. Both cues, vowel quality at the word level and pitch accents at the sentence level, reinforce each other to make stress perception more robust.

In stress-accent languages, such as English or Catalan, the presence of a pitch accent is signaled acoustically by the presence of a pitch movement on its landing site, i.e., the stressed syllable, causing co-variation between stress and pitch accents. However, not all syllables with word stress are accented in all discourse contexts. The presence or absence of a pitch accent depends on the larger prosodic structure in which the lexical item is found. Thus, as Beckman and Edwards (1994) pointed out, this account of English stress may help explain the often contradictory results of studies examining the phonetic correlates of stress since "previous phonetic studies have compared the intensities, durations, and F0 excursions in 'stressed' versus 'unstressed' syllables without controlling systematically for the levels of the stress hierarchy involved" [Beckman and Edwards, 1994, p. 17].

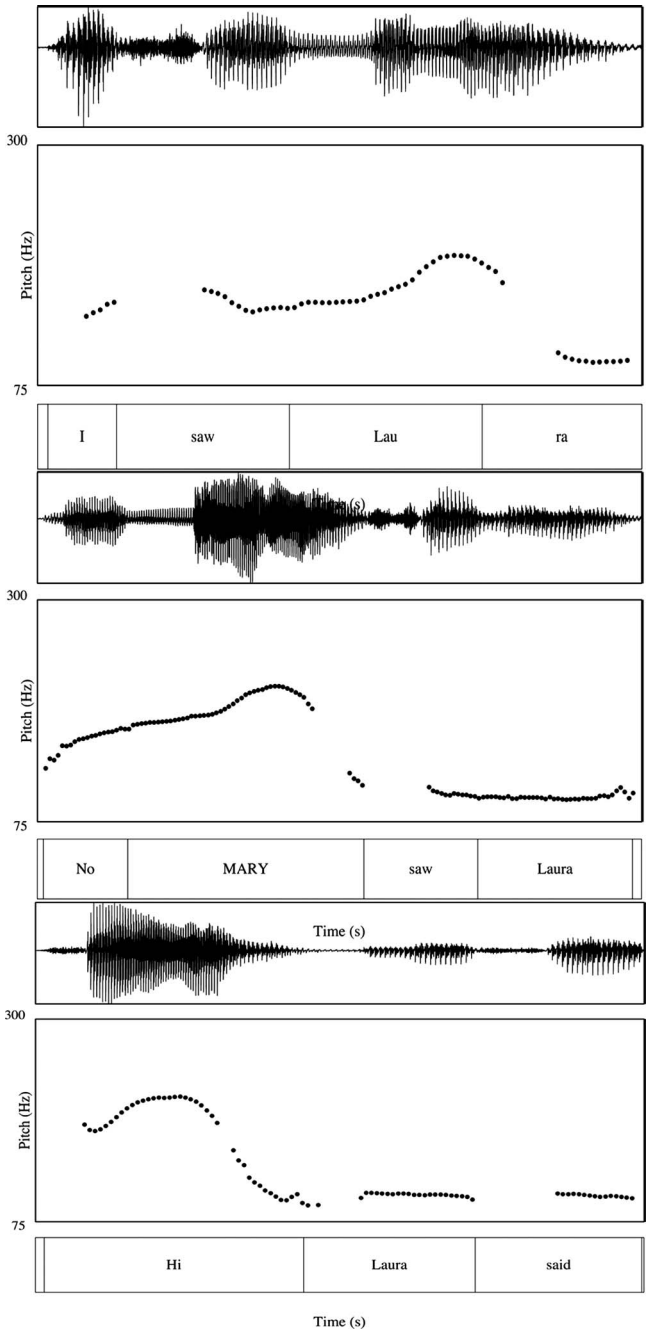


FIG. 1. Declarative sentences, post-focal contexts, and reporting clauses. The word "Laura" aligns with a pitch movement in the declarative sentence (top), while in post-focal contexts (center) and reporting clauses (bottom) "Laura" is produced with a flat *F*0 melody.

Some researchers, however, have controlled for levels of stress by restricting the contrast between stressed and unstressed syllables to the lower levels of the prominence hierarchy; that is, they examined cues to word stress only in unaccented contexts. These were created either by maintaining a constant *F*0 in synthesized stimuli (Fry, 1955, 1958; Turk and Sawusch, 1996), or by asking speakers to produce words in sentences whose intonation does not include pitch accents, such as post-focal contexts (Huss, 1978; Sluijter *et al.*, 1997) or reporting clauses (Ortega-Llebaria *et al.*, 2008). Post-focal contexts, such as "saw Laura" in the sentence "MARY saw Laura" when produced as an answer to "Did Anne see Laura?" (see Fig. 1), are produced with very reduced or no pitch movements, and they are placed just after the focal pitch accent, i.e., "MARY." Reporting clauses are used to report someone's speech, e.g., "He said," "She wrote," and "They shouted," and they are produced with a consistently flat pitch melody and a lower pitch register than the portion of direct speech they introduce (see Fig. 1).

In his pioneering studies on the perception of stress in English, Fry (1955, 1958) found that when listening to noun-verb word pairs that contrasted mainly in stress, i.e., subject noun–subject verb, English speakers relied more on duration cues than on intensity cues to perceive the stress contrast. Turk and Sawusch (1996) complemented Fry's (1955, 1958) results by showing that duration and intensity were perceived integrally, so that irrelevant variations in duration had an effect on the perception of loudness, and irrelevant variations in intensity had an effect on judgments of length. However, this effect was greater for duration, indicating that stress was effectively cued by either a combination of duration and intensity cues, or by duration cues alone, but not by intensity cues only.

The above results contrast with those of Huss (1978), who also examined minimal pairs that differed only in stress, such as import noun–import verb, embedded in post-focal phrases. Although English speakers produced small duration and intensity differences between stressed and unstressed syllables, they identified nouns versus verbs only at chance level. Huss (1978) concluded that in the absence of pitch accents and the formant frequency differences that accompany vowel reduction, duration and intensity cues were insufficient to cue word stress in English. However, these results may be related to the phrasing differences between the target sentences. Although a pause is likely to be inserted after "import" in sentences with a noun target, i.e., [The GERMANS' import] [sinks], a pause is placed before the target word in sentences with the verb target, i.e., [The GERMANS] [import sinks]. This pause causes the last syllable in the phrase to lengthen, making "port" in noun targets to become longer, which reduced the duration differences between nouns and verbs.

Sluijter *et al.* (1997) examined the perception of stress in Dutch by embedding [ˈnana] and [naˈna] in post-focal contexts, while controlling for phrasing effects. Similar to the English speakers in Fry's (1958) and Turk and Sawusch's (1996) experiments, Dutch speakers perceived stress by relying on duration and intensity cues. However, intensity differences related to word stress were perceived better with changes in spectral tilt, i.e., intensity differences between the higher and lower regions of the spectrum, than with changes in overall intensity, i.e., the mean intensity across all frequencies of the spectrum.

Ortega-Llebaria *et al.* (2008) embedded the words *mama* and *mimi* in reporting clauses to examine the perception of stress in Castilian Spanish, a language that, in contrast with Dutch and English, has no phonological vowel reduction. Spanish speakers perceived the stress differences between [ˈmama]-[maˈma] and [ˈmimi]-[miˈmi] by relying mainly on duration cues, and to a lesser extent, in overall intensity cues. Speakers did not use spectral tilt cues.

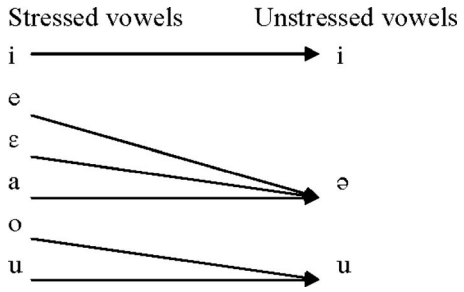


FIG. 2. Vowel inventory of Central Catalan. Lines indicate patterns of vowel reduction.

In general, the above studies provide cumulative evidence that word stress is perceived in the absence of pitch accents across languages by relying on duration. Stressed syllables are consistently longer than their unstressed counterparts. In contrast to duration, languages differ in their use of intensity cues; i.e., Spanish speakers use overall intensity while Dutch speakers rely on spectral tilt. However, these experiments limited their test materials to bi-syllabic words with open vowels. Although this constitutes a realistic strategy to stress perception in languages that have no vowel reduction, such as Spanish, in languages such as English or Dutch, where words alternate stressed open vowels with unstressed mid-central vowels, duration and intensity cues to word stress should also be examined in relation to vowel quality.

Campbell and Beckman (1997) highlighted the important role of vowel reduction in stress perception by suggesting that the different use that languages make of duration and spectral tilt cues to mark the stress contrast is a direct consequence of their differences in the degree of co-variation between stress and the formant frequency differences linked to vowel reduction. More specifically, they could not replicate Sluijter *et al.*'s (1997) results for Dutch in English, since unlike in Dutch, duration in English was not a consistent cue to stress and variations in spectral tilt correlated with pitch accents rather than stress proper. They argued that the different results obtained for Dutch and English are related "with differences between the two languages in the perceptibility of stress in the absence of accent or vowel reduction […] Dutch differs from English in having relatively fewer words in which unstressed syllables are reduced, particularly in word initial position" [Sluijter *et al.*, 1997, p. 70].

The formant frequency differences related to vowel reduction constitute a reliable cue to stress only in English, and, therefore, English speakers, unlike Dutch speakers, do not need to rely on other cues, such as duration or spectral tilt, to perceive the stress contrast. However, because these differences in stress and vowel reduction take place across languages, the different use that Dutch and English speakers make of duration and spectral tilt cues to stress could be related to other cross-language differences besides vowel reduction. Vowel reduction patterns in Central Catalan provide an appropriate context to examine different degrees of co-variation between stress and formant frequency differences between corresponding vowels within a single language. As shown in Fig. 2, Central Catalan allows seven vowels [i, e, ɛ, a, ɔ, o, u] to appear in stressed position and three vowels [i, ə, u] to appear in unstressed position (Wheeler, 2005; Herrick, 2003; Recasens, 1986; Recasens and Espinosa, 2006). In an unstressed position, the vowels /i/ and /u/ surface as [i] and [u], respectively (*vi* [ˈbi] "wine" > *vinet* [biˈnɛt] "wine.dim;" *suc* [ˈsuk] "juice" > *suquet* [suˈkɛt] "juice.dim") (Recasens, 1986, p. 131), the vowels /e, ɛ, a/ merge and surface as schwa [ə] (*peix* [ˈpeš] "fish" > *peixet* [pəˈšɛt] "fish.dim;" *sec* [ˈsek] "dry" > *sequet* [səˈkɛt] "dry.dim;" *sac* [ˈsak] "bag" > *saquet* [səˈkɛt] "bag.dim"), and the vowels /ɔ, o/ merge and surface as [u] (*poc* [ˈpɔk] "few" > *poquet* [puˈkɛt] "few.dim;" *boca* [ˈbokə] "mouth" > *boqueta* [buˈkɛtə] "mouth.dim"). Consequently, it is possible to have words in which vowel quality co-varies with stress, such as [ˈbokə] "mouth," [pəˈšɛt] "fish dim.," and [bəˈlentə] "courageous fem.," as well as words that contain no vowel quality cues to stress, i.e., [biˈžili] "watch out," [ˈlikit] "liquid," and [impulˈsiu] "impulsive."

Duration, overall intensity, and spectral correlates of word stress were measured in Catalan-Spanish cognates embedded in accented and in unaccented sentences in a pilot production experiment. Cognates consisted of four-syllable verbs that ended in [ˈminə] and [miˈna] in Catalan and in [ˈmino] and [miˈno] in Spanish. Measurements of target *–mina* and *–mino* endings showed that formant frequency differences between corresponding vowels took place only in Catalan [a], where *F*1 was 200 Hz higher than in [ə], its unstressed counterpart. No vowel quality differences in relation to stress or accent were found for Spanish [o] and [i], and for Catalan [i]. Moreover, no significant vowel quality differences were found between Catalan [i] and Spanish [i]. Speakers of both languages produced stressed vowels with durations 11 ms longer and overall intensities 1.7 dB greater than their unstressed counterparts in both accented and unaccented contexts. However, duration differences were significantly greater in Catalan [a] than in other vowels, suggesting that the [a]-[ə] alternation amplified the duration differences related to the stress contrast. No significant differences in spectral tilt were found in Catalan [i] and Spanish [o] and [i], the vowels that maintained the same quality across stress contexts. However, a gain of 4 dB differentiated the spectral tilt in stressed [a] from that in unstressed [ə], indicating that in our production data, spectral tilt correlated with formant frequency differences related to stress rather than stress proper. Spanish speakers' perception of word stress in unaccented contexts (Ortega-Llebaria *et al.*, 2008) confirmed that Spanish speakers perceived the stress contrast in [a] and [i] in the absence of pitch accents and vowel reduction patterns by relying on duration and overall intensity cues, but not on spectral tilt.

The goal of the present experiment is to examine how speakers of Central Catalan use duration, overall intensity, and spectral tilt cues to perceive word stress in unaccented contexts. In order to control for possible effects on stress of the formant frequency differences related to vowel reduction, these cues are examined in words with no vowel quality cues to stress and in words that contained open and mid-central vowels. Given results reported in the production literature, it

is of particular interest to examine how duration interacts with vowel reduction, and whether duration and spectral tilt differences to stress are perceived only in words that alternate open and mid-central vowels or also in words with no vowel quality cues to stress.

II. METHODOLOGY

A. Experimental design and statistics

In order to examine which cues and cue combinations Catalan speakers use to perceive lexical stress in unaccented contexts, Catalan subjects were asked to listen and identify the word *mamà* (or *mimí*), with stress on the last syllable, in a reporting clause. Reporting clauses are a particular clause type used to report someone's speech, (e.g., "He said," "She wrote," and "They shouted"), and they are produced with a flat pitch melody and a lower pitch register than the direct speech they introduce. They do not present any tonal movement, and in sentences up to ten syllables, there are no internal divisions or phrasing effects (for a detailed description of *F*0 in reporting sentences in Catalan, see Astruch, 2005). A native speaker of Central Catalan produced these items (Sec. II B), which were manipulated to create stimuli that combined various values of duration for the two syllables of *mama* (and *mimi*) in order to obtain a continuum that ranged from ratios typical of words with penultimate stress, i.e., [ˈmama] or [ˈmimi], to ratios of words with last syllable stress, i.e., [maˈma] or [miˈmi]. Similar continua were obtained for values of overall intensity and spectral tilt in *mama* and *mimi* words. Since in Central Catalan, [i] maintains similar vowel qualities across different stress and accent contexts while stressed open [a] alternates with an unstressed mid-central [ə] (Recasens, 1986; Recasens and Espinosa, 2006; Herrick, 2003), only *mama* was manipulated for vowel reduction (Sec. II C). Values of the duration continuum were crossed with those of overall intensity, spectral tilt, and the vowel reduction continua yielding grids typical of an expanded factorial design, i.e., the "duration × overall intensity" grid consisted of the 25 stimuli obtained by crossing the 5 levels of the duration continuum with the 5 levels of the overall intensity continuum (see, for instance, Massaro *et al.*, 1993; Chen and Massaro, 2004). Stimuli from each grid were presented in an identification task. The proportions of *mamà* (or *mimí*) identification answers were subjected to logistical regression analyses (GENMOD model in SAS statistics package) in order to obtain GEE parameter estimates and odd ratios for each cue to stress within each identification task. Significant interactions were further analyzed by calculating the regression equation within each level of duration, spectral tilt, and vowel reduction.

B. Recordings

A 41 year-old female native speaker of Central Catalan was recorded saying the sentence *Hola – saluda la mama contenta* "Hi – greets mom happily," where the reporting clause *saluda la mama contenta* was consistently pronounced with a flat pitch melody. The target word [ˈmamə] was replaced with [məˈma], [miˈmi], and [ˈmimi], yielding a total of 60 sentences (4 target words × 15 repetitions).

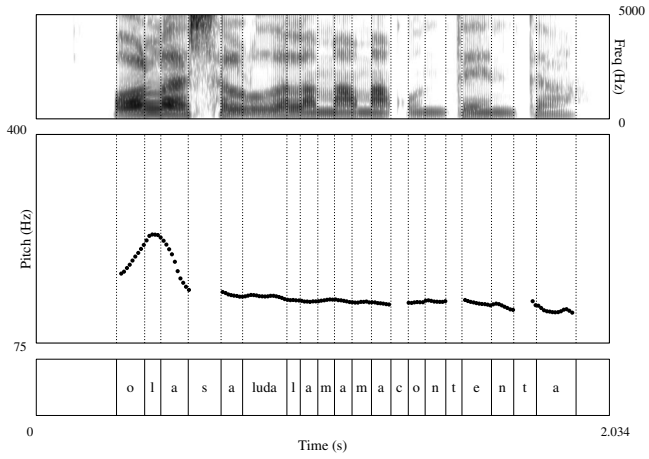


FIG. 3. Spectrogram, pitch track, and labels of the sentence *Hola – Saluda la mama contenta* "Hi – greets mom happily." Notice the pitch accent on "Hola" and the flat *F*0 melody for the reporting clause.

Measurements of duration, intensity, and spectral tilt were made on all target words, and the sentences containing the [ˈmamə] and [ˈmimi] items with values closest to the average were selected for further manipulation.

C. Materials

1. Duration and intensity

In order to examine the role of duration and intensity in the perception of primary stress, we controlled for formant frequency differences linked to vowel reduction. Recall that speakers from Central Catalan alternate stressed open [a] with unstressed mid-central [ə]. Thus, in order to avoid vowel reduction in the target word [ˈmamə], the first stressed vowel was copied to the second syllable, yielding two vowels with identical spectral characteristics, i.e., [mama]. For consistency, the same manipulations were performed on *mimi*. Afterwards, the target words were inserted in the carrier sentence, and, when necessary, their pitch was manipulated with the PSOLA utilities in PRAAT (Boersma and Weenink, 2005) to ensure that they had a nearly flat pitch trajectory. *F*0 values at the beginning, midpoint, and end of the target vowels were 138.5, 137.6, and 136.01 Hz for the first vowel of *mama*; 137.8, 138.04, and 137.9 Hz for the second vowel; and 135.5, 135.6, 135.5 Hz and 134, 134.4, and 133.5 Hz for the first and second vowels of *mimi* (Fig. 3).

A total of three five step [ˈmama]-[maˈma] continua (and three [ˈmimi]-[miˈmi] continua) were created by separately manipulating the cues of duration, overall intensity, and spectral tilt. For each continuum, stimulus 1 had a syllable ratio typical of words with penultimate stress, i.e., [ˈmama]. This ratio decreased in stimulus 2, was close to 1 in stimulus 3, and increased again but in the opposite direction for stimuli 4 and 5, with stimulus 5 replicating the ratio of words with last syllable stress, i.e., [maˈma]. For example, in the overall intensity continuum, syllable 1 was 3 dB louder than syllable 2 in stimulus 1. This difference decreased to 1.5 dB in stimulus 2, and became close to 0 in stimulus 3. In stimulus 4 the second syllable was 1.5 dB louder than the first, and this difference increased to 3 dB in

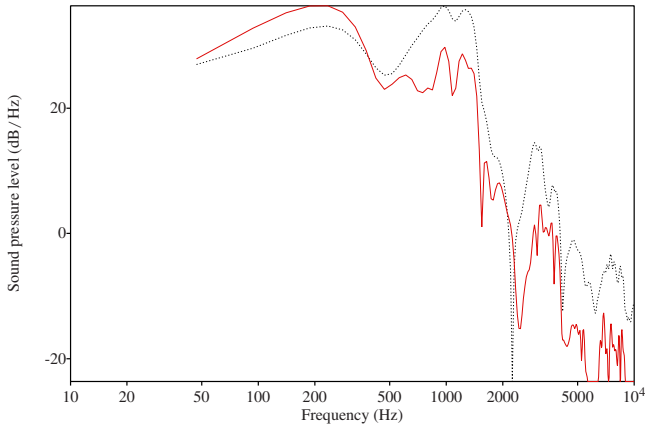


FIG. 4. (Color online) Spectral tilt of the vowels in *mama*. In stimulus 1 on the spectral tilt continuum, frequencies above 500 Hz have been amplified by 8 dB in the first vowel of *mama* producing a flatter spectral tilt in this vowel (dotted line) when compared to the spectral tilt of second vowel (compact line).

stimulus 5. These increments, which were achieved by multiplying all frequencies of the selected syllable by a constant factor with the multiply function in PRAAT (Boersma and Weenink, 2005), were modeled after our pilot production experiment, where Catalan stressed syllables in unaccented contexts were an average 1.57 dB (SD=1.81) louder than their unstressed counterparts.

In the spectral tilt continuum, amplitude differences between the lower and higher regions of the spectrum were achieved by manipulating the amplitude of the frequencies above 500 Hz in one of the *mama* (or *mimi*) vowels. Amplitude was increased to 4 dB by multiplying all frequencies above the 500 Hz range by a common factor using the finite impulse response filters with frequency bands spaced one-third of octave in the ADOBE AUDITION software. Then, the whole vowel was leveled for intensity with the scale intensity function in PRAAT (Boersma and Weenink, 2005) in order to maintain the same overall intensity as the other vowel in *mama* (or *mimi*). In stimulus 1, amplitude differences between frequencies below and above 500 Hz in the first vowel of *mama* (or *mimi*) were 8 dB greater than those in the second vowel (see Fig. 4). This difference was 4 dB in stimulus 2. Both vowels in stimulus 3 had identical spectral tilts, and in stimuli 4 and 5 the amplitude differences below and above 500 Hz were 4 and 8 dB greater in the second vowel than in the first vowel of *mama* (or *mimi*).

Boosting the amplitude of frequencies over 500 Hz was based on Sluijter and van Heuven's (1996) and Sluijter *et al.*'s (1997) experiments for Dutch. They found that intensity level differences between stressed and unstressed Dutch syllables were concentrated in the higher parts of the spectrum, whereas intensity differences below 500 Hz were negligible. In our stimuli, because formant values for [a] were above 500 Hz (*F*1: 869 Hz, *F*2: 1315 Hz, and *F*3: 3102 Hz) while for [i], *F*1 was below 500 Hz (*F*1: 324 Hz, *F*2: 2686 Hz, and *F*3: 3130 Hz), these manipulations boosted the amplitudes of *F*1, *F*2, and *F*3 with respect to *F*0 in [a], and *F*2 and *F*3 with respect to *F*0 and *F*1 in [i].

The value of the amplitude increments was fixed at 4 dB after measuring the Catalan speakers' productions in our pilot production experiment. On average, amplitude differences between frequencies below and above 500 Hz were 2.87 dB (SD=5.8) in stressed [a] and 7.01 dB (SD=5.65) in unstressed [ə], yielding an approximate gain of 4.14 dB for stressed [a] over its unstressed counterpart. In contrast, speakers produced remarkably similar spectral tilts for stressed and unstressed vowel [i] (M=6.56 dB and SD =5.73 for stressed [i], and M=6.79 dB and SD=5.51 for unstressed [i]).

Duration ratios between the two syllables of [mama] (or [mimi]) were modified by manipulating vowel duration while maintaining word duration of [mama] constant at 267 ms (or at 259 ms in [mimi]). Starting from stimulus 1, durations in milliseconds for each vowel in [mama] were 84-56, 77-63, 70-70, 63-77, and 56-84, respectively. In [mimi], durations in milliseconds were 77-49, 70-56, 63-63, 56-70, and 49-77. Durations were modified by cutting glottal cycles from one of the vowels in [mama] (or [mimi]) and adding cycles to the other vowel. Duration values were based on those obtained in our pilot production experiment. Stressed [a] was an average 82.5 ms long (SD=13.01) while unstressed [ə] was 61.9 ms (SD=10.15) yielding an average difference of 20.41 ms (SD=8.96). Stressed [i] was an average 58.56 ms long (SD=9.63) and unstressed [i] was 50.16 ms (SD=10.73) yielding a 6.39 ms difference (SD=8.26).

Finally, the five levels of the duration continuum were crossed with those of the overall intensity continuum, creating a 5×5 grid for [mama] and [mimi]. For example, when crossing overall intensity 1 with duration 1 of the continua, the resulting [mama] stimulus had a first vowel 84 ms long, a second vowel 56 ms long, and the first vowel was 3 dB higher than the second vowel. Similarly, the five levels of

TABLE I. Summary of identification tasks. There were five identification tasks, one per grid. Each grid contained the *mama* (or *mimi*) stimuli resulting from crossing different continua.

Crossed continua	Stimuli	No. of repetitions per stimulus	Subjects
Five step duration×five step overall intensity	25 mama	7	Group A=10 subjects
	25 mimi	7	Group B=10 subjects
Five step duration×five step spectral tilt	25 mama	7	Group A=10 subjects
	25 mimi	7	Group B=10 subjects
Seven step duration×two step vowel quality	14 mama	10	Group A+group B=20 subjects

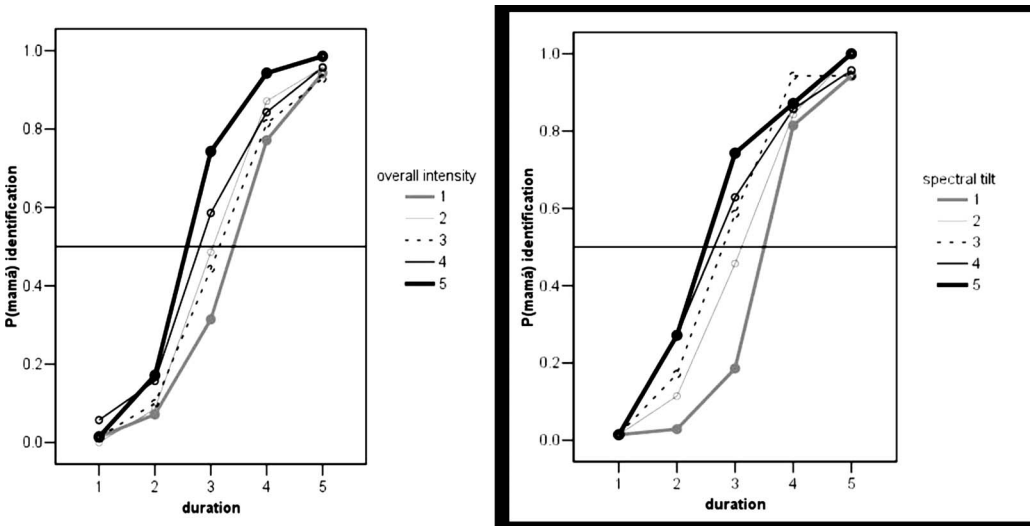


FIG. 5. Proportion of *mamà* responses for the 25 stimuli resulting from crossing the 5 levels of the duration continuum with the 5 levels of the overall intensity continuum in (a) [or the 5 levels of duration continuum with the 5 levels of spectral tilt continuum in (b)]. Duration and intensity ratios [or duration and spectral tilt ratios in (b)] reflect paroxytone stress in stimulus 1, i.e., [ˈmama], and oxytone stress in stimulus 5, i.e., [maˈma].

duration continuum were also crossed with the five of spectral tilt. A summary of the crossed continua is depicted in Table I.

2. Duration and vowel quality

In order to examine the effect of duration and vowel quality on the perception of primary stress, one [ˈməma] token and one [məˈma] token were selected for further manipulation. We made sure that the formant values for the stressed vowel in [ˈməma] showed an open quality (*F*1: 869 Hz, *F*2: 1315 Hz, and *F*3: 3102 Hz) while the values for the unstressed vowel in the second syllable were typical of the mid-central schwa (*F*1: 657 Hz, *F*2: 1237 Hz, and *F*3: 2944 Hz). In [məˈma], the vowel quality patterns reversed, yielding schwa values for the first vowel (*F*1: 668 Hz, *F*2: 1184 Hz, and *F*3: 2981 Hz) and an open quality for the second vowel (*F*1: 832 Hz, *F*2: 1486 Hz, and *F*3: 3040 Hz).

The vowel duration of each of these two tokens was manipulated to create two seven step continua while duration of [ˈməma] and [məˈma] remained constant at 235 ms. These manipulations aimed to have stimuli with short open [a] and long mid-central [ə] in addition to the stimuli with long open [a] and short mid-central [ə]. Starting from stimulus 1, the durations in milliseconds of the first and second vowels in [ˈməma] were 87-45, 80-52, 73-59, 66-66, 59-73, 52-80, and 45-87. A second continuum was created based on [məˈma]. Durations in milliseconds of the first and second vowels, starting from stimulus 1, were 45-85, 52-78, 59-71, 66-64, 73-57, 80-50, and 87-43.

In contrast to [a], Catalan [i] maintains the same vowel quality across stress contexts (Recasens, 1986, p. 131), which precludes Catalan listeners from identifying vowel nuclei with the quality of a schwa as possible realizations of Catalan [i]. Therefore, a duration × vowel quality identification task with [mimi] stimuli was not performed.

D. Subjects and identification tasks

20 native speakers of Central Catalan participated in the study. Their ages ranged from 19 to 41 years old, and they were born in areas where Central Catalan was spoken or had lived there for most of their lives. Catalan was the language they spoke with their families and friends, and they considered it their mother tongue. They had learned Spanish as a second language later in school, starting from age 6. None of them reported having any speech or hearing problems.

A group of ten subjects (group A in Table I) was instructed that they would hear the words *mamà* "mother," which is spoken with stress on the last syllable, or *mama* "mommy," which is pronounced with penultimate stress. Catalan orthographic conventions distinguish these two words by placing a stress mark in the stressed vowel of [məˈma], e.g., *mamà*. Subjects were asked to press the space bar in a keyboard as soon as they heard the word *mamà* in the sentence *Hola – saluda la ___ contenta* "Hi –greets _____ happily" over headphones. They listened to the randomized 175 sentences of the "duration × overall intensity" condition (25 stimuli of grid 1 × 7 repetitions) in 7 blocks of 25 stimuli with an inter stimuli interval (ISI) of 1500 ms and a 10-s pause between blocks. After a longer rest, they then listened to the 175 sentences of the "duration × spectral tilt" condition (25 stimuli of grid 3 × 7 repetitions). Presentation orders between the two tasks were counterbalanced across subjects. The remaining ten speakers (group B in Table I) received the same instructions that subjects in group A performed analogous tasks with the [ˈmimi]-[miˈmi] stimuli and spelled *mimi* and *mimí* in Catalan orthography. Finally, all 20 subjects (groups A and B) listened to the 10 repetitions of the 14 "duration × vowel quality" stimuli.

III. RESULTS

A. Duration, overall intensity, and spectral tilt in *mama*

The graphs in Fig. 5 illustrate the proportions of *mamà* answers for the 25 stimuli resulting from crossing the 5 steps of the duration with the 5 steps of the overall intensity continua [Fig. 5(a)], or with the 5 steps of the spectral tilt continuum [Fig. 5(b)]. Both graphs are similar in that the slopes

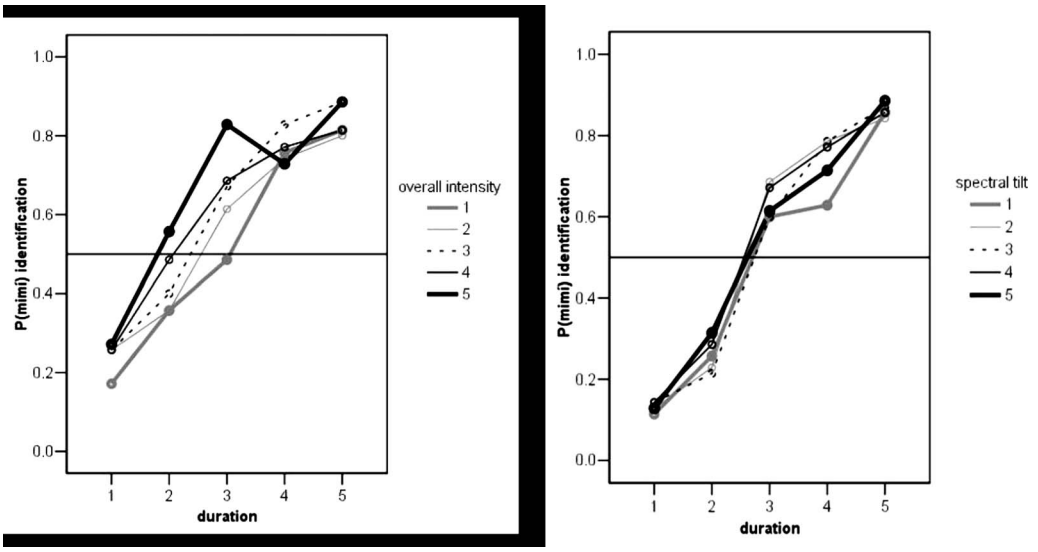


FIG. 6. Proportions of *mimí* responses for the 25 stimuli resulting from crossing the 5 levels of the duration continuum with the 5 levels of the overall intensity continuum in (a) [or the 5 levels of duration continuum with the 5 levels of spectral tilt continuum in (b)]. Duration and intensity ratios [or duration and spectral tilt ratios in (b)] reflect paroxytone stress in stimulus 1, i.e., [′mimi], and oxytone stress in stimulus 5, i.e., [mi′mi].

of the functions indicate that duration has a strong effect on the perception of primary stress: For each intensity level [measured as overall intensity in Fig. 5(a) or as spectral tilt changes in Fig. 5(b)], the [ma′ma] responses increase along the duration continuum regardless of whether they are in a competing or enhancing relationship with intensity. A second similarity between these two graphs is that the spread among curves, especially in duration stimuli 2 and 3, indicates that intensity also has an effect on the perception of stress.

Logistic regression analyses corroborate these visual trends. Analysis of GEE parameter estimates show a significant effect of duration ($Z=-8.97$, $p<0.0001$) and overall intensity ($Z=-2.36$, $p=0.018$) in Fig. 5(a), and of duration ($Z=-11.83$, $p<0.0001$), spectral tilt ($Z=-4.33$, $p<0.0001$), and duration × spectral tilt ($Z=2.01$, $p=0.044$) in Fig. 5(b), confirming that Catalan speakers use all the cues present in the stimuli to make stress judgments. Moreover, odd ratios of parameter estimates give us a reasonable estimate of cues' strength because the duration and intensity values used in this perception experiment approximate those obtained in our pilot production experiment, which averaged measurements on 1200 vowels produced by 20 speakers. They indicate that listeners rely more strongly on duration than on intensity cues. As duration increases by one whole unit in Fig. 5(a), listeners are 6.17 times more likely to perceive *mamà*, while increments in overall intensity show only a 1.34 increase in likelihood. Likewise, at each duration increment in Fig. 5(b), listeners are 8.39 times more likely to perceive *mamà*, but spectral tilt increments yield only a 1.9 likelihood increase.

The significant interaction between duration and spectral tilt was further analyzed. By investigating spectral tilt at each level of duration, it was found that spectral tilt has a significant effect only for durations 2 and 3 ($Z=-4.34$ and $Z=-6.67$ at $p<0.0001$; non-significant effect for durations 1, 4, and 5 scored $Z=-0.09$, $Z=-1.07$, and $Z=-2.1$ at $p>0.05$), in which both syllables of *mama* have similar or identical durations, showing that listeners rely on spectral tilt cues especially in those stimuli in which duration cues to stress are ambiguous. In contrast, when duration is examined at each level of spectral tilt, duration cues have a significant effect across the five levels of spectral tilt (starting from level 1, $Z=-8.03$, $Z=-15.69$, $Z=-7.76$, $Z=-13.42$, and $Z=-9.00$, with $p<0.0001$). Thus, Catalan speakers consistently rely on duration cues to perceive stress, not only for stimuli where spectral cues to stress become ambiguous, but also in those where spectral tilt is in either an enhancing or a competing relationship with duration.

The wider spread of the overall intensity curves for duration 3 in Fig. 5(a) is also examined by exploring the effect of overall intensity at each level of duration. Overall intensity has a significant effect only for duration 3 ($Z=-4.19$, $p<0.0001$; non-significant results for durations 1, 2, 4, and 5: $Z=-1.26$, $Z=-3.3$, $Z=-2.39$, and $Z=-1.46$ at $p>0.05$), indicating that listeners' use of overall intensity cues becomes evident when duration cues are ambiguous in the stimuli.

In summary, listeners rely heavily on duration to identify word stress, but the addition of overall intensity or spectral tilt as enhancing cues results in a shift of the response profile toward the alternative *mamà*, showing an additive relationship of duration with overall intensity and spectral tilt cues. However, listeners still perceived stress consistently in stimuli that contained duration cues alone, and also in those that contained spectral tilt cues alone, or overall intensity alone.

B. Duration, overall intensity, and spectral tilt in *mimi*

The graphs in Fig. 6 show the proportions of *mimí* answers for the 25 stimuli resulting from crossing the 5 steps of the duration with the 5 steps of the overall intensity continua [Fig. 6(a)], or with the 5 steps of the spectral tilt continua [Fig. 6(b)]. As in the graphs for *mama*, these graphs for *mimi* show that *mimí* responses increase along the duration continuum at each overall intensity [Fig. 6(a)] and spectral tilt

[Fig. 6(b)] curve, showing a consistent effect of duration cues in the perception of stress. However, unlike in *mama*, judgments at the end points are not unanimous possibly because of differences between subjects' groups. The spread of the curves is larger in Fig. 6(a) than in Fig. 6(b), indicating that overall intensity cues may have a stronger effect on the perception of stress in *mimi* than spectral tilt cues.

Analysis of GEE parameter estimates confirm the significant effect of duration ($Z=-4.60$, $p<0.0001$) and overall intensity ($Z=-2.93$, $p=0.003$) in Fig. 6(a), and of duration ($Z=-5.30$, $p<0.0001$) in Fig. 6(a). However, spectral tilt, as well as the interactions of duration with overall intensity and of duration with spectral tilt, yields non-significant results (spectral tilt: $Z=-0.29$, $p=0.77$; duration × intensity: $Z=1.17$, $p=0.24$; and duration × spectral tilt: $Z=-0.15$, $p=0.97$). Thus, Catalan listeners use duration and overall intensity cues to perceive lexical stress in *mimi*, but no significant effect is found for spectral tilt cues. Moreover, listeners rely more on duration than on overall intensity cues, because odd ratios of parameter estimates show that the probability of hearing *mimí* increases 2.17 times along each duration step, but it increases only 1.23 times at each overall intensity step. Finally, logistic regression analyses run separately at each duration level showing that overall intensity had a significant effect only for durations 2 and 3 ($Z=-3.44$, $p=0.0006$ and $Z=-3.76$, $p=0.0002$, respectively; non-significant effects at durations 1, 4, and 5: $Z=-1.53$, $Z=0.24$, and $Z=-2.28$ at $p<0.05$), while spectral tilt had non-significant effects at all levels at $p>0.05$ ($Z=-0.45$, $Z=-0.88$, $Z=-0.06$, $Z=-1.42$, and $Z=-0.58$).

When listening to the stress contrast in *mimi* stimuli, Catalan speakers relied strongly on duration, and to a lesser extent, overall intensity. Listeners still perceived the stress contrast in stimuli that contained duration cues alone or overall intensity cues alone. Spectral tilt had no significant effect.

C. Duration and vowel reduction in *mama*

Figure 7 shows the proportions of *mamá* answers for the 14 stimuli resulting from crossing the 7 step duration continuum with the [mamə] and [məma] items. Listeners identified *mamá* more often in the [məma] item (black line) than in the [mamə] item (gray line), showing that, in general, vowels with open qualities, i.e., [a], are perceived as stressed and vowels with mid-central qualities, i.e., [ə], are perceived as unstressed. However, the slopes of the curves showed that duration also had an effect on the perception of stress. Listeners perceived more *mamà* items when the second vowel was longer (stimuli 5–7) even in [mamə], where the longer vowel has a mid-central quality. In fact, listeners identified 58% of [mamə] items as *mamà* on duration 5, 68% on duration 6, and 66% on duration 7 while only 0.7% of [mamə] items were heard as *mamà* when [a] was longer than [ə]. They also heard fewer *mamà* items in stimuli 1–3 even in the [məma] continuum, where again [ə] is longer than [a]. For example, only 37% of [məma] items were identified as *mamà* in duration 1 against 97% in duration 7.

An analysis of GEE parameter estimates indicates that duration, vowel quality, and their interaction had a significant

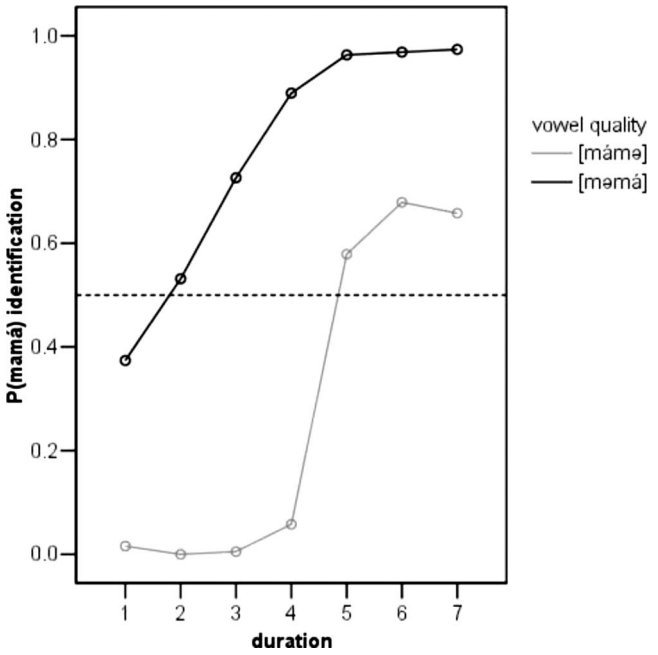


FIG. 7. Proportions of *mamà* responses for the 14 stimuli resulting from crossing the 7 levels of duration with the 2 levels of vowel quality in a [′mamə]-[mə′ma] continuum. Starting from stimulus 1, the durations in milliseconds of the first and second vowels in [′mamə] were 87-45, 80-52, 73-59, 66-66, 59-73, 52-80, and 45-87; and in [mə′ma] were 45-85, 52-78, 59-71, 66-64, 73-57, 80-50, and 87-43.

effect on the perception of stress ($Z=-11.48$, $p<0.0001$ for duration; $Z=-9.55$, $p<0.0001$ for vowel quality; and $Z=2.14$, $p=0.032$ for their interaction). The interaction was further analyzed by applying the logistic regression separately at each duration level and at each level of vowel reduction. Results showed a significant effect of vowel quality in each level of duration at $p<0.01$ ($Z=-5.83$, $Z=-5.97$, $Z=-6.12$, $Z=-8.95$, $Z=-7.68$, $Z=-7.78$, and $Z=-6.39$), and crucially, a significant effect of duration in the two levels of vowel quality ($Z=-9.43$ and $Z=-11.48$ at $p<0.01$), confirming, on the one hand, that the alternation of open and mid-central vowels had a strong effect on listeners' stress judgments. On the other hand, duration modified the perception of stress in [mamə] and [məma] items so that longer mid-central vowels were perceived as more stressed than shorter open vowels.

In summary, listeners consistently relied in both vowel quality and duration to perceive stress. Longer vowels with open qualities were perceived as more stressed than shorter vowels with mid-central qualities. However, in stimuli with conflicting cues, longer mid-central vowels were frequently perceived as more stressed than short open vowels, showing that the formant frequency differences related to vowel reduction did not inhibit duration cues to stress prominence.

IV. DISCUSSION

Speakers of Central Catalan perceived word stress in unaccented contexts in words that had no vowel quality cues to stress. More specifically, speakers used duration and overall intensity cues as effectively in strings of open vowels that were common in Catalan (i.e., [mimi]) as in those combinations of open vowels that were not possible Catalan words

(i.e., *[mama] instead of [mamə] or [məma]). Similar results have been obtained in languages such as English (Fry, 1955, 1958; Turk and Sawusch, 1996), Dutch (Sluijter *et al.*, 1997), and Spanish (Ortega-Llebaria *et al.*, 2008), showing that word stress is perceived cross-linguistically in the absence of pitch accents and vowel quality cues to stress by relying on duration and overall intensity cues. Thus, duration and intensity cues are used by speakers of languages that differ greatly in their degree of co-variation between vowel reduction and stress, that is, languages such as English, with a strong co-variation between stress and vowel quality as well as languages such as Spanish, where vowels maintain the same quality across stress contexts or Catalan, where only some stressed open vowels alternate with unstressed mid-central vowels.

Moreover, the present results show that Catalan speakers relied strongly on duration to perceive stress in words that alternated vowels with open and mid-central qualities as well. When the duration continuum was crossed with [mamə] and [məma] stimuli, Catalan speakers relied on both vowel quality and duration cues to perceive stress, showing that these cues were in an additive relationship. In stimuli with conflicting cues, listeners perceived longer [ə] as more stressed than shorter [a] in both [mamə] and [məma] items, revealing that not only vowels with open qualities are perceived as stressed, but that mid-central vowels of sufficient length also induce the perception of stress. Similar perception patterns have been found in European Portuguese, where Delgado Martins (1975) showed that duration affects stress perception not only in words such as *splicitu*, where there are no vowel quality cues to stress, but also in words such as *splicito* when the full [o] is shortened more than [i]. Interestingly, Fry (1964) showed a similar effect on English. He crossed a duration continuum with a vowel reduction continuum and found that English speakers relied more strongly on duration than on vowel quality cues to perceive word stress. He cautioned that his results were related to the different scales used in the two continua. Thus, if vowels with non-peripheral, mid-central qualities are *sufficiently* long, they induce stress perception not only in languages such as Catalan and European Portuguese, but also in languages such as English, where there is a strong co-variation between stress and vowel quality.

Therefore, the observations that longer mid-central vowels induce the perception of stress and that duration and intensity cues are sufficient to perceive stress in words made up of open vowels even in languages with a strong co-variation between stress and vowel quality do not support the idea that the degree of co-variation between vowel reduction and stress in a language determines the specific cues that speakers of that language use to perceive stress. Rather, they indicate that the phonetic realization of stress is made up of a cluster of cues with trading relationships.

Results from Catalan illustrate the trading relationships between duration cues and vowel quality, overall intensity, and spectral tilt cues. Duration cues are used in all stimuli, regardless of whether or not they contain vowel quality, overall intensity, or spectral tilt cues in a competing or in an enhancing relationship with duration. In contrast with duration, overall intensity and spectral tilt cues are most apparent in stimuli in which other cues to the stress contrast are ambiguous. These results for Catalan support those for English (Fry, 1955, 1958; Turk and Sawusch, 1996) and Spanish (Ortega-Llebaria *et al.*, 2008) in that duration and intensity cues are in an additive relationship, and, yet, there is an asymmetry in favor of duration. Variations in duration have a stronger effect than variations in intensity in stress perception. However, contrary to Turk and Sawusch's (1996) results, speakers of Catalan were able to perceive stress in stimuli that contained duration cues to stress, but also in those that only had overall intensity or spectral tilt cues to stress, indicating that Catalan speakers used any cue available in the speech signal.

In contrast with overall intensity cues, which were used to perceive stress in both [mama] and [mimi] stimuli, Catalan speakers used spectral tilt cues to perceive stress in [mama] but not in [mimi]. This asymmetry may relate to the spectral tilt manipulations used in this perception experiment, which boosted $F1$, $F2$, and $F3$ amplitudes in [a] but only $F2$ and $F3$ amplitudes in [i]. However, in Spanish (Ortega-Llebaria *et al.*, 2008), where vowels maintain the same quality across stress contexts, speakers did not use either spectral tilt cues to perceive stress in [mama] in spite of performing spectral tilt manipulations identical to those performed in this experiment for Catalan [mama]. Moreover, Catalan speakers in our pilot production experiment produced significant spectral tilt differences in relation to stress only when spectral tilt was measured in stressed [a] and unstressed [ə] without correcting for formant frequency differences. After correcting for formant frequency differences using Fulop *et al.*'s (1998) corrections, there were no clear differences in spectral tilt between these two vowels. Thus, the different use that Catalan and Spanish speakers make of spectral tilt cues to perceive stress in [mama] indicates that, in these languages, spectral tilt changes as a consequence of the formant frequency differences between corresponding vowels rather than in relation to the stress contrast, because only in Catalan stressed [a] reduces to an unstressed [ə] by lowering $F1$ by 200 Hz.

Spectral tilt results for Catalan and Spanish contrast with those obtained for Dutch (Sluijter and van Heuven, 1996; Sluijter *et al.*, 1997), where spectral tilt, together with duration, was a consistent cue to the stress contrast. However, Sluijter *et al.* (1997) examined the perception of spectral tilt in relation to stress only in the vowel [a], which in Dutch reduces to [•] by lowering $F1$ and $F2$ and raising $F3$ (Sluijter and van Heuven, 1996, p. 2481). Consequently, they could not disentangle the effects of vowel reduction in spectral tilt from those of stress.

There is no general agreement cross-linguistically that spectral tilt is a consistent cue to word stress. Although Sluijter (1995) showed that English behaves like Dutch with respect spectral tilt, Campbell and Beckman (1997) could not replicate Sluijter's (1995) results, and Kochanski *et al.* (2005) showed that variations in loudness and duration, but not in spectral tilt or $F0$, were reliable cues to word stress. Other perception studies could not show that spectral tilt is a cue to word stress in Spanish (Ortega-Llebaria *et al.*, 2008)

or to accent in Swedish (Heldner, 2003). In order to disentangle whether spectral tilt is a cue to stress rather than a consequence of the formant frequency differences related to vowel reduction, future research needs to examine vocal fold movement in a set of stressed and unstressed vowels that vary in their qualities. Thus, in Cutler's (2005) (p. 270) words "on this issue, the last word may not yet been spoken."

V. CONCLUSION

Results from Catalan agree with research in other languages that shows that listeners perceive word stress in the absence of pitch accents by relying on a cluster of cues in which duration plays an important role (see, for instance, Fry, 1955, 1958, 1964; Turk and Sawusch, 1996; and Kochanski *et al.*, 2005 for English; Sluijter *et al.*, 1997 for Dutch; and Ortega-Llebaria *et al.*, 2008 for Spanish). They also agree with Lieberman's (1960) finding for English that cues from this cluster trade off, so that in the absence of one cue, speakers rely on others. For example, our results show that in the absence of pitch accents, duration, and vowel reduction cues, listeners perceive stress by relying on overall intensity. Or in the absence of pitch accents, overall intensity, and vowel reduction, speakers rely on duration cues.

Trading relations between duration and vowel reduction were also examined in this experiment. In Catalan, as in English and other languages, vowel reduction implies duration and formant frequency differences between corresponding vowels so that vowels with open qualities are longer than their mid-central counterparts (Lindblom, 1963; Moon and Lindblom, 1994). In spite of this consistent co-variation between duration and vowel quality, Catalan speakers, such as Fry's (1964) English speakers, perceived longer mid-central vowels as more stressed than shorter open vowels, proving that duration differences in relation to stress are not perceived only through vowel quality differences linked to vowel reduction. Thus, cross-linguistic research provides cumulative evidence that word stress is perceived independently of pitch accents by relying on a set of cues with trading effects, so that no single cue, including vowel reduction, is absolutely necessary for the perception of stress.

ACKNOWLEDGMENTS

We thank Mary Beckman and Scott Myers for their insightful comments on earlier versions of this paper. Preliminary results were presented in PaPI 2007. We thank the audience for their comments. Heartfelt thanks to Jane Opie and Michael Mahometa for their valuable help with English and statistics. This research was funded by grants from the Ministerio de Educación y Ciencia-FEDER HUM2006-01758/PHYLLO and by a URAP grant from the College of Liberal Arts, UT-Austin.

Astruch, L. (**2005**). "The intonation of extra-sentential elements in Catalan and English," Ph.D. thesis, University of Cambridge, Cambridge.
Beckman, M. E. (**1986**). *Stress and Non-Stress Accent* (Foris, Dordrecht).
Beckman, M. E., and Edwards, J. (**1994**). "Articulatory evidence for differentiating stress categories," in *Phonological Structure and Phonetic Form. Papers in Laboratory Phonology III*, edited by P. A. Keating (Cambridge University Press, Cambridge), pp. 7–33.
Boersma, P., and Weenink, D. (**2005**). PRAAT: doing phonetics by computer (Version 4.3.01).
Bolinger, D. L. (**1958**). "A theory of pitch accent in English," Word **14**, 109–149.
Campbell, N., and Beckman, M. E. (**1997**). "Stress, prominence and spectral tilt," *Intonation: Theory, Models and Applications*, *Proceedings of an ESCA Workshop*, Athens, Greece, 18–20 September 1997, edited by A. Botinis, G. Kouroupetroglou, and G. Carayiannis (Department of Informatics, University of Athens, Greece), pp. 67–70.
Chen, T., and Massaro, D. (**2004**). "Mandarin speech perception by ear and eye follows a universal principle," Percept. Psychophys. **66**, 820–836.
Cutler, A. (**2005**). "Lexical stress," in *The Handbook of Speech Perception*, edited by D. B. Pisoni and R. E. Remez (Blackwell, Oxford), pp. 264–289.
de Jong, K. (**1991**). "The oral articulation of English stress accent," Ph.D. thesis, The Ohio State University, Columbus, OH.
Delgado Martins, M. R. (**1975**). "Perception of Portuguese stress," Proceedings of the VIII International Congress of Phonetic Sciences, Leeds, United Kingdom.
Fry, D. B. (**1955**). "Duration and intensity as physical correlates of linguistic stress," J. Acoust. Soc. Am. **27**, 765–768.
Fry, D. B. (**1958**). "Experiments in the perception of stress," Lang Speech **1**, 126–152.
Fry, D. B. (**1964**). "The dependence of stress judgments on vowel formant structure," Proceedings of the 5th International Congress in Phonetic Sciences, Münster, Germany, edited by S. Karger (Basel, New York), pp. 306–311.
Fulop, S. A., Kari, E., and Ladefoged, P. (**1998**). "An acoustic study of the tongue root contrast in Degema vowels," Phonetica **55**, 80–98.
Heldner, M. (**2003**). "On the reliability of overall intensity and spectral emphasis as acoustic correlates of focal accents in Swedish," J. Phonetics **31**, 39–62.
Herrick, D. (**2003**). "An acoustic analysis of phonological vowel reduction in six varieties of Catalan," Ph.D. thesis, University of California, Santa Cruz, CA.
Huss, V. (**1978**). "English word stress in the post-nuclear position," Phonetica **35**, 86–105.
Kochanski, G., Grabe, E., Coleman, J., and Rosner, B. (**2005**). "Loudness predicts prominence: Fundamental frequency lends little," J. Acoust. Soc. Am. **118**, 1038–1054.
Lieberman, P. (**1960**). "Some acoustic correlates of word stress in American English," J. Acoust. Soc. Am. **32**, 451–454.
Lindblom, B. (**1963**). "Spectrographic study of vowel reduction," J. Acoust. Soc. Am. **35**, 1773–1781.
Massaro, D., Cohen, M., Gesi, A., Heredia, M., and Tsuzaki, M. (**1993**). "Bimodal speech perception: An examination across languages," J. Phonetics **21**, 445–478.
Moon, S., and Lindblom, B. (**1994**). "Interaction between duration, context, and speaking style in English stressed vowels," J. Acoust. Soc. Am. **96**, 40–55.
Ortega-Llebaria, M., Prieto, P., and Vanrell, M. M. (**2008**). "Perceptual evidence for direct acoustic correlates of stress in Spanish," Proceedings of the 16th International Congress of Phonetic Sciences, Saarbrücken, Germany, 6–10 August 2007, edited by J. Trouvain and W. J. Barry, pp. 1121–1124.
Pierrehumbert, J. B. (**1980**). "The phonetics and phonology of English intonation," Ph.D. thesis, Massachusetts Institute of Technology, Cambridge, MA.
Recasens, D. (**1986**). *Estudis de Fonètica Experimental del Català Oriental Central (Experimental Phonetic Studies on East-Central Catalan)* (Publicacions de l'Abadia de Montserrat, Barcelona, Spain).
Recasens, D., and Espinosa, A. (**2006**). "Dispersion and variability of Catalan vowels," Speech Commun. **48**, 645–666.
Sluijter, A. M. C. (**1995**). *Phonetic Correlates of Stress and Accent* (Holland Academic Graphics, The Hague).
Sluijter, A. M. C., and van Heuven, V. (**1996**). "Spectral balance as an acoustic correlate of linguistic stress," J. Acoust. Soc. Am. **100**, 2471–2485.
Sluijter, A. M. C., van Heuven, V., and Pacilly, J. A. (**1997**). "Spectral balance as a cue in the perception of linguistic stress," J. Acoust. Soc. Am. **101**, 503–513.
Turk, A., and Sawusch, J. (**1996**). "The processing of duration and intensity cues to prominence," J. Acoust. Soc. Am. **99**, 3782–3790.
Wheeler, M. W. (**2005**). *The Phonology of Catalan* (Oxford University Press, Oxford, UK).

Spatial selective attention in a complex auditory environment such as polyphonic music

Katja Saupe[a)]
Department of Neurobiology, Institute of Biology II, University of Leipzig, Talstrasse 33, Leipzig D-04103, Germany

Stefan Koelsch[b)]
Junior Research Group Neurocognition of Music, Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Stephanstrasse 1a, Leipzig D-04103, Germany

Rudolf Rübsamen
Department of Neurobiology, Institute of Biology II, University of Leipzig, Talstrasse 33, Leipzig D-04103, Germany

(Received 4 June 2009; revised 13 November 2009; accepted 13 November 2009)

To investigate the influence of spatial information in auditory scene analysis, polyphonic music (three parts in different timbres) was composed and presented in free field. Each part contained large falling interval jumps in the melody and the task of subjects was to detect these events in one part ("target part") while ignoring the other parts. All parts were either presented from the same location (0°; *overlap condition*) or from different locations (−28°, 0°, and 28° or −56°, 0°, and 56° in the azimuthal plane), with the target part being presented either at 0° or at one of the right-sided locations. Results showed that spatial separation of 28° was sufficient for a significant improvement in target detection (i.e., in the detection of large interval jumps) compared to the overlap condition, irrespective of the position (frontal or right) of the target part. A larger spatial separation of the parts resulted in further improvements only if the target part was lateralized. These data support the notion of improvement in the suppression of interfering signals with spatial sound source separation. Additionally, the data show that the position of the relevant sound source influences auditory performance. © *2010 Acoustical Society of America.* [DOI: 10.1121/1.3271422]

PACS number(s): 43.75.Cd, 43.66.Pn, 43.66.Jh [DD] Pages: 472–480

I. INTRODUCTION

Our everyday listening environment is often highly complex, with many sounds occurring at the same time. Sitting in an office, for example, you might hear the telephone ringing, people talking, and traffic noise outside. Such a setting was exemplified by Cherry (1953) as a cocktail party situation. Because all the surrounding sound signals arrive at the cochlea as a composite, a preliminary analysis of the incoming sound is required to divide the auditory input into distinct perceptual objects (also referred to as auditory scene analysis; see Bregman, 1990, for review). To select relevant information from concurrent, irrelevant sound streams, spectral, temporal, and spatial cues are analyzed and integrated (for a review of selective attention to auditory objects see Alain and Arnott, 2000). As long as only two sound sources are present, the influence of spatial information on the segregation of auditory objects is often relatively small if other stimulus parameters are available instead (Butler, 1979; Deutsch, 1975; Shackleton *et al.*, 1994; Yost *et al.*, 1996). However, the benefit from spatial information increases significantly with three simultaneously active sound sources (Eramudugolla *et al.*, 2008; Hawley *et al.*, 2004; Yost *et al.*, 1996). Previous studies investigating spatial auditory attention with more than two sound sources often presented the competing stimuli successively rather than simultaneously (e.g., Münte *et al.*, 2001; Nager *et al.*, 2003; Teder-Sälejärvi *et al.*, 1999). Treisman (1964) was one of the first to present up to three sources simultaneously for the investigation of selective filtering in auditory attention. While retaining a fixed attended sound source, the number (0–2) and simulated spatial location of irrelevant sound sources, as well as the simulated distance between sound sources, were varied. However, because of the dichotic presentation of sound stimuli through earphones instead of free field stimulation, the acoustical percept was somewhat unnatural with sound sources being either located directly at the two ears or along an intracranial axis between the two ears. Yost *et al.* (1996) created a natural listening condition by using up to three simultaneously active sound sources in free field and used spoken words, letters, and numbers as acoustic stimuli. Divided attention to all simultaneously active stimuli was tested for different positions and different spatial separations of the sound sources. Several follow-up studies investigated the masking influence of task-irrelevant acoustic stimuli on the speech reception threshold in a multi-source environment (Culling *et al.*, 2004; Hawley *et al.*, 1999, 2004; Kidd *et al.*, 2005; Peissig and Kollmeier, 1997).

[a)]Author to whom correspondence should be addressed. Present address: Institute of Psychology I, University of Leipzig, Seeburgstrasse 14-20, Leipzig D-04103, Germany. Electronic mail: saupe@rz.uni-leipzig.de
[b)]Present address: Department of Psychology, Pevensey Building, University of Sussex, Falmer, Brighton, BN1 9QH, UK.

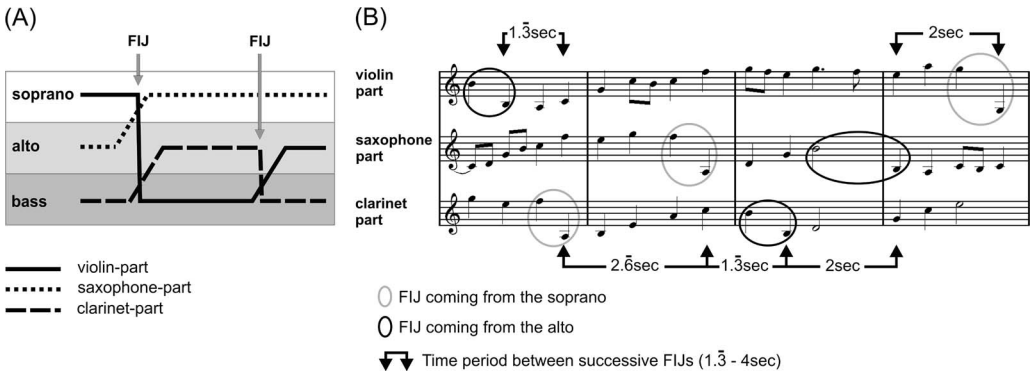


FIG. 1. (A) Example of the melody course of the three parts. The violin part (target part) is initially in the soprano register, saxophone part in the alto, and the clarinet part in the bass register. A FIJ in the melodic contour of at least 1 octave in the violin part (FIJ in the soprano) brought that part into the bass register; isochronal, the saxophone part turns into the soprano and the clarinet part into the alto register. Later in this example, a FIJ occurs in the clarinet part, which is in the alto register before that FIJ, and in the bass register after that FIJ. The different registers are indicated by different shades of gray; (B) an excerpt sheet of music. Note that the same number of FIJs occurred in each part. Targets were defined as FIJs occurring in the violin part, and distracters were FIJs occurring in the saxophone and clarinet parts.

While the cocktail party phenomenon in a complex listening environment has mostly been described for language comprehension, listening to music often requires similar processes. Polyphonic music (i.e., multi-part music) can also contain several simultaneously active sound sources, creating multiple auditory streams. When different parts are played, for example, by different instruments, it is possible to focus attention selectively to one instrument and to follow the melody played by this instrument (Janata *et al.*, 2002). A first step in investigating the contribution of spatial information in selective listening to musical patterns was the "scale illusion" (Deutsch, 1975). In this paradigm, dichotic tonal sequences consisting of the repetitive presentation of an ascending and a descending scale were presented dichotically in a way that adjacent tones of the scales were switched from ear to ear. Most of the subjects heard instead of alternating scales, two streams, one from high- to medium-pitched and back to high notes, and a second from low- to medium-pitched and back to low notes. This indicates that stimuli were channeled mainly by pitch range rather than by the ear of input. Butler (1979) extended this paradigm to the use of other melodic materials and demonstrated that the effects can also be transferred to free field stimulation. Later, it has been shown that introducing differences in timbre can cause a degradation of the scale illusions (Smith *et al.*, 1982).

It has also been shown that a decrement in the integration of melodic patterns occurs if the tones of the pattern were distributed pseudorandomly between the ears compared to when presented binaurally (Deutsch, 1979). When a lower frequency tone (drone) was simultaneously presented to the ear opposite to that receiving the melody component, performance was largely improved again but not if the drone and the melody tone were delivered to the same ear. The authors suggested competition between two organizing principles: "Where input is to one ear at a time, localization cues are very compelling, so that linkages are formed on the basis of ear input and not frequency proximity. However, when both ears receive input simultaneously, an ambiguity arises as to the sources of these inputs, so that organization by frequency proximity becomes a more reasonable principle."

The present study specifically investigates the influence of spatial information for selective attention to one of three simultaneously presented melody parts of polyphonic music (see also Fig. 1). This phenomenon bears some similarities to spatial release from masking for which two mechanisms are currently discussed (Bronkhorst, 2000; Durlach, 1963; vom Hövel, 1984): (i) the binaural unmasking of lower frequencies facilitated due to different interaural time differences (ITDs) between competing sound sources; (ii) "best ear" listening, i.e., a benefiting signal-to-noise ratio at the ear ipsilateral to the target sound source and contralateral to the interfering sound sources caused by the headshadow effect. If binaural unmasking is effective, an improvement in attending to the relevant stream with increasing distance of the sources (i.e., increasing differences between ITDs related to the competing sound sources) would be expected. Indeed, Drennan *et al.* (2003) demonstrated an increase in the ability to segregate two competing speech sounds with increasing angle between the sources in the acoustic free field, as well as with increasing interaural time differences under headphone conditions.

In the present study, all three melodic parts were either presented from the same or from different locations, with the to-be-attended part being presented either at 0° or at one of the right-sided locations (i.e., +28° or +56° in the azimuthal plane). Due to an increasing influence of binaural unmasking, we hypothesized that the detection of targets occurring in the attended part would improve with increasing spatial separation of the different parts. An additional aspect of the present study was to find out whether the position of the relevant part, either frontal or lateral to the subject, influences it target detections.

II. METHODS

A. Subjects

The data of 20 right-handed and normal-hearing subjects (nine females) aged 22–30 years (mean age of 25.8 years) were included in this study (two female subjects were excluded after the practice blocks because they were unable to

detect the targets). All subjects were non-musicians, i.e., they had no formal musical training (apart from normal school education). None of the participants had a history of a neurological disease or injury. All subjects participated on a voluntary basis, gave written informed consent, and received monetary reimbursement.

B. Stimuli

Eight polyphonic music pieces, each consisting of three parts, and each with a length of approximately 3:45 min (ranging from 3:41 to 3:53 min) were created using the software CUBASE SX 2.01 (Steinberg Media Technology GmbH, Hamburg, Germany). Each part had a different computer-generated timbre (Violin, Saxophone, and Clarinet) and was synthesized into a single wav-file. The assignment of the three timbres to the different parts was the same for each subject and each music piece. The three single wav-files of one composition were merged into one multi-channel wav-file using MATLAB 7.1 (The MathWorks, Nattick, MA). The three different parts of one composition played in three different registers (soprano, alto, and bass). The register, in which a part plays, is defined by the pitch of the part in relation to the pitch of the other two parts (with the soprano playing the highest and the bass playing the lowest pitch).

The melodic contour of each part was non-monotonically ascending until interrupted by a target/distractor, which is a falling interval jump (FIJ) in the melody of at least 1 octave. The FIJs occurred in all three parts and always turned the part in which a FIJ occurred into the bass register [Fig. 1(A)]. Due to the general pattern of melodic ascent interrupted by sudden descents (FIJs), each part played in different registers during each composition [see Fig. 1(A) for a schematic illustration].

An important requirement concerning the stimulus design was that target detection should only be possible by selectively attending to the relevant part, instead of global listening to the overall sound of the music pieces. To control for this (i) FIJs did not violate the harmony of the overall sound (i.e., they did not induce dissonances); (ii) FIJs occurred only in a part, which was in the soprano or in the alto register (to avoid large changes in the frequency range of the overall sound); and (iii) no discernible breaks occurred in the melodic contour of any register even when a crossover of parts occurred due to a FIJ. For example, in Fig. 1(A), during the first FIJ the melodic contour in the soprano register does not change noticeably although the violin part has been "replaced" by the saxophone part, and likewise the saxophone by the clarinet and clarinet by the violin in the alto and bass registers, respectively.

If downward movements in the melodic contour occurred apart from FIJs, they only spanned an interval of at most three semitones and were therefore hardly confused with targets [Fig. 1(B)].

Each part in each piece of music contained 38 FIJs (19 while the part was in the soprano, and 19 while the part was in the alto register). The time period between successive FIJs (irrespective of the part) was 1333–4000 ms [Fig. 1(B)], with the first one in any part occurring not earlier than 4000 ms after the onset of the music piece.

The duration of single tones used for the compositions was 333–2667 ms, which is equivalent to an eighth note and a whole note played at a tempo of 90 beats/s. The tones directly followed each other, i.e., there was no interstimulus interval. The second tone of a FIJ always occurred on a beat and its duration was at least 667 ms (equivalent with a quarter note).

The three parts of a music piece were similar in style and contained the same components (tone durations, frequencies, and rhythmic patterns) in the same proportion. Our stimuli were major-minor tonal music and—except for the frequent crossings of parts—composed by following the classical theory of harmony (e.g., Hindemith, 1940).

Stimulus intensity was 50 dB sensation level (SL), i.e., 50 dB above the individual sensation threshold.

C. Task

The subjects were instructed to focus their attention on the melodic contour of the violin part (which was always the target part) and to detect the FIJs occurring in this part (targets). Subjects indicated detection as fast as possible by pressing a button on a response box. FIJs in the distractor parts (i.e., in the saxophone and clarinet parts) were to be ignored.

D. Apparatus and procedure

Testing was performed in an echo- and sound-attenuated room with walls, ceiling, and floor covered by acoustic foam with 5 cm^3 wedges. The setup consisted of a semicircular platform of 2.3 m in radius, raised 1.13 m above the floor, with speakers (Control IG JBL) positioned at 0° azimuth, 28° to the left and to the right (−28° and +28°, respectively), and 56° to the left and to the right (−56° and +56°, respectively). The complete setup was covered by black gauze, so the speakers were not visible to the subjects. During the tests, the subjects were comfortably seated on an adjustable chair in the center of the loudspeaker array. Subjects were asked to focus a fixation point at 0°. Head movements were prevented by fixing the head to the backrest. During testing, subjects were observed from a neighboring control room through a semitransparent mirror.

By use of the software PRESENTATION Version 9.07 (Neurobehavioral Systems, Inc., Albany, CA) single instrumental parts could be assigned to each of the five loudspeakers through an eight-channel soundcard (SB Audigy 2Zs Audio). Five experimental conditions were defined: an overlap condition
(0°/0°/0°), in which all instrumental parts were presented through one loudspeaker at 0°, and four separation conditions, in which the three different parts were presented from three different loudspeakers. The four separation conditions were subdivided into two with slight (+28°/−28°/0° and 0°/−28°/+28°) and two with wide (+56°/−56°/0° and 0°/−56°/+56°) speaker separations. Part locations are de-

TABLE I. Loudspeaker configurations for the different stimulus conditions. Part locations are depicted in the order: violin part (target part)/saxophone part/clarinet part as degrees from front.

Conditions	Violin/saxophone/clarinet
Overlap condition	0°/0°/0°
Separation conditions, target part frontal	0°/−28°/+28°
	0°/−56°/+56°
Separation conditions, target part lateral	+28°/−28°/0°
	+56°/−56°/0°

picted in the order: violin part (target part)/saxophone part/clarinet part as degrees from front.

Three different compositions (pseudorandomly chosen for each subject from the set of the eight compositions) were presented for each of the five loudspeaker configurations (see Table I). The order of the resultant 15 experimental blocks was randomized between subjects. During short breaks between the blocks, the subjects were informed whether the violin part (target part) will be presented up front or from the right side in the upcoming presentation.

Prior to data acquisition, subjects performed training blocks to assure that they were able to detect the targets. The training was subdivided into three phases: first and second, only the violin part was presented from 0° and subjects were instructed to just listen to the part (first phase) or to indicate target detection by pressing a button in response to targets (second phase). Subjects were qualified for the experiment only if they showed a success rate of at least 85% in the second phase. At last, a composition with all three parts was presented in the 0°/−56°/+56° loudspeaker configuration, with the target part being 10 dB louder than the distracting parts (to familiarize the subjects with the polyphonic sound of the compositions). For each phase of the training a different compositions out of the experimental set was chosen, which was balanced over subjects.

E. Data analysis

Behavioral responses were analyzed for the frequency of hits, selection errors, and detection errors. The respective response categorization was based on the reaction times to targets and distracters. Key presses 200–1100 ms after occurrence of targets were categorized as "hits," after distracters as "selection errors," and key presses outside these time windows as "detection errors."

For a better comparability of the conditions, all respective response categories were related to each other by means of a non-dimensional qualifier, in the following denoted as q-index. The range of the q-index is confined to an interval (0,…,1), with "1" indicating accurate detection of all targets and no responses in the selection and detection error categories. On the contrary, the q-index becomes "0" if at least one of the following three cases was observed: (i) no target was detected, (ii) the subject reacted to all distracters, or (iii) if the total number of responses equals the number of detection errors. This is formalized by the following equation:

$$q = \frac{H}{N_H} \cdot \frac{(N_{ES} - ES)}{N_{ES}} \cdot \frac{(n - ED)}{n}, \quad (1)$$

with H representing the response number of hits, ES is the response number of selection errors, ED is the number of detection errors, N_H is the maximum attainable number of hits, N_{ES} is the maximum attainable number of selection errors, and n the total of responses ($n=H+ES+ED$).

Differences in the q-index and in the reaction times between the overlap and separation conditions were quantified by paired t-tests under consideration of the alpha level Bonferroni corrections. For the separation conditions, the q-indices and the reaction times of the hits were subjected to two-way analyses of variance (ANOVAs) (factors: separation × position of the target part); Bonferroni multiple comparisons post hoc tests were used for comparisons between the conditions. The Greenhouse–Geisser correction was applied when the assumption of sphericity was violated. Differences were assessed as significant at an alpha level of 0.05.

The acoustic stimulation lasted for several minutes and was not subdivided into separate trials. In addition, subjects could respond at any time. As a consequence, a considerable number of hits could be achieved by an excessively high response rate. Because the number of responses was not limited, it was necessary to calculate the chance level, which takes these aspects into consideration. Therefore, the number of responses in the hits category and the q-index were tested against chance expectation with the chance level calculated based on an urn model (Johnson and Kotz, 1977). For that, each of the three answer categories (H=hits, ES=selection errors, and ED=detection errors) was assigned to a defined number (N_H, N_{ES}, and N_{ED}) of objects. N_H corresponds to the number of targets that are included in the target part (violin part), while N_{ES} corresponds to the number of distracters that are included in the saxophone and clarinet parts. N_{ED} will be ascertained, assuming that the number of n-responses (the total number of responses measured for a real subject) is uniformly distributed over the whole stimulus duration, with no relation to the stimulus structure, in a hypothetical subject (Poisson process, see Pitman, 1993). The distributions of the responses in the hits category correspond to a distribution while drawing n samples (without replacement), which is known in the literature as a hypergeometrical distribution (Pitman, 1993; Sachs, 1992).

III. RESULTS

A. Response rate

1. Influence of spatial separation of the parts

In all five conditions the target detection (number of hits and values for the q-index) was above chance level. The means in the three response categories for all five conditions are indicated in Table II. For a better comparability of the subjects' performance, the respective response categories were combined for the calculation of the non-dimensional qualifier, which is the q-index (see Sec. II for details) [Fig. 2(A)]. A higher q-index was found for the separation conditions than for to the overlap condition [0°/0°/0° ×

TABLE II. Response rate of hits, selection errors, and detection errors for all five conditions of part locations (mean, with SD in parentheses). Part locations are depicted in the order: violin part (target part)/saxophone part/clarinet part as degrees from front. The maximal attainable number of hits was 114 and number of selection errors was 228.

Part location (deg from front)	Hits	Selection errors	Detection errors
	Mean (SD)	Mean (SD)	Mean (SD)
0/0/0	62.7(16.8)	12.2(6.9)	7.5(4.9)
0/−28/+28	94.9(14.1)	2.3(1.8)	4.3(3.3)
+28/−28/0	93.7(14.3)	5.8(3.2)	3.8(3.1)
0/−56/+56	93.7(14.9)	2.9(−3.0)	4.9(5.3)
+56/−56/0	101.2(11.5)	2.9(2.7)	2.6(2.5)

separation conditions: $t_{0°/-28°/+28}(19)=-13.44$, $p_{0°/-28°/+28°}<0.001$; $t_{0°/-56°/+56°}(19)=-10.92$, $p_{0°/-56°/+56°}<0.001$; $t_{+28°/-28°/0°}(19)=-10.27$, $p_{+28°/-28°/0°}<0.001$; and $t_{+56°/-56°/0°}(19)=-15.17$, $p_{+56°/-56°/0°}<0.001$]. In general, the sensitivity for target detection differed strongly between subjects, as indicated by the high standard deviation. Still, in all subjects, the same tendency for an increase in the q-index with spatial separation of the parts was observed (for single subject data, see Fig. 3).

2. Increasing spatial separation at different positions of the target part

For the different separation conditions, the influence of the separation and position factors of the target part (position) was evaluated based on the respective q-indices [Fig. 2(B)]. Two-way ANOVAs (factors: separation × position) resulted in a significant main effect of separation [$F(1,19)=5.48$, $p<0.05$] but not of position [$F(1,19)=3.45$, $p=0.79$], and in a significant interaction between the separation and position factors [$F(1,19)=9.71$, $p<0.01$]. The Bonferroni multiple comparisons *post hoc* tests indicated a significant increase in the q-index with increasing spatial separation for lateral target positions [(+28°/−28°/0°) × (+56°/−56°/0°): $t(19)=-3.93$, $p=0.005$] but no effect of separation for frontal target positions [(0°/−28°/+28°) × (0°/−56°/+56°): $t(19)=0.36$, $p=0.999$]. Also, an effect of position could not be observed for moderate signal separations [(0°/−28°/+28°) × (+28°/−28°/0°): $t(19)=0.93$, $p=0.999$], while larger separation led to better target detection of laterally presented parts [(0°/−56°/+56°) × (+56°/−56°/0°): $t(19)=-3.42$, $p<0.05$].

B. Reaction time

The reaction times (RTs) for hits (Fig. 4) were significantly longer in the overlap condition compared to the separation conditions [(0°/0°/0°) × separation conditions: $t_{0°/-28°/+28°}(19)=3.79$, $p_{0°/-28°/+28°}<0.005$; $t_{+28°/-28°/0°}(19)=4.65$, $p_{+28°/-28°/0°}<0.005$; $t_{0°/-56°/+56°}(19)=3.51$, $p_{0°/-56°/+56°}<0.01$; and $t_{+56°/-56°/0°}(19)=4.32$, $p_{+56°/-56°/0°}<0.005$]. Comparisons within the different separation conditions showed neither a main effect of separation [$F(1,19)=0.15$, $p=0.7$] nor of position [$F(1,19)=1.69$, $p=0.21$] and also no interaction of these two factors [separation × position: $F(1,19)=1.48$, $p=0.24$].

IV. DISCUSSION

Using a target detection task, the influence of spatial information for selective attention to one of multiple streams was investigated with polyphonic music.

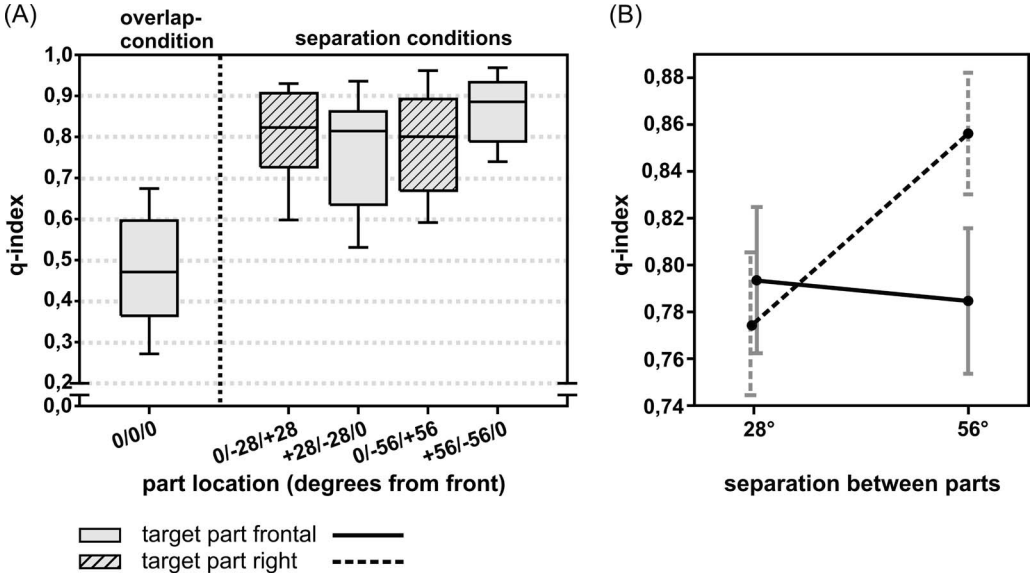


FIG. 2. (A) q-indices for performance in the different stimulus conditions and (B) the interactions of the q-indices between the factors separation and position of the target part in the separation conditions. The separation factor is divided into moderate (conditions 0°/−28°/+28° and +28°/−28°/0°) and large separations (conditions 0°/−56°/+56° and +56°/−56°/0°) of the three parts. The position factor is divided into frontal presentation of the target part (conditions 0°/−28°/+28° and 0°/−56°/+56°; solid line) and presentation of the target part on the right side (conditions +28°/−28°/0° and +56°/−56°/0°; hatched line). Box plots in (A) show medians, interquartile, and interdecile ranges. Target parts frontal are indicated by gray boxes, target parts lateral are indicated by hatched boxes. Mean values and standard error of the means are shown in (B). Part locations are depicted in the order: violin part (target part)/saxophone part/clarinet part as degrees from front.

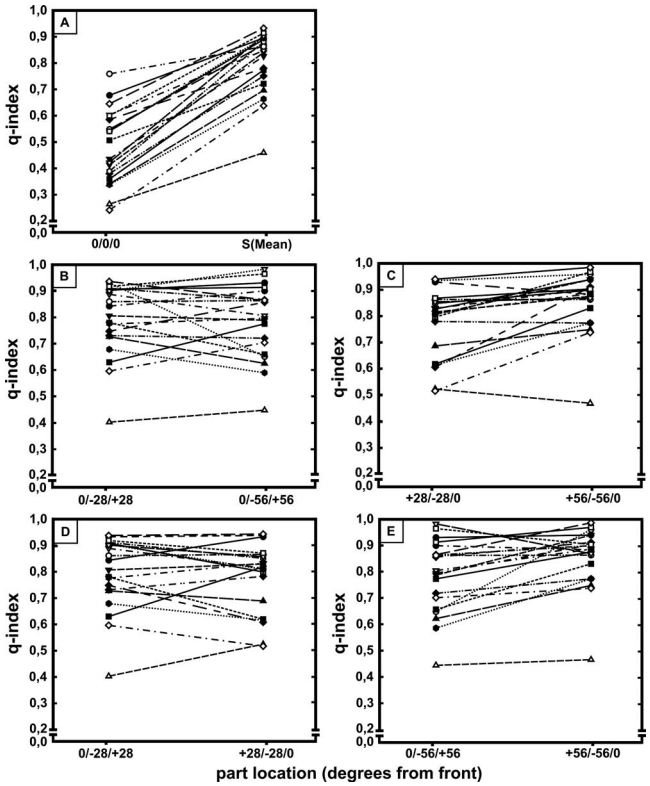


FIG. 3. Q-index per subject: (A) overlap condition (0°/0°/0°) compared to the mean of all separation conditions [S (mean)]; (B) increasing spatial separation of the three parts with target part frontal (condition 0°/−28°/+28° and 0°/−56°/+56°) and (C) lateral (condition +28°/−28°/0° and +56/−56/0°); comparison between target part frontal and lateral while same spatial separation of the parts for (D) moderate (condition 0°/−28°/+28° and +28°/−28°/0°) and (E) large separations of the parts (condition 0°/−56°/+56° and +56°/−56°/0°). Each line indicates the mean values for one subject. Part locations are depicted in the order: violin part (target part)/saxophone part/clarinet part as degrees from front.

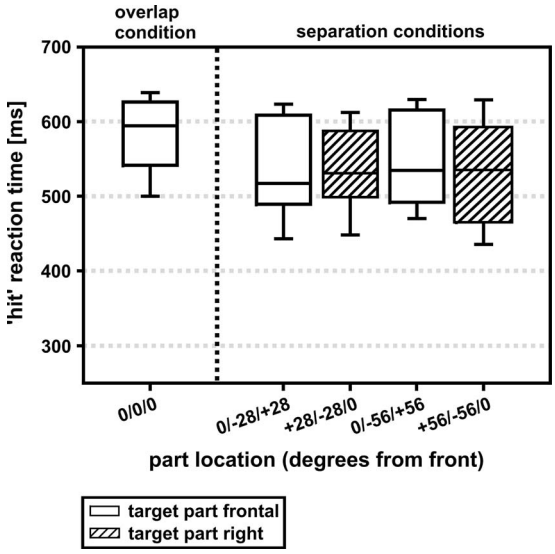


FIG. 4. Distribution of mean reaction time for hits in the overlap (0°/0°/0°) and the separation conditions (0°/−28°/+28°, +28°/−28°/0°, 0°/−56°/+56°, and +56°/−56°/0°). Median, interquartile, and interdecile values are shown. Part locations are depicted in the order: violin part (target part)/saxophone part/clarinet part as degrees from front.

A. Influence of spatial separation of the parts

In the overlap condition, 55% of the targets were detected, which is consistent with results of Janata *et al.* (2002), who also studied a selective attention task to polyphonic music. In that study, subjects were asked to detect timbral deviants in an attended stream during the presentation of three different instruments (different timbres). While in the selective attention condition in Janata *et al.*, 2002, timbral deviants only occurred in the attended stream, in the present study, the FIJs (in the melodic contour) occurred in all three parts. Still, only those in the violin part (targets) were task-relevant, whereas those in the saxophone and clarinet parts (distracters) had to be ignored. Moreover, the fact that FIJs did neither violate the frequency range nor the harmony of the overall sound impeded target recognition. Thus, targets could not be detected by integrative listening to global changes in the overall sound structure but only by selective listening to the relevant part and taking no account of the respective structures in irrelevant parts.

The possibility of explaining the target detection level by random responses could be ruled out for all subjects and for all conditions. Thus, in agreement with previous studies (Janata *et al.*, 2002; Yost *et al.*, 1996), the present results also show that, in a more challenging experimental setting, basal stream segregation of multiple sound sources is possible even without spatial separation of the sources just by the use of spectral information. However, additionally adding spatial information resulted in a significant increase in performance and decrease in reaction times. Even the relatively small spatial signal separation of 28° facilitates the allocation to different sound sources and improves listeners' ability to follow the relevant melodic contour. This result is consistent with previous findings based on the comprehension of spoken language, which showed a large influence of spatial separation for the processing of single sound sources in complex auditory environments (Hawley *et al.*, 1999, 2004; Peissig and Kollmeier, 1997; Yost *et al.*, 1996). Because in some of these studies the subjects had to reproduce the sentences/words after listening (Hawley *et al.*, 1999, 2004; Yost *et al.*, 1996), some constraints have to be made regarding the interpretation of these earlier results. Spoken language bears a lot of redundancy and this might have aided the performance of the subjects. On the other hand, reproduction of aurally perceived speech material relies on memory processes, for which was not controlled in these experiments. Our own endeavor was not hampered by any of these constraints and thus clearly disclosed the positive effects of spatial segregation on selective attention in complex acoustic environments.

B. Increasing spatial separation at different positions of the target part

Our results demonstrate that 28° spatial separation of the parts leads to significant improvement in target detection compared to the overlap condition, irrespective of the position (frontal or right) of the target part.

We hypothesized that, with increasing spatial separation of the relevant and the distracting parts, the influence of spatial unmasking further increases. As mentioned above, two mechanisms—best-ear listening and binaural unmasking—

are discussed to be responsible for spatial release from masking (Bronkhorst, 2000; Durlach, 1963; vom Hövel, 1984). During frontal presentation of the target part, masking signals were presented in both hemifields with equal distances to the target part. Note that, in each of the three parts, subsequent tones directly followed each other (with no silent gap between tones). This led to a continuous masking of the target signal from both sides during frontal presentation of the target part. For this reason, best-ear listening should not be very effective in this condition (even if it cannot be excluded entirely due to differences in the interfering parts). In contrast, during lateral presentation of the target signal, head-shadow effects might additionally have improved the signal-to-noise ratio for the relevant part, resulting in a monaural advantage at the best ear. Specifically, such a monaural advantage becomes effective if the interfering signals are in the same hemifield, instead of being located in different hemifields with the relevant sound source in-between (Hawley *et al.*, 1999, 2004; Peissig and Kollmeier, 1997; Zurek, 1992). Thus, during lateral presentation of the target part, both mechanisms—binaural unmasking and best-ear listening—should have been available for spatial unmasking, whereas during frontal presentation of the target part, binaural unmasking should be the dominant mechanism in the present experiment.

In agreement with our hypothesis, an increase in spatial separation of the parts resulted in a further improvement of target detection if the target part was lateralized (+56° to the right side). This goes in line with previous studies, which demonstrated an improvement of signal processing with increasing interaural time and intensity differences in acoustic speech material (Yost *et al.*, 1996) or noise-band material (Drennan *et al.*, 2003).

Previous reports showed that the ability to focus on a specific location is strongly reduced in the periphery compared to the central auditory space (Teder-Sälejärvi *et al.*, 1998, 1999). Also, spatial processing of sound sources, e.g., localization and spatial discrimination, is much better for frontal compared to lateral sound positions (Oldfield and Parker, 1984; Perrott *et al.*, 1993). Because of this enhanced frontal focus of selective auditory attention, one might expect that the processing of the relevant stream coming from the frontal loudspeaker is at least as good as when presented laterally. However, if in the present study the target part was presented frontally, no further improvement of target detection with increasing spatial separation from the irrelevant parts was observed. This lack of effect in the frontal target position suggests that, in the present experiment, the increase in spatial unmasking may be driven mainly by the best-ear listening.

However, signal modification by the pinnae might be the base for an alternative explanation of the results. Previous experiments demonstrated an amplification of the incoming signal, depending on the angle of incidence at the pinnae (Saberi *et al.*, 1991; Shaw, 1974). The highest amplification for the respective ear was demonstrated for lateral angles of 45° and 50°. Thus, presenting the target part from the lateral loudspeaker led to an additional increase in signal-to-noise ratio, which might help to separate the relevant part from the three concurrent streams. The other way around, during frontal presentation of the target part (with distracting parts on the left and right sides), the distracting parts will be amplified compared to the target part due to the more favorable angle of incidence at the pinnae. This effect should even increase with increasing spatial separation in the present experiment, which means that the intensity of the distracting parts increase compared to the target part. Interestingly, this would conflict with a possible benefit received by increasing binaural unmasking during the frontal presentation of the target part. But even if during frontal presentation of the target part, the signal-to-noise ratio decreases with increasing spatial separation (due to the pinnae related amplification of the distracting parts), the supportive influence of binaural unmasking should increase, and also the focus of auditory attention to frontal location should increasingly aid the selection of the target part at this position. Previous studies describe a gradual distribution of auditory attention (Mondor and Zatorre, 1995; Teder-Sälejärvi and Hillyard, 1998), which means that the amount of attention decreases with increasing distance of the sound source to the focus of attention. Because of the ability to specifically focus attention to the central auditory space, processing of the relevant part at frontal position (and ignoring of the lateral irrelevant parts) with increasing spatial separation should be facilitated. However, it is possible that the influence of these competing mechanisms (increasing binaural advantage and the gradient of attention on one side, monaural disadvantage on the other side) is equally strong in condition 0°/−56°/+56°. This might be the reason why during frontal presentation of the target part, we observed no difference in target detection with increasing spatial separation.

Also, impeded signal processing for frontal signals in a masking situation has been reported earlier. For example, deterioration in speech reception thresholds was documented for frontal target signals when the two masking signals are distributed in both hemifields, as compared to when they originate from one hemifield (Culling *et al.*, 2004; Hawley *et al.*, 1999, 2004; Peissig and Kollmeier, 1997). Similarly, using headphone stimulation, Treisman (1964) observed better word identification when the relevant message was presented to the right ear instead of being presented binaurally, i.e., perceived centrally. These results were explained by assuming object formation occurring outside the focus of attention, which improves with increasing similarity of the irrelevant sound sources, and which allows listeners to group them into a separate object distinct from the relevant sound source (Alain and Arnott, 2000; Alain *et al.*, 1996; Alain and Woods, 1993; Treisman, 1964).

No difference in performance was observed between frontal and lateral presentations of the target part during small separation angle of the parts (0°/−28°/+28° and +28°/−28°/0°, respectively). Considering the discussion above, one might expect better task performance during lateralized presentation of the target part also for small separation angles because the two distinct mechanisms (best-ear listening and the signal amplification by the pinnae) should be available in addition to binaural unmasking. The fact that this was not the case implies that these two monaural mecha-

nisms were only of minor significance in the small separation condition. However, it is still possible that the results indicate balanced effects of opposite mechanisms: While the focus of attention might, in addition to binaural unmasking, facilitate the processing of the frontal sound source, best-ear listening and the monaural advantage due to the spectral influence of the pinnae might be in favor of processing of the target part presented laterally.

In contrast to the target detection rate, response times seem to be unaffected by the position of the target part as soon as all three parts are presented from different locations.

V. SUMMARY AND CONCLUSIONS

Sustained selective attention to one musical instrument in polyphonic music is possible even without spatial separation, only by using spectral information. Introducing spatial separation between the relevant and the masking parts improved the selective processing of the relevant information. Increasing spatial separation from 28° to 56° between the different sound sources caused further improvements in target detection, if the relevant signal is presented laterally (with one masking signal presented from the opposite hemifield and one from 0°). If, under the identical arrangement of the sound sources, the relevant part is in front (with one masking signal presented from the right and one from the left side), no further improvement in target detection became evident with increasing separation of the distracters from 28° to 56° in the present experiment.

The competing influences of (i) best-ear listening, (ii) binaural unmasking, (iii) focus of attention, and (iv) pinna-related signal amplification on target detection were discussed. Still, the present stimulus design did not allow us to distinguish between the contributions of these alternatives. To make a distinction between the influences of binaural unmasking and monaural mechanisms, the influence of the interaural intensity differences (IIDs) and ITDs could be tested separately, following the experimental design employed by Culling *et al.* (2004). The usage of larger angles of separation would be helpful to prove the influence of the filter characteristics of the pinna. If the angle of incidence to the pinna has an effect, this influence should decrease with variations in the sound source around the angle of highest pinneal signal amplification.

Concluding, the present study demonstrates that the underlying mechanisms for solving the cocktail party problem are universal and can be utilized in selective listening to music as has been demonstrated for selective language processing earlier (e.g., Kidd *et al.*, 2005; Yost *et al.*, 1996).

ACKNOWLEDGMENTS

The authors wish to thank Frank-Steffen Elster for helping in composing the stimuli, Sven Gutekunst and Maren Grigutsch for the support in programming, Gerd Joachim Dörrscheidt for providing the calculation of the *q*-index and the coincidence level, and Sven Gutekunst and Peter Wolf for technical support, as well as Ramona Menger for acquisition and scheduling of subjects.

This work was supported by a stipend of the German Research Foundation (Postgraduate Training Program "Function of attention in cognition," Grant No. DFG 1182).

Alain, C., and Arnott, S. R. (**2000**). "Selectively attending to auditory objects," Front. Biosci. **5**, d202–212.

Alain, C., Ogawa, K. H., and Woods, D. L. (**1996**). "Aging and the segregation of auditory stimulus sequences," J. Gerontol. B Psychol. Sci. Soc. Sci. **51**, P91–P93.

Alain, C., and Woods, D. L. (**1993**). "Distractor clustering enhances detection speed and accuracy during selective listening," Percept. Psychophys. **54**, 509–514.

Bregman, A. S. (**1990**). *Auditory Scene Analysis: The Perceptual Organization of Sound* (MIT Press, Cambridge, MA).

Bronkhorst, A. W. (**2000**). "The cocktail party phenomenon: A review of research on speech intelligibility in multiple-talker conditions," Acustica **86**, 117–128.

Butler, D. (**1979**). "Further study of melodic channeling," Percept. Psychophys. **25**, 264–268.

Cherry, E. C. (**1953**). "Some experiments on the recognition of speech, with one and with 2 ears," J. Acoust. Soc. Am. **25**, 975–979.

Culling, J. F., Hawley, M. L., and Litovsky, R. Y. (**2004**). "The role of head-induced interaural time and level differences in the speech reception threshold for multiple interfering sound sources," J. Acoust. Soc. Am. **116**, 1057–1065.

Deutsch, D. (**1975**). "Two-channel listening to musical scales," J. Acoust. Soc. Am. **57**, 1156–1160.

Deutsch, D. (**1979**). "Binaural integration of melodic patterns," Percept. Psychophys. **25**, 399–405.

Drennan, W. R., Gatehouse, S., and Lever, C. (**2003**). "Perceptual segregation of competing speech sounds: The role of spatial location," J. Acoust. Soc. Am. **114**, 2178–2189.

Durlach, N. I. (**1963**). "Equalization and cancellation theory of binaural masking-level differences," J. Acoust. Soc. Am. **35**, 1206–1218.

Eramudugolla, R., McAnally, K. I., Martin, R. L., Irvine, D. R. F., and Mattingley, J. B. (**2008**). "The role of spatial location in auditory search," Hear. Res. **238**, 139–146.

Hawley, M. L., Litovsky, R. Y., and Colburn, H. S. (**1999**). "Speech intelligibility and localization in a multi-source environment," J. Acoust. Soc. Am. **105**, 3436–3448.

Hawley, M. L., Litovsky, R. Y., and Culling, J. F. (**2004**). "The benefit of binaural hearing in a cocktail party: Effect of location and type of interferer," J. Acoust. Soc. Am. **115**, 833–843.

Hindemith, P. (**1940**). *Unterweisung im Tonsatz (The Craft of Musical Composition)* (Schott, Mainz).

Janata, P., Tillmann, B., and Bharucha, J. J. (**2002**). "Listening to polyphonic music recruits domain-general attention and working memory circuits," Cogn. Affect. Behav. Neurosci. **2**, 121–140.

Johnson, N. L., and Kotz, S. (**1977**). *Urn Models and Their Application: An Approach to Modern Discrete Probability Theory (Probability & Mathematical Statistics)* (Wiley, New York).

Kidd, G., Arbogast, T. L., Mason, C. R., and Gallun, F. J. (**2005**). "The advantage of knowing where to listen," J. Acoust. Soc. Am. **118**, 3804–3815.

Mondor, T. A., and Zatorre, R. J. (**1995**). "Shifting And focusing auditory spatial attention," J. Exp. Psychol. Hum. Percept. Perform. **21**, 387–409.

Münte, T. F., Kohlmetz, C., Nager, W., and Altenmüller, E. (**2001**). "Neuroperception—Superior auditory spatial tuning in conductors," Nature (London) **409**, 580.

Nager, W., Kohlmetz, C., Altenmüller, E., Rodriguez-Fornells, A., and Münte, T. F. (**2003**). "The fate of sounds in conductors' brains: An ERP study," Brain Res. Cogn. Brain Res. **17**, 83–93.

Oldfield, S. R., and Parker, S. P. A. (**1984**). "Acuity of sound localization—A topography of auditory space.1. Normal hearing conditions," Perception **13**, 581–600.

Peissig, J., and Kollmeier, B. (**1997**). "Directivity of binaural noise reduction in spatial multiple noise-source arrangements for normal and impaired listeners," J. Acoust. Soc. Am. **101**, 1660–1670.

Perrott, D. R., Costantino, B., and Cisneros, J. (**1993**). "Auditory and visual localization performance in a sequential discrimination task," J. Acoust. Soc. Am. **93**, 2134–2138.

Pitman, J. (**1993**). *Probability* (Springer-Verlag, New York).

Saberi, K., Dostal, L., Sadralodabai, T., Bull, V., and Perrott, D. R. (**1991**).

"Free-field release from masking," J. Acoust. Soc. Am. **90**, 1355–1370.
Sachs, L. (**1992**). *Angewande Statistik* (Springer-Verlag, Berlin).
Shackleton, T. M., Meddis, R., and Hewitt, M. J. (**1994**). "The role of binaural and fundamental-frequency difference cues in the identification of concurrently presented vowels," Q. J. Exp. Psychol. A, **47**, 545–563.
Shaw, E. A. G. (**1974**). *The External Ear* (Springer-Verlag, New York).
Smith, J., Hausfeld, S., Power, R. P., and Gorta, A. (**1982**). "Ambiguous musical figures and auditory streaming," Percept. Psychophys. **32**, 454–464.
Teder-Sälejärvi, W. A., and Hillyard, S. A. (**1998**). "The gradient of spatial auditory attention in free field: An event-related potential study," Percept. Psychophys. **60**, 1228–1242.
Teder-Sälejärvi, W. A., Hillyard, S. A., Röder, B., and Neville, H. J. (**1999**). "Spatial attention to central and peripheral auditory stimuli as indexed by event-related potentials," Brain Res. Cognit. Brain Res. **8**, 213–227.
Treisman, A. M. (**1964**). "The Effect of irrelevant material on the efficiency of selective listening," Am. J. Psychol. **77**, 533–546.
vom Hövel, H. (**1984**). "Zur Bedeutung der Übertragungseigenschaften des Außenohres sowie des binauralen Hörsystems bei gestörter Sprachübertragung (The influence of the transfer characteristics of the external ear and the binaural hearing aid during defective speech transmission)," in *Fakultät für Elektrotechnik* (RWTH, Aachen).
Yost, W. A., Dye, R. H., and Sheft, S. (**1996**). "A simulated 'cocktail party' with up to three sound sources," Percept. Psychophys. **58**, 1026–1036.
Zurek, P. M. (**1992**). "Binaural advantages and directional effects in speech intelligibility," in *Acoustical Factors Affecting Hearing Aid Performance*, edited by G. A. Studebaker and I. Hochberg (Allyn & Bacon, Boston), pp. 255–276.

Perception of musical pitch and lexical tones by Mandarin-speaking musicians

Chao-Yang Lee[a] and Yuh-Fang Lee
School of Hearing, Speech and Language Sciences, Ohio University, Athens, Ohio 45701

(Received 9 July 2009; revised 8 October 2009; accepted 16 October 2009)

The relationship between music and language processing was explored in two perception experiments on the identification of musical notes and Mandarin tones. In the music task, Mandarin-speaking musicians were asked to identify musical notes of three timbres without a reference pitch. 72% of the musicians met the criterion for absolute pitch when an exact match was required, and 82% met the criterion when one-semitone errors were allowed. Accuracy of identification was negatively correlated with age of onset of musical training, and piano notes were identified more accurately than viola and pure tone stimuli. In the Mandarin task, the musicians were able to identify, beyond chance, brief Mandarin tone stimuli that were devoid of dynamic F0 information and cues commonly considered necessary for speaker normalization. Although F0 height detection was involved in both musical note and Mandarin tone identification, performances in the two tasks were not correlated. The putative link between absolute pitch and tone language experience was discussed. © *2010 Acoustical Society of America.* [DOI: 10.1121/1.3266683]

PACS number(s): 43.75.Cd, 43.71.Hw [DD] Pages: 481–490

I. INTRODUCTION

Music and language are distinctively human activities. Their parallels and contrasts have garnered much attention (Patel, 2008). However, the relationship between music and language processing remains underspecified. In this study, we explored the music-language relationship by examining pitch perception in music and lexically contrastive tones. Pitch perception is an integral part of music perception. Pitch variations can also signal a difference in word meaning in lexical tone languages. For example, the syllable *ma* in Mandarin means "mother" with a level tone (tone 1), "hemp" with a rising tone (tone 2), "horse" with a low tone (tone 3), and "scorn" with a falling tone (tone 4). In other words, music perception and lexical tone perception both involve mapping pitch onto discrete musical or linguistic categories. Given the role of pitch in musical and linguistic contrasts, whether the same perceptual mechanism is implicated in both musical and linguistic pitch perception has important implications for the music-language relationship.

In the current study, we conducted two perception experiments to address this issue: one on the perception of musical notes without a reference pitch, and the other on the perception of brief, isolated Mandarin tones produced by multiple speakers. Our main goal was to evaluate the hypothesis that absolute pitch is associated with tone language experience (Deutsch, 2002, 2006; Deutsch *et al.*, 2009, 2004a, 2006). The musical note identification experiment was a variant of the absolute pitch task developed by Deutsch *et al.* (2006), which has been used to gauge the presence of absolute pitch (Deutsch *et al.*, 2009, 2006; Lee and Hung, 2008). The Mandarin tone identification experiment was developed to assess how well Mandarin tones could be identified without dynamic F0 information (Lee, 2009; Lee and Hung, 2008; Lee *et al.*, 2008, 2009, 2010). In particular, isolated Mandarin syllables produced by multiple speakers were digitally processed to generate "onset-only" syllables. As the majority of the F0 contour was neutralized in these syllables, listeners were forced to use F0 height information for tone identification. Since the detection of F0 height is involved in both musical note identification and Mandarin tone identification, we examined the correlation of performances between the two tasks. If a common perceptual mechanism is implicated, performances in the two tasks should be correlated.

A. Absolute pitch and tone language

The link between absolute pitch and tone language was first suggested by Deutsch and colleagues (Deutsch, 2002, 2006; Deutsch *et al.*, 2004a, 2006, 2009). Absolute pitch is "the ability to identify the frequency or musical name of a specific tone or, conversely, the ability to produce some designated frequency, frequency level, or musical pitch without comparing the tone with any objective reference tone" (Ward, 1999). Only a small percentage of individuals in North America and Europe possess absolute pitch (Profita and Bidder, 1988), and the genesis of this ability remains unclear. The idea proposed by Deutsch and colleagues is that individuals with absolute pitch are able to associate a particular pitch with a verbal label, just as tone language speakers can associate a particular pitch or a combination of pitches with a lexical tone category. Considering this functional similarity, tone language speakers could be said to possess a form of absolute pitch. In other words, absolute pitch is treated by tone language speakers as a feature of speech.

Deutsch and colleagues further hypothesized that infants raised in a tone language environment acquire absolute pitch

[a]Author to whom correspondence should be addressed. Electronic mail: leec1@ohio.edu.

for lexical tones just like they would acquire other features of speech. When they receive musical training, they acquire absolute pitch for musical tones in the same way as they would acquire the tones of a second tone language. By contrast, for individuals reared in a nontone language environment, since no linguistically relevant absolute pitch has been acquired, those individuals would acquire absolute pitch for musical tones as if they would acquire the tones of a first tone language. Deutsch and colleagues therefore predicted more stable and precise pitch templates as well as higher prevalence of absolute pitch in speakers of a tone language.

The evidence for the hypothesis has come from acoustic studies comparing the consistency of lexical pitch production by tone and nontone language speakers. In particular, Deutsch *et al.* (2004a) found that the average F0 of word enunciations across recording sessions was more consistent for tone language speakers than nontone language speakers. In addition, while the nontone language speakers showed a significant decrease in pitch consistency across days relative to immediate repetition, the tone language speakers were equally consistent across days as on immediate repetition. This finding suggests that the pitch templates used by the tone language speakers were more stable over the long term than those used by the nontone language speakers. Burnham *et al.* (2004) replicated the findings of Deutsch *et al.* (2004a) with a variant of the word production task, but noted that the difference between tone and nontone languages was fairly small.

Studies on the prevalence of absolute pitch in tone and nontone language speakers have also provided support for the putative link between absolute pitch and tone language. Deutsch *et al.* (2006) asked English- and Mandarin-speaking musicians to identify piano notes without a reference pitch. They found a significantly higher percentage of absolute pitch possessors in the Mandarin speakers. For both groups of musicians, absolute pitch was more prevalent in individuals with an earlier onset of musical training. When the age of onset of musical training was controlled, absolute pitch was still more prevalent in the Mandarin speakers, further supporting the role of tone language experience in absolute pitch.

To rule out the potential contribution of ethnic heritage, Deutsch *et al.* (2009) used the same task to examine absolute pitch in musicians of Caucasian heritage (without tone language experience) and of East Asian heritage (with three levels of tone language proficiency: very fluent, fairly fluent, and nonfluent). They found that identification accuracy was positively associated with tone language experience. Furthermore, the accuracy of the East Asian nonfluent group was comparable to the accuracy of the Caucasian group, effectively ruling out a genetic interpretation. In addition, performance of the very fluent group was comparable to the performance of the Chinese musicians reported in Deutsch *et al.* (2006), indicating that the superior performance was not likely due to the country where the musicians received their music education.

To obtain a direct measure of both musical and lexical pitch perception abilities, Lee and Hung (2008) examined Mandarin tone identification by English-speaking musicians and nonmusicians, and musical note identification by the musicians. In the Mandarin task, all participants were given a brief tutorial on Mandarin tones and were asked to identify the tones of the syllable *sa* produced by 32 speakers. The stimuli included intact syllables and acoustically modified syllables with limited F0 information (i.e., silent-center and onset-only syllables). Despite having no prior experience with Mandarin, the musicians identified the intact and silent-center tones beyond chance. The musicians also outperformed the nonmusicians, indicating that some kind of musical ability facilitated lexical tone identification. In the music task, the musicians were asked to identify synthesized musical notes of three timbres without a reference pitch. Their performance was at chance level, suggesting that their superior performance over nonmusicians in Mandarin tone identification did not arise from absolute pitch. However, since none of the musicians actually possessed absolute pitch, the lack of correlation found between the two tasks could not be fully interpreted.

It should be noted that the absolute pitch hypothesis proposed by Deutsch and colleagues does not make explicit predictions about the association between lexical and musical pitch perception performances as were examined in Lee and Hung (2008). In particular, Deutsch and colleagues hypothesized that for Mandarin speakers, absolute pitch for lexical tones is acquired as tones in a first tone language, whereas absolute pitch for musical tones is acquired as tones in a second tone language. Given the distinction between first and second languages, the hypothesis does not necessarily predict an association between absolute pitch and lexical tone processing in tone language users. The hypothesis, however, should favor such an association if absolute pitch for music and absolute pitch for speech share common brain mechanisms (Deutsch, 2002).

B. Perception of vocal pitch height

The perception of pitch height is not only relevant for music perception, but also implicated in lexical tone perception. For languages with level tones (e.g., Cantonese), detection of F0 height is necessary for distinguishing the level tones. For languages with contour tones (e.g., Mandarin), where tones could be identified based on dynamic F0 information, F0 height can still influence tone perception (Gandour, 1983; Gandour and Harshman, 1978; Lee, 2009; Massaro *et al.*, 1985). If the same perceptual mechanism is involved in musical pitch and lexical tone perception, one would expect a positive correlation between pitch height detection performance for musical notes and lexical tones.

The use of F0 height is particularly relevant when dealing with lexical tones produced by multiple speakers. Since speaking F0 range differs across individuals, absolute F0 values of a particularly tone are likely to vary by speaker. For example, a phonologically low tone produced by a female speaker could be acoustically equivalent to a phonologically high tone for a male speaker. That means F0 height judgment will have to be determined with respect to the perceived F0 range of a speaker, a process commonly referred to as speaker normalization (Johnson, 2005). Potential sources of

information for speaker normalization include syllable-internal F0 contours, external context, and familiarity with speaker voice through repeated exposure (Fox and Qi, 1990; Leather, 1983; Lin and Wang, 1984; Moore and Jongman, 1997; Wong and Diehl, 2003; Zhou *et al.*, 2008). To isolate the contribution of F0 height detection in lexical tone identification, it would be important to control for these factors.

Lee (2009) examined Mandarin tone identification from stimuli deprived of F0 contour, external context, and familiarity with speaker voice. The Mandarin syllable *sa*, produced by 32 speakers of both genders, was digitally processed to generate stimuli that included only the fricative and the first six glottal periods, thereby neutralizing the dynamic F0 contrasts among the four tones. Each stimulus was presented just once in isolation to native Mandarin listeners, who had no prior exposure to any of the speaker voices. Despite the impoverished nature of the stimuli, tone classification accuracy exceeded chance. Acoustic analyses showed that the high-onset (tones 1 and 4) and low-onset tones (tones 2 and 3) contrasted in F0, duration, and two voice quality measures (i.e., F1 bandwidth and spectral tilt). Correlation analyses showed that F0 covaried with the voice quality measures and that tone classification based on F0 height also correlated with the voice quality measures. Since the same acoustic measures consistently distinguished the female from the male stimuli, speaker gender detection was proposed to be the basis of the F0 height judgment. Similar gender-based accounts for vocal pitch detection have also been noted in literature (Deutsch, 1991; Deutsch *et al.*, 1990; Dolson, 1994; Honorof and Whalen, 2005).

What remains unresolved is whether the ability to use F0 height for lexical tone identification is associated with the ability to identify F0 height in music. Although Lee and Hung's (2008) onset-only Mandarin tone stimuli were used to evaluate the use of F0 height for lexical tone identification, those English-speaking participants were already exposed to all speaker voices from the intact and silent-center syllable presentations. In addition, none of the musicians actually possessed absolute pitch, making it difficult to interpret the lack of correlation between the performance in the absolute pitch task and the Mandarin tone task. In the current study, we extended Lee and Hung (2008) and Lee (2009) by investigating absolute pitch and identification of onset-only Mandarin tones in Mandarin-speaking musicians.

C. Summary and predictions

Our main goal in this study was to evaluate the hypothesis that absolute pitch is associated with tone language experience. To that end, two perception experiments were conducted with Mandarin-speaking musicians. In the absolute pitch task, participants were asked to identify synthesized musical tones of three timbres (piano, pure tone, and viola) in the absence of a reference pitch. It was expected that the prevalence of absolute pitch in these musicians would be comparable to the Mandarin speaker data reported in Deutsch *et al.* (2006, 2009). It was also expected that performance in the absolute pitch task would be negatively correlated with the age of onset of musical training. In the Mandarin tone task, participants were asked to identify isolated, onset-only, multispeaker Mandarin tones. It was expected that the musicians would be able to identify the tones with accuracy beyond chance, just as their nonmusician counterparts did (Lee, 2009). Finally, if a common perceptual substrate existed between pitch perception in music and language, performances in musical note identification and Mandarin tone identification should be positively correlated.

II. EXPERIMENT 1: MUSICAL NOTE IDENTIFICATION

A. Method

1. Materials

The materials used in this experiment were identical to those used in Lee and Hung (2008). In particular, 36 notes that spanned a three-octave range from C_3 (131 Hz) to B_5 (988 Hz) were synthesized with three timbres (pure tone, piano, and viola) for a total of 108 notes. The pure tones were synthesized using MATLAB (The MathWorks, Natick, MA). The piano and viola notes were synthesized with a Kurzweil K2000 synthesizer tuned to the standard A_4 at 440 Hz. The duration for all notes was 500 ms.

For each timbre, the 36 notes were ordered such that any two consecutive notes were separated by more than an octave. The purpose of the separation was to prevent listeners from developing relative pitch as a reference for the task (Deutsch *et al.*, 2006). The 36 notes of a given timbre were divided into 3 blocks of 12 notes, with a 5 s interstimulus interval and a 10 s break between the blocks. Notes of a given timbre were always presented together. Six lists of stimulus presentation were created to counterbalance the order of timbre presentation. The 72 participants were randomly assigned to 1 of the 6 lists such that a given list was used for 12 participants.

2. Participants

Seventy-two musicians were recruited from the student population at National Taiwan Normal University in Taipei, Taiwan. They included 52 females and 17 males with an age range from 18 to 24 years (mean=20.6, SD=1.6). All reported normal hearing and all identified Mandarin as their native language. The age at which they first received musical training ranged from 4 to 13 years (mean=6.7, SD=2.2). Their major instrument or area of concentration included cello (2 individuals), clarinet (6), composition (4), double bass (3), erhu (5), euphonium (1), flute (5), harp (1), horn (3), liuqin (1), oboe (2), percussion (3), piano (21), saxophone (1), trombone (1), trumpet (4), tuba (1), viola (1), violin (4), and vocal performance (3).

3. Procedure

The synthesized musical notes were saved as individual audio files and were imported to the Brown Laboratory interactive speech system (BLISS) (Mertus, 2000) for stimulus presentation. Twelve practice trials, synthesized as oboe sounds with the Kurzweil K2000 synthesizer, were given prior to the actual experiment to familiarize the participants with the presentation format. The practice notes were se-

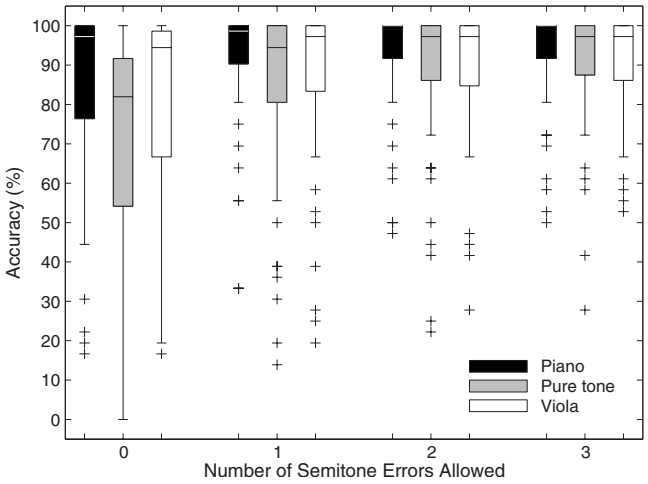


FIG. 1. A box plot showing the accuracy of musical note identification for piano, pure tone, and viola stimuli as a function of number of semitone errors allowed.

lected from the same pitch range (C_3–B_5) and were also 500 ms long. As in the actual experiment, any two consecutive notes were separated by more than an octave. None of the practice notes appeared in the actual experiment. No feedback was given.

Participants were tested individually in a sound-treated booth in the Phonetics Laboratory at National Taiwan Normal University. They listened to the stimuli through a pair of headphones (Philips SHP2500) connected to a Windows laptop computer (Panasonic CF-Y2). The participants were told that they would be listening to 9 blocks of 12 notes of 3 timbres ranging from C_3 to B_5. Their task was to notate the notes that they had heard on a customized staff paper immediately after each note was played, and to apply accidental signatures if applicable. The participants were also told that there would be 5 s to respond to each note and a 10 s break between blocks.

4. Data analysis

The written responses were graded by the second author of this article (a musician). One-way repeated measures analyses of variance (ANOVAs) were conducted on response accuracy with timbre (pure tone, piano, and viola) as a within-subject factor and participants as a random factor. To allow a direct comparison with Lee and Hung (2008), four dependent variables were used: (1) percentage of accurate note identification, (2) percentage of accurate note identification allowing one-semitone errors, (3) percentage of accurate note identification allowing two-semitone errors, and (4) percentage of accurate note identification allowing three-semitone errors. When a main effect from the ANOVAs was significant, the Bonferroni *post-hoc* test was used for pair-wise means comparisons to keep the family-wise type I error rate at 5%.

B. Results

Figure 1 shows a box plot of the accuracy of musical note identification, arranged by timbre and the number of semitone errors allowed. When an exact match was required, average accuracy was highest for piano notes (mean =86%, SD=21.3), followed by viola (mean=81%, SD =24.1) and pure tone (mean=72%, SD=25.5) notes. The ANOVA revealed a significant timbre effect [$F(2,142)=35.34$, $p<0.0001$]. All three pair-wise means comparisons were significant. Since there are 12 semitones within an octave, chance level performance is 8.3%. These averages therefore indicated that the musicians' identification accuracy was substantially beyond chance. Inspection of individual data showed that 52 of the 72 participants (72%) met the criterion for absolute pitch as defined in Deutsch *et al.* (2006), i.e., achieving at least 85% accuracy in the piano task.

When one-semitone errors were allowed, accuracy was expectedly higher. In particular, piano stimuli generated higher average accuracy (mean=91%, SD=16.3) than viola (mean=88%, SD=19.3) and pure tone (mean=86%, SD =20.9) stimuli. The ANOVA revealed a significant timbre effect [$F(2,142)=8.7$, $p<0.0005$]. Pair-wise means comparisons showed that the two comparisons involving piano were significant. Permitting one-semitone errors effectively allows a range of three semitones, indicating a chance level of 25%. These averages indicated that the musicians' identification accuracy was again substantially beyond chance, as expected from the results based on an exact match. Inspection of individual data showed that 59 of the 72 participants (82%) met the 85% accuracy criterion for absolute pitch.

When two-semitone errors were allowed, piano stimuli still generated higher average accuracy (mean=93%, SD =12.9) than viola (mean=90%, SD=15.6) and pure tone (mean=90%, SD=17.5) stimuli. The ANOVA revealed a significant timbre effect [$F(2,142)=5.68$, $p<0.005$]. Pair-wise means comparisons showed that the two comparisons involving piano were significant. Permitting two-semitone errors effectively allows a range of five semitones, indicating a chance level of 41.7%. These averages indicated that the musicians' identification accuracy was again substantially beyond chance, as expected from the results based on an exact match. We do not report the number of individuals attaining 85% accuracy when two- or three-semitone errors are allowed because accuracy measured with these deviations is no longer considered qualifying for absolute pitch.

When three-semitone errors were allowed, identification accuracy became comparable among the three timbres: piano (mean=93%, SD=11.8), viola (mean=92%, SD=12.1), and pure tone (mean=91%, SD=15). The timbre effect was not significant [$F(2,142)=3.02$, $p=0.0052$]. Permitting three-semitone errors effectively allows a range of five semitones, indicating a chance level of 58.3%. These averages indicated that the musicians' identification accuracy was again substantially beyond chance, as expected from the results based on an exact match.

To evaluate the hypothesis that absolute pitch is associated with the age of onset of musical training (Deutsch *et al.*, 2006), Pearson's correlation coefficients were derived between the four dependent variables and the age of onset of musical training. Fisher's r to z transformation was then carried out to evaluate if the correlation coefficients were significantly different from zero. Consistent with the hypoth-

esis, the age of onset of musical training was negatively correlated with all four dependent measures: accuracy requiring exact match ($r=-0.222$, $p<0.005$), allowing one-semitone errors ($r=-0.237$, $p<0.0005$), two-semitone errors ($r=-0.232$, $p<0.001$), and three-semitone errors ($r=-0.232$, $p<0.0005$). These results indicated that the earlier the onset of musical training, the better the performance in the absolute pitch task.

C. Discussion

A considerable number of the Mandarin-speaking musicians in the current study met the criterion for absolute pitch: 72% when an exact match was required, and 82% when one-semitone errors were allowed. In contrast, with identical testing materials and procedure, Lee and Hung (2008) found that none of the English-speaking musicians possessed absolute pitch. With piano stimuli only, Deutsch *et al.* (2006) found significantly higher occurrence of absolute pitch in Mandarin-speaking musicians (ranging approximately from 40% to 75%) than in English-speaking musicians (ranging approximately from 0% to 15%) depending on age of onset of musical training. Deutsch *et al.* (2009) showed that accuracy rate (not percentage of listeners with absolute pitch) was at approximately 90% for the "tone very fluent" musicians, again depending on age of onset of musical training. This value is comparable to the accuracy obtained from the Mandarin-speaking musicians in the current study (86%). Taken together, these results indicate that the occurrence of absolute pitch is positively associated with tone language fluency.

However, there are differences other than tone language experience that could account for some of the aforementioned contrasts between tone and nontone language users in the present study. Age of onset of musical training, for example, appears to be an important factor. For example, the Mandarin-speaking musicians in the current study on average started their musical training earlier (mean =6.7 years, SD=2.2) than the English-speaking musicians in Lee and Hung (2008) (mean=9.4 years, SD=3.2). Inspection of data from Deutsch *et al.* (2006, 2009) further showed that the occurrence of absolute pitch dropped significantly for musicians who started musical training after 8 or 9 years of age. Therefore, the lack of absolute pitch possessors in Lee and Hung (2008) could reflect a joint contribution of tone language experience and age of onset of musical training. Nonetheless, the negative correlation between the age of onset of musical training and musical note identification accuracy is consistent with Deutsch *et al.* (2006, 2009).

The timbre effect found in the current study contrasted with the null finding from Lee and Hung (2008). In particular, piano stimuli in the current study consistently resulted in the most accurate responses except when three-semitone errors were allowed. Viola stimuli also generated more accurate responses when an exact match was required. Compared to pure tones, complex sounds include a myriad of extraneous cues that may facilitate pitch identification (Ward, 1999). The advantage of piano stimuli, in particular, has been noted in literature (Lockhead and Byrd, 1981; Miyazaki, 1989). The current results are consistent with these findings, suggesting that the presence of timbre facilitates pitch height judgments.

III. EXPERIMENT 2: MANDARIN TONE IDENTIFICATION

A. Method

1. Materials

The materials used in this experiment were identical to the onset-only stimuli used in Lee and Hung (2008). In particular, the Mandarin syllable *sa*, produced with all 4 tones by 16 female and 16 male native speakers, was used to generate the stimuli for this experiment. The recordings were made in a sound-treated booth with an Audio-technica AT825 field recording microphone connected through a preamplifier and analog-to-digital converter (USBPre microphone interface) to a Windows personal computer. The speakers were instructed to read the syllables in citation form. The recordings were digitized with BLISS (Mertus, 2000) at 44.1 kHz with 16-bit quantization. Each syllable was identified from the BLISS waveform display, excised from the master file, and saved as an audio file. The peak amplitude was normalized across syllables.

Each *sa* syllable was digitally processed with BLISS to generate the onset-only syllables. In particular, all but the fricative and the first 15% of the voiced portion of the syllable was digitally silenced. There were no perceptible clicks as a result of the signal processing; therefore no further tapering procedure was applied. A total of 128 stimuli (4 tones $\times$ 32 speakers) were generated. Detailed acoustic analyses on these stimuli have been reported in Lee (2009) and Lee and Hung (2008).

2. Participants

The 72 musicians who participated in the musical note identification task also participated in this experiment.

3. Procedure

The stimuli were saved as individual audio files and imported to BLISS for stimulus presentation and response data acquisition. The 128 stimuli produced by the 32 speakers were assigned to 4 blocks such that each block included only 1 stimulus from a given speaker. In other words, each block had 32 stimuli and all stimuli were produced by different speakers. The purpose of this arrangement was to minimize familiarity with individual speaker voices that could be used for speaker normalization. Within each block, the number of male and female speakers was balanced (16 males and 16 females), so was the number of the four tones (8 stimuli for each of the 4 tones). Each participant was assigned a uniquely randomized presentation order such that no two participants received the same order of presentation. The order of presentation of the blocks was also randomized for each participant. A 10 s break was given between blocks.

Participants were tested individually in the same sound-treated booth with the same equipments as in the previous

experiment. The participants were told that they would be listening to the syllable *sa* with all four tones produced by 32 female and male speakers. They were also told that the syllables had been digitally processed such that only the very beginning of the syllables was audible. Their task was to identify the tone of each stimulus by pressing the buttons labeled "1," "2," "3," and "4" on the computer keyboard, representing the four Mandarin tones. All participants indicated that they were familiar with the convention of designating Mandarin tones with the numbers. The participants were further told that they had 5 s to respond to each stimulus and that their response would be timed; therefore they should response as quickly as possible.

4. Data analysis

Analyses were conducted on response accuracy, reaction time, and tone confusion patterns. Identification response and reaction time were automatically recorded by BLISS. Since the duration of individual stimulus varied, reaction time was measured from stimulus offset to avoid the potential confound of the durational differences. Missing cells due to incorrect responses constituted 2.6% of the reaction time data; only correct responses were included in the reaction time analysis. ANOVAs were conducted on accuracy and reaction time with speaker gender (female and male) and tone (1, 2, 3, and 4) as within-subject factors, and participants (F1) and stimulus items (F2) as random factors. Responses to female and male stimuli were evaluated separately to allow a direct comparison with Lee (2009). When a main effect from the ANOVAs was significant, the Bonferroni *post-hoc* test was used for pair-wise means comparisons to keep the family-wise type I error rate at 5%. Tone confusion patterns were also analyzed to evaluate the types of tone errors made by the participants. Contingency table analyses with χ^2 tests were conducted to evaluate the dependency of tone identification responses on the stimuli.

B. Results

1. Accuracy

Figure 2 shows the accuracy of tone identification responses. The ANOVAs revealed significant main effects of speaker gender [$F1(1,71)=44.77$, $p<0.0001$; $F2(1,120)=47.71$, $p<0.001$] and tone [$F1(3,213)=77.23$, $p<0.0001$; $F2(3,120)=40.5$, $p<0.0001$] and a significant gender × tone interaction [$F1(3,213)=38.47$, $p<0.0001$; $F2(3,120)=10.63$, $p<0.0001$]. In particular, female stimuli (mean=41%, SD=22.7) were identified more accurately than male stimuli (mean=35%, SD=19.2). When all stimuli were considered, tone 1 (52%, SD=20.9) and tone 3 (47%, SD=16.8) were identified more accurately than tone 4 (30%, SD=18.3) and tone 2 (22%, SD=12.9). All pair-wise means comparisons were significant except the tone 1-tone 3 comparison. The interaction arose from the result that male tone 3 was identified more accurately than female tone 3; in contrast, other female stimuli were identified more accurately than the male stimuli.

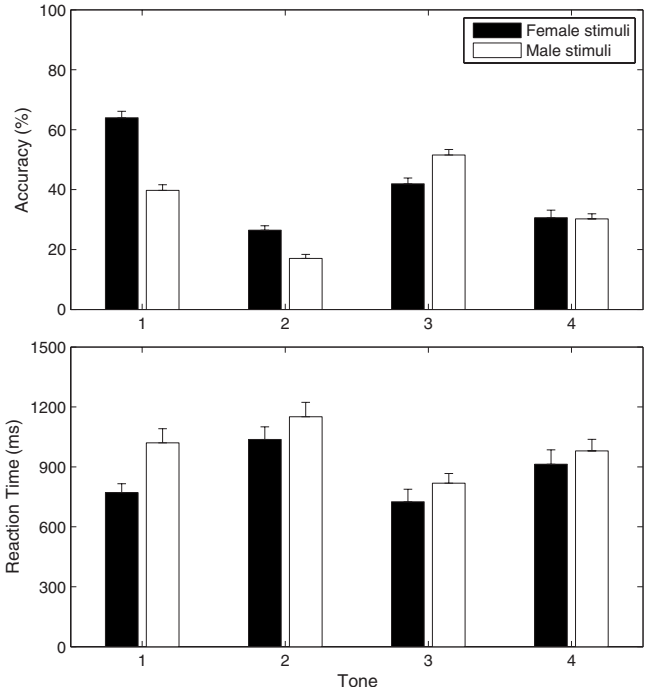


FIG. 2. Average accuracy and reaction time (+SE) of identification of onset-only Mandarin tones for female and male stimuli ($N=72$).

2. Reaction time

Figure 2 also shows the reaction time of tone identification responses. The ANOVAs revealed significant main effects of speaker gender [$F1(1,57)=26.39$, $p<0.0001$; $F2(1,120)=23.31$, $p<0.0001$] and tone [$F1(3,171)=19.17$, $p<0.0001$; $F2(3,120)=9.82$, $p<0.0001$]. There was no interaction effect. Consistent with the accuracy results, responses to female stimuli (mean=861 ms, SD=480) were faster than those to male stimuli (mean=992 ms, SD=492). When all stimuli were considered, responses to tone 3 (772 ms, SD=427) were fastest, followed by those to tone 1 (896 ms, SD=464), tone 4 (946 ms, SD=498), and tone 2 (1094 ms, SD=517). All pair-wise means comparisons were significant except the tone 1-tone 4 comparison. These results are generally consistent with the accuracy findings that higher accuracy is associated with faster reaction time.

3. Tone confusion patterns

To inspect the type of tone errors made in this task, tone responses were tabulated to generate contingency tables. To evaluate the null hypothesis that tone identification responses were not related to the stimulus tones, expected responses were also calculated for χ^2 tests. Figure 3 shows the numbers of expected and actual tone responses to the tone stimuli. The expected frequencies were calculated as follows: Suppose E_{ij} is the expected frequency for the cell in row i and column j, R_i and C_j are the corresponding row and column totals (marginal totals), and N is the total number of observations, then $E_{ij}=R_iC_j/N$ (Cohen, 1996).

Separate χ^2 tests were conducted for the female and male stimuli to allow a direct comparison with Lee (2009). Both tests showed that the null hypothesis should be re-

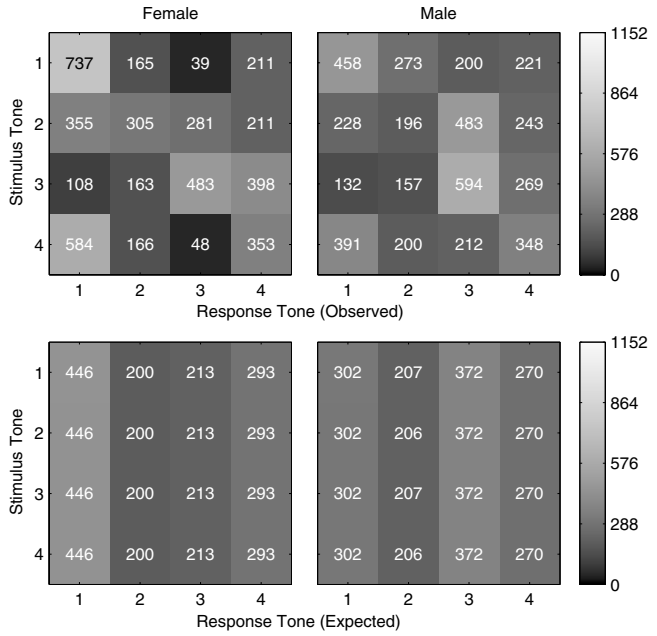


FIG. 3. Contingency tables showing the number of observed (top) and expected (bottom) Mandarin tone identification responses. The expected frequencies are theoretical frequencies derived from the row and column totals from the observed contingency tables. See text for details.

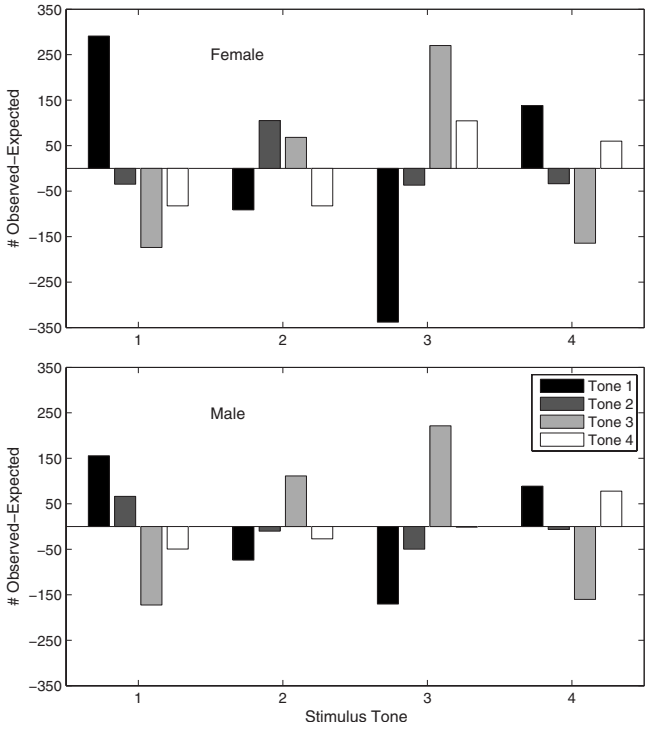


FIG. 4. Difference between the numbers of observed and expected Mandarin tone identification responses as a function of stimulus tone. Greater values indicate more confusing tone pairs, whereas smaller values indicate less confusing tone pairs. See text for details.

jected: female, χ^2 $(9, N=4607)=1311.53$, $p<0.0001$; male, χ^2 $(9, N=4605)=602.11$, $p<0.0001$. (The unequal N resulted from one and three missing responses in the female and male data, respectively.) These results indicated that the listeners' tone identification responses were related to the tone stimuli. That is, the overall accuracy was different from chance level performance.

A second set of contingency tables was generated based on specific tones. To evaluate tone 1 identification, for example, both the stimuli and responses were coded as tone 1 or nontone 1. Therefore, all responses could be classified into "hit" (tone 1 stimulus→tone 1 response), "miss" (tone 1 stimulus→nontone 1 response), "false alarm" (nontone 1 stimulus→tone 1 response), and "correct rejection" (nontone 1 stimulus→nontone 1 response). For the remaining three tones, the same coding procedure was applied to generate similar 2×2 contingency tables. As before, χ^2 tests were performed to test the null hypothesis that the responses were not related to the stimuli. For the female stimuli, all four χ^2 tests indicated that the null hypothesis should be rejected: tone 1, χ^2 $(1, N=4607)=412.81$, $p<0.0001$; tone 2, χ^2 $(1, N=4607)=89.37$, $p<0.0001$; tone 3, χ^2 $(1, N=4607)=561.15$, $p<0.0001$; tone 4, χ^2 $(1, N=4607)=21.93$, $p<0.0001$. For the male stimuli, all but the tone 2 result indicated that the null hypothesis should be rejected: tone 1, χ^2 $(1, N=4605)=144.68$, $p<0.0001$; tone 2, χ^2 $(1, N=4605)=0.83$, n.s.; tone 3, χ^2 $(1, N=4605)=259.61$, $p<0.0001$; tone 4, χ^2 $(1, N=4605)=39.04$, $p<0.0001$. In sum, these results showed that tone identification performance was above chance level except for the responses to tone 2 stimuli produced by the male speakers.

A final set of χ^2 tests was conducted based on F0 height of the tones: high-onset tones included tones 1 and 4, and low-onset tones included tones 2 and 3. As the acoustic analyses in Lee (2009) showed, dynamic F0 information was absent from these brief stimuli, but the F0 height difference remained between the high- and low-onset tones. If the identification data show that the high-low distinction can be detected beyond chance, the ability to estimate F0 height is most likely the basis for the tone identification performance. Indeed, when both the stimuli and responses were coded by F0 height, χ^2 tests showed that the high-low judgments were not random: For the female stimuli, χ^2 $(1, N=4607)=625.1$, $p<0.0001$; for the male stimuli, χ^2 $(1, N=4605)=258.49$, $p<0.0001$. These results indicate that the listeners were capable of judging the high-low distinction above chance.

The contingency tables also provided information about the confusion patterns among the four tones. To facilitate the interpretation of the patterns, Fig. 4 shows the difference between the expected and actual response counts as a function of stimulus tone. Each bar represents the result of subtracting the number of expected responses from the number of actual responses in each cell of Fig. 3. Positive or greater values indicate that the response tone is more confusable with the stimulus tone. In contrast, the smaller or more negative a value is, the less confusable the response tone is with the stimulus tone.

There are slight differences between responses to the female and male stimuli, but the overall patterns are similar. In particular, tone 1 was least often misidentified as tone 3; tone 2 was most confused with tone 3 and less often identified as tone 1 or 4; tone 3 was rarely misidentified as tone 1; tone 4 was most often misidentified as tone 1 and least misidentified as tone 3. These results are consistent with the ob-

servation that tone stimuli sharing a similar F0 height are more confusable, indicating that the high-low distinction could be detected by the listeners.

C. Discussion

The response patterns of the Mandarin-speaking musicians closely resembled those of the Mandarin-speaking nonmusicians reported in Lee (2009). In particular, they were able to identify onset-only, isolated, multispeaker Mandarin tones with accuracy beyond chance. With the exception of male tone 2 stimuli, tone identification responses were contingent on the tone stimuli, indicating the listeners' ability to distinguish between target and non-target tones. Given the brevity of the stimuli, dynamic F0 information was not available (Greenberg and Zee, 1979; Lee, 2009); therefore F0 height detection was the most likely basis for the identification performance. Indeed, the χ^2 analyses by F0 height showed that the listeners were able to distinguish between high- and low-onset tones beyond chance. This finding was further corroborated by the confusion pattern that tones with distinct F0 heights tended to be less confusable. In sum, musicians or not, native Mandarin listeners were able to use F0 height for the identification of these brief Mandarin tones.

IV. CORRELATION BETWEEN MANDARIN TONE IDENTIFICATION AND MUSICAL NOTE IDENTIFICATION

Since F0 height detection was involved in both musical note identification and Mandarin tone identification, did those musicians with absolute pitch perform better in the Mandarin tone identification task? Correlation analyses were conducted to quantitatively evaluate the relationship between the musical note and Mandarin tone identification performances. Pearson's correlation coefficients were derived between accuracy of musical note identification (averaged across all three timbres; exact match required or one-semitone errors allowed) and that of Mandarin tone identification (averaged across all stimuli). Fisher's r to z transformation was carried out to evaluate if the correlation coefficients were significantly different from zero. Two sets of analyses were conducted, one with all musicians (72 individuals) and the other with those who possessed absolute pitch (the 53 individuals who achieved an average accuracy of 85% or higher in musical note identification).

The correlation coefficients were generally low and none of them turned out significantly different from zero. In particular, with all musicians included, the correlation coefficients were −0.03 (exact match required) and −0.03 (one-semitone errors allowed). When only individuals with absolute pitch were included, the coefficients were −0.16 (exact match required) and −0.07 (one-semitone errors allowed). In other words, we did not find evidence that performance in the musical note identification task was associated with performance in the Mandarin tone identification task.

V. GENERAL DISCUSSION

We evaluated Mandarin-speaking musicians' identification of musical notes without a reference pitch and their identification of onset-only, isolated, multispeaker Mandarin tone stimuli that required detection of F0 height. The data supported several predictions made earlier. First, absolute pitch was quite prevalent in the Mandarin-speaking musicians, particularly when compared to the English-speaking musicians tested with the same materials and procedure (Lee and Hung, 2008). The percentage of absolute pitch possessors among the Mandarin-speaking musicians was also comparable to the Mandarin-speaking musicians in Deutsch *et al.* (2006) and the tone very fluent musicians in Deutsch *et al.* (2009). These results, based on samples of musicians drawn from China, Taiwan, and the United States, consistently suggest that absolute pitch is associated with tone language experience; i.e., absolute pitch is more prevalent in musicians who speak a tone language fluently. As Deutsch *et al.* (2009) noted, genetic factors are not likely the reason for the contrast because the "tone nonfluent" musicians of East Asian heritage in their study showed comparably low accuracy as the Caucasian musicians did in the absolute pitch task. The country where the musicians received their education is also not likely the cause because the tone very fluent United States musicians displayed comparable level of performance as the musicians did who were educated in China.

Second, performance in the absolute pitch task was negatively correlated with the age of onset of musical training. This finding is consistent with Deutsch *et al.* (2006, 2009) showing greater occurrences of absolute pitch in musicians who started their musical training early. The age of onset effect also supports the idea that the acquisition of absolute pitch is analogous to the acquisition of speech in that a critical period is implicated (Deutsch *et al.*, 2006). As noted earlier, the age of onset of musical training could have contributed to the absence of absolute pitch in the English-speaking musicians in Lee and Hung (2008). In addition, Deutsch *et al.* (2006) showed that absolute pitch was consistently more prevalent in Mandarin-speaking musicians when the age of onset of musical training was controlled, further affirming the contribution of tone language experience to absolute pitch.

Third, the Mandarin-speaking musicians were able to identify the onset-only, isolated, multispeaker Mandarin tone stimuli with accuracy beyond chance, just as their nonmusician counterparts did (Lee, 2009). As noted, tone identification from these brief stimuli was quite challenging because dynamic F0 information was neutralized. Consequently, listeners had to rely primarily on F0 height for tone identification. Furthermore, the decision of whether a tone is phonologically high or low had to be made with respect to individual speaking F0 range, which varied across speakers. For example, Lee's (2009) acoustic analyses showed considerable overlap between the female low-onset tones and the male high-onset tones. To decipher these acoustically similar tones, listeners had to come up with some estimate of speaking F0 range in order to decide whether a tone is phonologically high or low. Since the tone stimuli were presented in isolation and without repetition, there was no external context or familiarity with speaker voices to facilitate F0 range estimation. Nonetheless, both Mandarin-speaking nonmusicians (Lee, 2009) and musicians (the current study) were

able to identify the tones of the stimuli with accuracy exceeding chance, indicating their ability to judge F0 height from these rather impoverished tone stimuli.

Since F0 height judgment was involved in both absolute pitch and Mandarin tone identification, it seems reasonable to assume that performance in the two tasks would be correlated. That is, musicians who scored higher in the absolute pitch task should be more accurate in the Mandarin tone identification task. However, results from the correlation analyses did not support this prediction. Correlation coefficients, whether obtained with all musicians or just those with absolute pitch, were generally low and did not differ significantly from zero. One possibility for the lack of correlation is that there was not sufficient performance variability in the absolute pitch and Mandarin tone identification measures. In particular, accuracy in the absolute pitch task was generally high and only a small number of individuals did not qualify for absolute pitch. In contrast, Mandarin tone identification accuracy was generally low even though it did exceed chance. Perhaps one or both of these measures were simply not sensitive enough to detect the correlation.

However, the nonsignificant correlation could also reflect a true lack of association between absolute pitch and Mandarin tone identification. That is, even though both require associating pitch with a verbal label, distinct processing mechanisms may be implicated. Considering the functional difference between absolute pitch and lexical tone, this seems to be an equally plausible interpretation. In particular, while absolute pitch involves associating a particular pitch with a verbal label, lexical tone involves associating a pitch pattern and other acoustic cues to phonetic categories that bear linguistic significance. In addition to F0, other acoustic parameters such as duration and amplitude also contribute to the perception of lexical tone (e.g., Whalen and Xu, 1992; Liu and Samuel, 2004). In other words, lexical tone perception is not just pitch perception, rather it involves mapping all relevant acoustic information from the output of a vocal tract onto linguistically significant phonetic categories.

Furthermore, all normally developing speakers of a tone language will acquire the association between acoustic correlates of lexical tone and tonal categories without explicit instruction. The verbal labels for absolute pitch, however, have to be taught. Anecdotal experience also suggests that many tone language users are not able to identify or produce musical pitch reliably (e.g., singing), yet they have no trouble identifying or producing lexical tones. These observations suggest that linguistic and nonlinguistic pitches may involve distinct processing mechanisms. These observations are also consistent with behavioral and neurophysiological evidence showing that how an acoustic attribute is processed depends on whether it is interpreted linguistically (e.g., Remez *et al.*, 1981; Gandour *et al.*, 1998).

However, the above interpretation does not imply that pitch processing in music and lexical tone is completely dissociated. Many studies have shown the advantage of musical training in lexical tone production and perception (Alexander *et al.*, 2005; Gottfried, 2007; Gottfried and Riester, 2000; Gottfried *et al.*, 2001; Lee and Hung, 2008; Wong *et al.*, 2007). Lee and Hung's (2008) results are particularly worth revisiting because the same test materials and procedure were used as in the current study. In particular, they showed that English-speaking musicians were more accurate than nonmusicians in identifying Mandarin tones from intact and silent-center syllables. However, none of the musicians were found to possess absolute pitch, suggesting that their superior performance over nonmusicians did not arise from absolute pitch. Interestingly, the musicians were no more accurate than the nonmusicians in the onset-only tone stimuli, which, as we noted, were deprived of F0 contour and had to be identified based on F0 height. It would be of interest to find out if English-speaking musicians who actually possess absolute pitch will outperform nonmusicians in identifying onset-only Mandarin tones.

The finding from the current Mandarin tone experiment replicated Lee (2009) and is consistent with the idea that speakers of a linguistic community can acquire pitch class templates through long-term exposure to the prevalent speaking F0 of the community, and that these templates can influence both speech production and perception (Deutsch, 1991; Deutsch *et al.*, 2004b, 2009, 1990; Dolson, 1994; Honorof and Whalen, 2005). In particular, to judge whether a tone is high or low, a listener will have to consider the speaking F0 range of a speaker because a phonologically low tone for a female speaker could be acoustically equivalent to a phonologically high tone for a male speaker (Lee, 2009). In the absence of cues commonly considered necessary for speaker normalization (F0 contour, external context, and familiarity with speaker voice through repeated exposure), Mandarin speakers were nonetheless able to identify the tones beyond chance. A similar result was obtained by Honorof and Whalen (2005), who showed that English-speaking listeners were able to locate an F0 reliably within a speaker's F0 range without context or prior exposure to the speaker's voice. Since F0 height judgments were positively correlated with the female and male F0 range values, Honorof and Whalen (2005) noted that the ability to detect speaker gender could be the basis for the F0 judgment performance.

This speculation was supported by Lee's (2009) acoustic data and correlation analyses. In particular, Lee (2009) showed that F0 covaried with two voice quality measures (F1 bandwidth and spectral tilt) and that Mandarin tone classification based on F0 height also correlated with these acoustic measures. These results suggest that the Mandarin tone identification performance could be due to the listeners' successful identification of speaker gender as a precursor. That is, listeners could identify speaker gender based on the voice quality measures and then exploit the covariation between voice quality and F0 for tone identification. Once the gender decision is made, pitch class templates that are gender-specific could be invoked to compare to the tone stimuli. The height of the tone stimuli and thus tone identity could then be inferred with the templates as a reference frame.

VI. CONCLUSIONS

With the use of pitch for lexical contrasts, tone languages provide a unique opportunity for exploring the

music-language relationship. The present study contributes to literature by providing further evidence on the relationship between absolute pitch and lexical tone perception. We showed that the majority of the Mandarin-speaking musicians met the criteria for absolute pitch in the musical note identification experiment. The performance in the musical note identification task also correlated negatively with the age of onset of musical training. Like their nonmusician counterparts (Lee, 2009), the musicians were able to identify onset-only, isolated, multispeaker Mandarin tones beyond chance. Although the ability to detect F0 height was implicated in both musical note identification and Mandarin tone identification, performances in the two tasks were not correlated. With results from this series of studies (Lee, 2009; Lee and Hung, 2008), we suggest that speaker gender identification is a precursor to Mandarin tone identification from multispeaker input, and that musical abilities other than absolute pitch contribute to the musicians' superior performance in Mandarin tone identification.

ACKNOWLEDGMENTS

We would like to thank Naomi Ogasawara of National Taiwan Normal University for providing sound booths for the perception experiments. We also thank Diana Deutsch and an anonymous reviewer for their helpful comments. This research was supported by professional development funds from Ohio University.

Alexander, J., Wong, P. C. M., and Bradlow, A. (**2005**). "Lexical tone perception in musicians and nonmusicians," Proceedings of Interspeech' 2005—Eurospeech–9th European Conference on Speech Communication and Technology, Lisbon, Portugal.

Burnham, D., Peretz, I., Stevens, K., Jones, C., Schwanhäusser, B., Tsukada, K., and Bollwerk, S. (**2004**). "Do tone language speakers have perfect pitch?," in *Proceedings of the 8th International Conference on Music Perception and Cognition*, Evanston, IL, edited by S. D. Libscomb, R. Ashley, R. O. Gjerdingen, and P. Webster (Causal Productions, Adelaide, Australia).

Cohen, B. H. (**1996**). *Explaining Psychological Statistics* (Brooks/Cole, Pacific Grove, CA).

Deutsch, D. (**1991**). "The tritone paradox: An influence of language on music perception," Music Percept. **8**, 335–347.

Deutsch, D. (**2002**). "The puzzle of absolute pitch," Curr. Dir. Psychol. Sci. **11**, 200–204.

Deutsch, D. (**2006**). "The enigma of absolute pitch," Acoust. Today **2**, 11–19.

Deutsch, D., Dooley, K., Henthorn, T., and Head, B. (**2009**). "Absolute pitch among students in an American music conservatory: Association with tone language fluency," J. Acoust. Soc. Am. **125**, 2398–2403.

Deutsch, D., Henthorn, T., and Dolson, M. (**2004a**). "Absolute pitch, speech, and tone language: Some experiments and a proposed framework," Music Percept. **21**, 339–356.

Deutsch, D., Henthorn, T., and Dolson, M. (**2004b**). "Speech patterns heard early in life influence later perception of the tritone paradox," Music Percept. **21**, 357–372.

Deutsch, D., Henthorn, T., Marvin, E., and Xu, H. (**2006**). "Absolute pitch among American and Chinese conservatory students: Prevalence differences, and evidence for a speech-related critical period," J. Acoust. Soc. Am. **119**, 719–722.

Deutsch, D., North, T., and Ray, L. (**1990**). "The tritone paradox: Correlate with the listener's vocal range for speech," Music Percept. **7**, 371–384.

Dolson, M. (**1994**). "The pitch of speech as a function of linguistic community," Music Percept. **11**, 321–331.

Fox, R. A., and Qi, Y.-Y. (**1990**). "Context effects in the perception of lexical tones," J. Chin. Linguist. **18**, 261–284.

Gandour, J. (**1983**). "Tone perception in Far Eastern languages," J. Phonetics **11**, 149–175.

Gandour, J., and Harshman, R. A. (**1978**). "Crosslanguage differences in tone perception: A multidimensional scaling investigation," Lang Speech **22**, 1–33.

Gandour, J., Wong, D., and Hutchins, G. (**1998**). "Pitch processing in the human brain is influenced by language experience," NeuroReport **9**, 2115–2119.

Gottfried, T. L. (**2007**). "Effects of musical training on learning L2 speech contrasts," in *Language Experience in Second Language Speech Learning*, edited by O.-S. Bohn and M. J. Munro (John Benjamins, Amsterdam), pp. 221–237.

Gottfried, T. L., and Riester, D. (**2000**). "Relation of pitch glide perception and Mandarin tone identification," J. Acoust. Soc. Am. **108**, 2604.

Gottfried, T. L., Staby, A. M., and Ziemer, C. J. (**2001**). "Musical experience and Mandarin tone discrimination and imitation," J. Acoust. Soc. Am. **115**, 2545.

Greenberg, S., and Zee, E. (**1979**). "On the perception of contour tones," UCLA Working Papers in Phonetics **45**, 150–164.

Honorof, D. N., and Whalen, D. H. (**2005**). "Perception of pitch location within a speaker's F0 range," J. Acoust. Soc. Am. **117**, 2193–2200.

Johnson, K. A. (**2005**). "Speaker normalization in speech perception," in *The Handbook of Speech Perception*, edited by D. B. Pisoni and R. E. Remez (Blackwell, Malden, MA), pp. 363–389.

Leather, J. (**1983**). "Speaker normalization in perception of lexical tone," J. Phonetics **11**, 373–382.

Lee, C.-Y. (**2009**). "Identifying isolated, multispeaker Mandarin tones from brief acoustic input: A perceptual and acoustic study," J. Acoust. Soc. Am. **125**, 1125–1137.

Lee, C.-Y., and Hung, T.-H. (**2008**). "Identification of Mandarin tones by English-speaking musicians and nonmusicians," J. Acoust. Soc. Am. **124**, 3235–3248.

Lee, C.-Y., Tao, L., and Bond, Z. S. (**2008**). "Identification of acoustically modified Mandarin tones by native listeners," J. Phonetics **36**, 537–563.

Lee, C.-Y., Tao, L., and Bond, Z. S. (**2009**). "Speaker variability and context in the identification of fragmented Mandarin tones by native and non-native listeners," J. Phonetics **37**, 1–15.

Lee, C.-Y., Tao, L., and Bond, Z. S. "Identification of acoustically modified Mandarin tones by non-native listeners," Lang. Speech **53** (to be published).

Lin, T., and Wang, W. S.-Y. (**1984**). "Shengdiao ganzhi wenti (The issue of tone perception)," Zhongguo Yuyan Xuebao (Bull. Chinese Linguistics), **2**, 59–69.

Liu, S., and Samuel, A. G. (**2004**). "Perception of Mandarin lexical tones when F0 information is neutralized," Lang Speech **47**, 109–138.

Lockhead, G. R., and Byrd, R. (**1981**). "Practically perfect pitch," J. Acoust. Soc. Am. **70**, 387–389.

Massaro, D. W., Cohen, M. M., and Tseng, C.-Y. (**1985**). "The evaluation and integration of pitch height and pitch contour in lexical tone perception in Mandarin Chinese," J. Chin. Linguist. **13**, 267–289.

Mertus, J. A. (**2000**). *The Brown Lab Interactive Speech System* (Brown University, Rhode Island).

Miyazaki, K. (**1989**). "Absolute pitch identification: Effects of timbre and pitch region," Music Percept. **7**, 1–14.

Moore, C. B., and Jongman, A. (**1997**). "Speaker normalization in the perception of Mandarin Chinese tones," J. Acoust. Soc. Am. **102**, 1864–1877.

Patel, A. D. (**2008**). *Music, Language, and the Brain* (Oxford University Press, New York).

Profita, J., and Bidder, T. G. (**1988**). "Perfect pitch," Am. J. Med. Genet. **29**, 763–771.

Remez, R. E., Rubin, P. E., Pisoni, D. B., and Carrell, T. D. (**1981**). "Speech perception without traditional speech cues," Science **212**, 947–950.

Ward, W. D. (**1999**). "Absolute pitch," in *The Psychology of Music*, 2nd ed., edited by D. Deutsch (Academic, New York), pp. 265–298.

Whalen, D. H., and Xu, Y. (**1992**). "Information for Mandarin tones in the amplitude contour and in brief segments," Phonetica **49**, 25–47.

Wong, P. C. M., and Diehl, R. L. (**2003**). "Perceptual normalization for inter- and intra-talker variation in Cantonese level tones," J. Speech Lang. Hear. Res. **46**, 413–421.

Wong, P. C. M., Skoe, E., Russo, N. M., Dees, T., and Kraus, N. (**2007**). "Musical experience shapes human brainstem encoding of linguistic pitch patterns," Nat. Neurosci. **10**, 420–422.

Zhou, N., Zhang, W., Lee, C.-Y., and Xu, L. (**2008**). "Lexical tone recognition with an artificial neural network," Ear Hear. **29**, 326–335.

Major and minor music compared to excited and subdued speech

Daniel L. Bowling, Kamraan Gill, Jonathan D. Choi, Joseph Prinz, and Dale Purves
Department of Neurobiology and Center for Cognitive Neuroscience, Duke University Durham, North Carolina 27708

(Received 3 June 2009; revised 15 October 2009; accepted 16 October 2009)

The affective impact of music arises from a variety of factors, including intensity, tempo, rhythm, and tonal relationships. The emotional coloring evoked by intensity, tempo, and rhythm appears to arise from association with the characteristics of human behavior in the corresponding condition; however, how and why particular tonal relationships in music convey distinct emotional effects are not clear. The hypothesis examined here is that major and minor tone collections elicit different affective reactions because their spectra are similar to the spectra of voiced speech uttered in different emotional states. To evaluate this possibility the spectra of the intervals that distinguish major and minor music were compared to the spectra of voiced segments in excited and subdued speech using fundamental frequency and frequency ratios as measures. Consistent with the hypothesis, the spectra of major intervals are more similar to spectra found in excited speech, whereas the spectra of particular minor intervals are more similar to the spectra of subdued speech. These results suggest that the characteristic affective impact of major and minor tone collections arises from associations routinely made between particular musical intervals and voiced speech.
 [DOI: 10.1121/1.3268504]

PACS number(s): 43.75.Cd, 43.75.Bc, 43.71.Es [DD] Pages: 491–503

I. INTRODUCTION

The affective impact of music depends on many factors including, but not limited to, intensity, tempo, rhythm, and the tonal intervals used. For most of these factors the way emotion is conveyed seems intuitively clear. If, for instance, a composer wants to imbue a composition with excitement, the intensity tends to be forte, the tempo fast, and the rhythm syncopated; conversely, if a more subdued effect is desired, the intensity is typically piano, the tempo slower, and the rhythm more balanced (Cohen, 1971; Bernstein, 1976; Juslin and Laukka, 2003). These effects on the listener presumably occur because in each case the characteristics of the music accord with the ways that the corresponding emotional state is expressed in human behavior. The reason for the emotional effect of the tonal intervals used in music, however, is not clear.

Much music worldwide employs subsets of the chromatic scale, which divides each octave into 12 intervals defined by the frequency ratios listed in Table IA (Nettl, 1956; Randel, 1986; Carterette and Kendall, 1999; Burkholder *et al.*, 2005). Among the most commonly used subsets are the diatonic scales in Table IB (Pierce, 1962; Bernstein, 1976; Randel, 1986; Burns, 1999; Burkholder *et al.*, 2005). In recent centuries, the most popular diatonic scales have been the Ionian and the Aeolian, usually referred to simply as the major and the minor scale, respectively (Aldwell and Schachter, 2003) (Fig. 1). Other things being equal (e.g., intensity, tempo, and rhythm), music using the intervals of the major scale tends to be perceived as relatively excited, happy, bright, or martial, whereas music using minor scale intervals tends to be perceived as more subdued, sad, dark, or wistful (Zarlino, 1571; Hevner, 1935; Cooke, 1959; Crowder, 1984; Krumhansl, 1990; Gregory and Varney, 1996; Peretz *et al.*, 1998; Burkholder *et al.*, 2005). There has been, however, no agreement about how and why these scales and the intervals that differentiate them elicit distinct emotional effects (Heinlein, 1928; Hevner, 1935; Crowder, 1984; Carterette and Kendall, 1989; Schellenberg *et al.*, 2000; Gabrielsson and Lindström, 2001). Based on the apparent role of behavioral mimicry in the emotional impact produced by other aspects of music, we here examine the hypothesis that the same framework pertains to the affective impact of the different tone collections used in melodies.

To evaluate the merits of this hypothesis, we compared the spectra of the intervals that, based on an empirical analysis, specifically distinguish major and minor music with the spectra of voiced speech uttered in different emotional states. There are several reasons for taking this approach. First, voiced speech sounds are harmonic and many of the ratios between the overtones in any harmonic series correspond to musical ratios (Helmholtz 1885; Bernstein, 1976; Rossing 1990; Crystal, 1997; Stevens, 1998; Johnston, 2002). Second, most of the frequency ratios of the chromatic scale are statistically apparent in voiced speech spectra in a variety of languages (Schwartz *et al.*, 2003). Third, we routinely extract biologically important information about the emotional state of a speaker from the quality of their voice (Johnstone and Scherer, 2000; Scherer *et al.*, 2001; Juslin and Laukka, 2003; Thompson and Balkwill, 2006). Fourth, the physiological differences between excited and subdued affective states alter the spectral content of voiced speech (Spencer, 1857; Juslin and Laukka, 2003; Scherer, 2003). Fifth, with few exceptions, the only tonal sounds in nature are the vocalizations of animals, the most important of these being conspecific (Schwartz *et al.*, 2003). Sixth, a number of non-musical phe-

TABLE I. Western musical scales (also called modes). (A) The 12 intervals of the chromatic scale showing the abbreviations used, the corresponding number of semitones, and the ratio of the fundamental frequency of the upper tone to the fundamental frequency of the lower tone in just intonation tuning. (B) The seven diatonic scales/modes. As a result of their relative popularity, the Ionian and the Aeolian modes are typically referred to today as the major and minor scales, respectively. Although the Ionian and Aeolian modes and the scales they represent have been preeminent in Western music since the late 16th century, some of the other scales/modes continue to be used today. For example, the Dorian mode is used in plainchant and some folk music, the Phrygian mode is used in flamenco music, and the Mixolydian mode is used in some jazz. The Locrian and Lydian are rarely used because the dissonant tritone takes the place of the fifth and fourth scale degrees, respectively.

(A) Chromatic scale			(B) Diatonic scales						
Interval Name	Semitones	Frequency ratio	"MAJOR" Ionian	Dorian	Phrygian	Lydian	Mixolydian	"MINOR" Aeolian	Locrian
Unison (Uni)	0	1:1	M2	M2	m2	M2	M2	M2	m2
Minor second (m2)	1	16:15	M3	m3	m3	M3	M3	m3	m3
Major second (M2)	2	9:8	P4	P4	P4	tt	P4	P4	P4
Minor third (m3)	3	6:5	P5	P5	P5	P5	P5	P5	tt
Major third (M3)	4	5:4	M6	M6	m6	M6	M6	m6	m6
Perfect fourth (P4)	5	4:3	M7	m7	m7	M7	m7	m7	m7
Tritone (tt)	6	7:5	Oct	Oct	Oct	Oct	Oct	Oct	Oct
Perfect fifth (P5)	7	3:2							
Minor sixth (m6)	8	8:5							
Major sixth (M6)	9	5:3							
Minor seventh (m7)	10	9:5							
Major seventh (M7)	11	15:8							
Octave (Oct)	12	2:1							

nomena in pitch perception, including perception of the missing fundamental, the pitch shift of the residue, spectral dominance, and pitch strength can be rationalized in terms of spectral similarity to speech (Terhardt, 1974; Schwartz and Purves, 2004). And finally, as already mentioned, other aspects of music appear to convey emotion through mimicry of human behaviors that signify emotional state. It therefore makes sense to ask whether spectral differences that specifically distinguish major and minor melodies parallel spectral differences that distinguish excited and subdued speech.

II. METHODS

A. Overview

The intervals that distinguish major and minor music were determined from classical and folk melodies composed in major and minor keys. The notes in these melodies were extracted and the intervals they represent were calculated and tallied by condition. Excited and subdued speech samples were obtained by recording single words and monologs spoken in either an excited or subdued manner. From these recordings only the voiced segments were extracted and analyzed. The spectra of the distinguishing musical intervals were then compared with the spectra of the voiced segments in excited and subdued speech according to fundamental frequency and frequency ratios.

B. Acquisition and analysis of the musical databases

A database of classical Western melodies composed in major and minor keys over the past three centuries was compiled from the electronic counterpart of Barlow and Morgenstern's Dictionary of Musical Themes (Barlow and Morgenstern, 1948). The online version (http://www.multimedialibrary.com/Barlow) includes 9825 monophonic

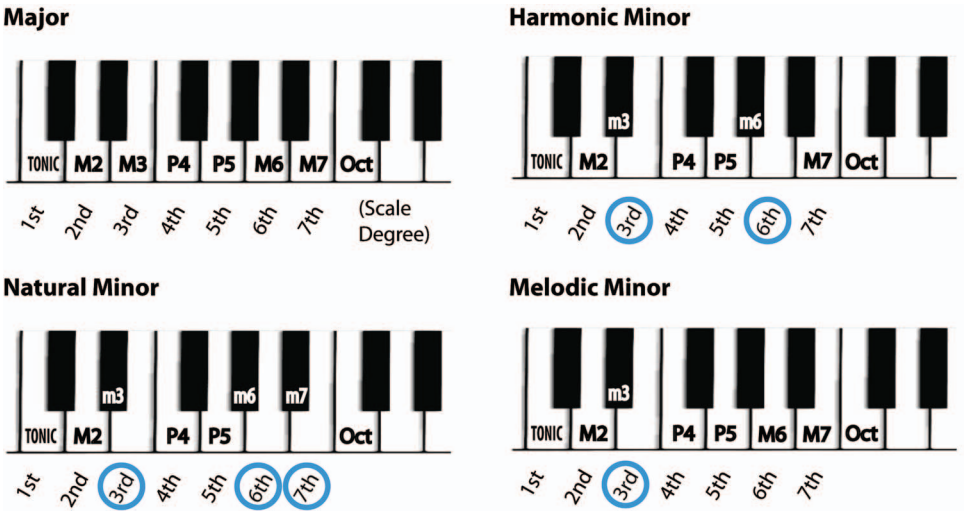
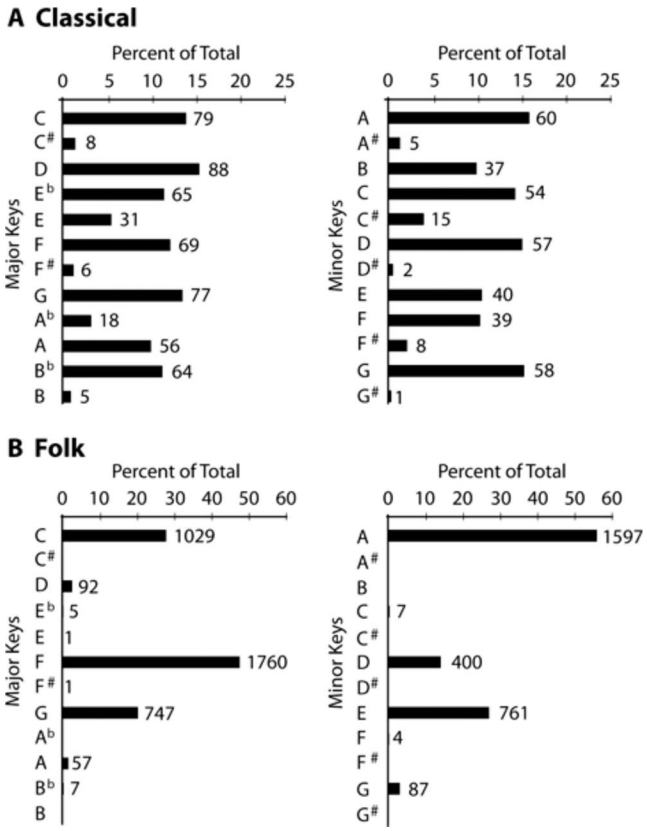


FIG. 1. The major and minor scales in Western music, shown on a piano keyboard (the abbreviations follow those in Table IA). The diatonic major scale contains only major intervals, whereas the all three diatonic minor scales substitute a minor interval at the third scale degree. Two also substitute a minor interval at the sixth scale degree, and one a minor interval at the seventh scale degree (the melodic minor scale is shown ascending; when descending it is identical to the natural minor). Thus the formal differences between the major and minor scales are in the third, sixth, and seventh scale degrees; the first, second, fourth, and fifth scale degrees are held in common.

TABLE II. The distribution of the key signatures of the major and minor melodies we compiled from Barlow and Morgenstern's Dictionary of Musical Themes (A) and the Finnish Folk Song Database (B). The number of melodies analyzed in each key is indicated to the right of each bar.



melodies in MIDI format comprising 7–69 notes. From these, we extracted 971 initial themes from works that were explicitly identified in the title as major or minor (themes from later sections were often in another key). The mean number of notes in the themes analyzed was 19.9 (SD =7.5). To ensure that each initial theme corresponded to the key signature in the title, we verified the key by inspection of the score for accidentals; 29 further themes were excluded on this basis. Applying these criteria left 942 classical melodies for analysis of which 566 were major and 376 minor. The distribution of key signatures for these melodies is shown in Table IIA.

To ensure that our conclusions were not limited to classical music, we also analyzed major and minor melodies in the Finnish Folk Song Database (Eerola and Toiviainen, 2004a). These melodies are from the traditional music of that region published in the late 19th century and first half of the 20th century (many of the themes derive from songs composed in earlier centuries). The full database contains 8614 songs in MIDI format comprising 10–934 notes each annotated by key and designated as major or minor. To make our analysis of folk songs as comparable as possible to the analysis of classical melodies and to avoid modulations in later sections of the pieces, we excluded all songs comprising more than 69 notes [the maximum number in the classical database melodies; the mean number of notes for the Finnish songs we analyzed was 39.0 (SD=13.2)]. We also excluded any songs that were annotated as being in more than one key or that were polyphonic. Applying these criteria left 6555 melodies for analysis of which 3699 were major and 2856 minor. The distribution of key signatures for these melodies is shown in Table IIB.

To assess the tonal differences between the major and minor melodies, the chromatic intervals represented by each melody note were determined (1) with respect to the annotated tonic of the melody and (2) with respect to the immediately following melody note. Chromatic intervals based on the tonic (referred to as "tonic intervals") were defined as ascending with respect to the tonic by counting the number of semitones between each note in a melody and its annotated tonic (see Table IA); tonic intervals larger than an octave were collapsed into a single octave. Chromatic Intervals between melody notes (referred to as "melodic intervals") were determined by counting the number of semitones between each note in a melody and the next. The number of occurrences of each chromatic interval was tabulated separately for tonic and melodic intervals in major and minor classical and folk music using MIDI toolbox version 1.0.1 (Eerola and Toiviainen, 2004b) for MATLAB version R2007a (Mathworks Inc., 2007). The associated frequencies of occurrence were then calculated as percentages of the total number of intervals counted. For example, the frequency of tonic major thirds in classical major melodies was determined by dividing the total number of tonic major thirds in these melodies by the total number of all tonic intervals in these melodies.

The MIDI melodies were coded in equal tempered tuning; however, we converted all intervals to just intonation for analysis since this tuning system is generally considered more "natural" than equal temperament (see Sec. IV).

C. Recording and analysis of voiced speech

Single word utterances were recorded from ten native speakers of American English (five females) ranging in age from 18–68 and without significant speech or hearing pathology. The participants gave informed consent, as required by the Duke University Health System. Monologs were also recorded from ten speakers (three of the original participants plus seven others; five females). Single words enabled more accurate frequency measurements, whereas the monologs produced more typical speech. All speech was recorded in a sound-attenuating chamber using an Audio-Technica AT4049a omni-directional capacitor microphone and a Marantz PMD670 solid-state digital recorder (Martel Electronics, Yorba Linda CA). Recordings were saved to a Sandisk flash memory card in .wav format at a sampling rate of 22.05 kHz and transferred to an Apple PowerPC G5 computer for analysis. The background noise in the recording chamber was measured in one-third-octave bands with center frequencies from 12.5 Hz to 20 kHz, using a Larson–Davis System 824 Real Time Analyzer averaging over periods of 10 s. The noise level was less than NC-15 for frequencies up to and including 500 Hz, falling to ~NC-25 by 5 kHz (NC-20 is a typical specification for an empty concert hall).

The spectra were analyzed using the "to pitch" autocor-

relation algorithm (Boersma 1993), the "to formant" LPC algorithm (Press *et al.*, 1992), and the "get power" algorithm in PRAAT version 5.0.43 (Boersma and Weenik, 2008). PRAAT default settings were used for both pitch and formant estimates, with the exception of the pitch ceiling which was set at 500 rather than 600 Hz (for pitch: floor=75 Hz, ceiling=500 Hz, time step=10 ms, window length=40 ms; for formants: number=5, ceilings for male/female =5 kHz/5.5 kHz, respectively, window length=50 ms; since there is no standard window length for power analyses we determined power every 10 ms using a 10 ms window).

Single words. Ten words that each had a different vowel embedded between the consonants /b/ and /d/ (i.e., bead, bid, bed, bad, bod, bud, booed, bawd, bird, and "bood," the last pronounced like "good") were repeated by participants. These vowels (/i, ɪ, ɛ, œ, ɑ, ʌ, u, ɔ, ɝ, ʊ/) were chosen as a representative sample of voiced speech sounds in English; the consonant framing (/b···d/) was chosen because it maximizes vowel intelligibility (Hillenbrand and Clark, 2000). Each participant repeated each word seven times so that we could analyze the central five utterances, thus avoiding onset and offset effects. This sequence was repeated four times (twice in each emotional condition) using differently ordered lists of the words, pausing for 30 s between the recitations of each word list. Participants were instructed to utter the words as if they were excited and happy, or conversely as if they were subdued and sad. The quality of each subject's performance was monitored remotely to ensure that the speech was easily recognized as either excited or subdued. The fundamental frequency values were checked for accuracy by comparing the pitch track in PRAAT with the spectrogram. In total, 1000 examples of single word vowel sounds in each experimental condition were included in the database (i.e., 100 from each speaker).

A PRAAT script was used to automatically mark the pauses between each word; vowel identifier and positional information were then inserted manually for each utterance (the script can be accessed at http://www.helsinki.fi/~lennes/praat-scripts/public/mark_pauses.praat). A second script was written to extract the fundamental frequency (using "to pitch") and the frequencies of the first and second formants (using "to formant") from a 50 ms segment at the mid-point of each vowel utterance.

Monologs. The participants read five monologs with exciting content and five monologs with subdued content (Table III). Each monolog was presented for reading in a standard manner on a computer monitor and comprised about 100 syllables with an approximately equal distribution of the ten different American English vowels examined in the single word analysis. The instructions given the participants were to utter the monologs with an emotional coloring appropriate to the content after an initial silent reading. Performance quality was again monitored remotely.

Analysis of the monologs was complicated by the fact that natural speech is largely continuous; thus the "mark pauses" script could not be used. Instead we wrote another PRAAT script to extract the fundamental frequency, first and second formant frequencies, and amplitude values using a time step of 10 ms and the PRAAT functions listed above (amplitude was obtained by taking the square root of the values returned by "get power"). In order to avoid the inclusion of silent intervals, time points where the amplitude was less than 5% of a speaker's maximum were rejected. To avoid including unvoiced speech segments, time points at which no fundamental frequency was identified were also removed (as the to pitch function steps through the speech signal it assigns time points as either voiced or unvoiced depending on the strength of identified fundamental frequency candidates; for details, see Boersma, 1993). On average, 32% of the monolog recordings was silence and 13% was unvoiced; thus 55% of the material was voiced and retained. Each monolog was spoken over 10–20 s and yielded approximately 380 voiced data points for analysis.

TABLE III. Examples of the excited and subdued monologues read by the speakers.

(A) Excited
I won the lottery and I still can't believe it! I always bought a ticket but you never expect to win. I don't known what I'm going to do with ten million dollars but I'm sure going to have fun finding out. I've had some good luck in my day, but this tops it all.
(B) Subdued
The papers finalizing the divorce came today. Sometimes I think that it is better this way but still I am torn between hurting the kids and staying in a hopeless marriage, there is really no solution. They are still so young and I know that this will hurt them a lot.

D. Comparison of speech and musical spectra

Spectral comparisons were based on fundamental frequency and frequency ratios; these two acoustic features were chosen because of the critical roles they play in the perception of both voiced speech sounds and musical intervals. In speech, fundamental frequency carries information about the sex, age, and emotional state of a speaker (Hollien, 1960; Crystal, 1990; Protopapas and Lieberman, 1996; Banse and Scherer, 1996; Harrington *et al.*, 2007); frequency ratios between the first and second formants (F1, F2) differentiate particular vowel sounds, allowing them to understood across speakers with anatomically different vocal tracts (Delattre, 1952; Pickett *et al.*, 1957; Petersen and Barney, 1962; Crystal, 1990; Hillenbrand *et al.*, 1995). In music, the fundamental frequencies of the notes carry the melody; the frequency ratios between notes in the melody and the tonic define the intervals and provide the context that determines whether the composition is in a major or minor mode.

Comparison of fundamental frequencies. The fundamental frequency of each voiced speech segment was determined as described [Fig. 2(a)]. The comparable fundamental frequency of a musical interval, however, depends on the relationship between two notes. The harmonics of two notes can be thought of as the elements of a single harmonic series with a fundamental defined by their greatest common divisor [Fig. 2(b)]. Accordingly, for each of the intervals that distinguished major and minor music in our databases (see Fig. 1 and below), we calculated the frequency of the greatest common divisor of the relevant notes. For tonic intervals these notes were the melody note and its annotated tonic, and for

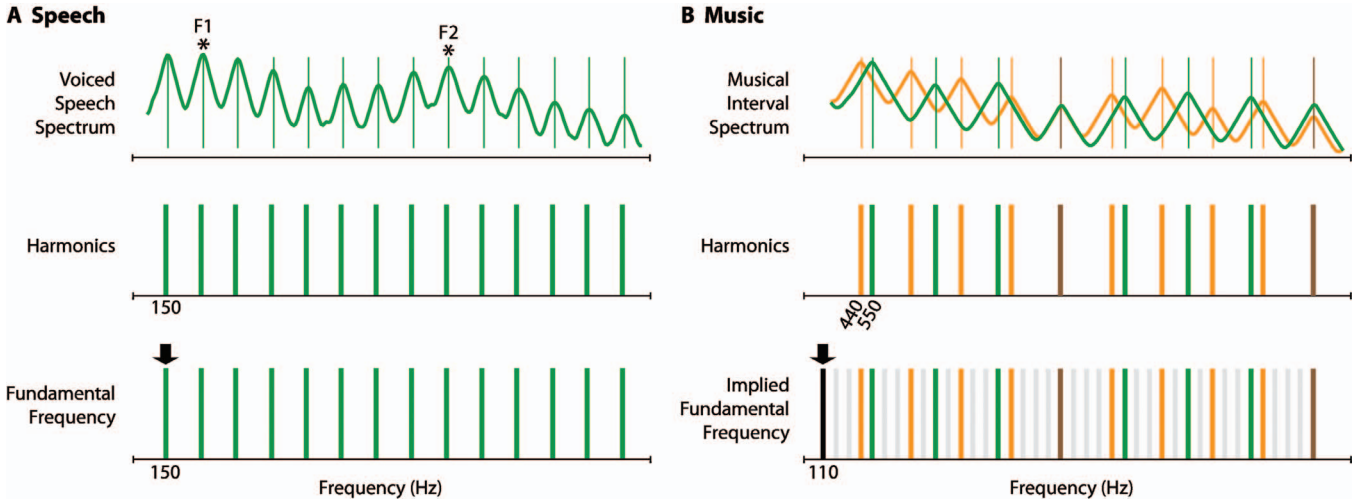


FIG. 2. The harmonic structure of speech sounds and musical intervals. (A) The spectrum of a voiced speech sound comprises a single harmonic series generated by the vibration of the vocal folds (the vertical green lines indicate the loci of harmonic peaks); the relative amplitude of the harmonics is modulated by the resonance of the rest of the vocal tract, thus defining speech formants (asterisks indicate the harmonic peaks of the first two formants, F1 and F2). The voiced speech segment shown here as an example was taken from the single word database and has a fundamental frequency of 150 Hz (black arrow in the lower panel). (B) The spectra of musical intervals entail two harmonic series, one from each of the two relevant notes (see text). The example in the upper panel shows the superimposed spectra of two musical notes related by a major third (the harmonic series of the lower note is shown in orange, and the harmonic series of the higher note in green, and the harmonics common to both series in brown). Each trace was generated by averaging the spectra of 100 recordings of tones played on an acoustic guitar with fundamentals of ~440 and ~550 Hz, respectively; the recordings were made under the same conditions as speech (see Sec. II). The implied fundamental frequency (black arrow in the lower panel) is the greatest common divisor of the two harmonic series.

melodic intervals these notes were the two adjacent notes in the melody. Differences between the distributions of fundamental frequencies in excited and subdued speech, and between the distributions of implied fundamental frequencies in major and minor music, were evaluated for statistical significance by independent two-sample t-tests.

The perceptual relevance of implied fundamental frequencies at the greatest common divisor of two tones is well documented (virtual pitch; reviewed in Terhardt, 1974). However, in order for an implied fundamental to be heard the two tones must be presented simultaneously or in close succession (Hall and Peters, 1981; Grose *et al.*, 2002). These criteria are not met by the tonic intervals we examined here because for many melody notes the tonic will not have been sounded for some time in the melody line. Instead the justification for using implied fundamentals with tonic intervals depends on the tonic's role in providing the tonal context for appreciating the other melody notes (Randel, 1986; Aldwell and Schacter, 2003). Each note in a melody is perceived in the context of its tonic regardless of whether or not the tonic is physically simultaneous. This is evident from at least two facts: (1) if the notes in a melody were not perceived in relation to the tonic, there would be no basis for determining the interval relationships that distinguish one mode from another, making it impossible to distinguish major compositions from minor ones (Krumhansl, 1990; Aldwell and Schachter, 2003); (2) because a note played in isolation has no affective impact whereas a note played in the context of a melody does, the context is what gives individual notes their emotional meaning (Huron, 2006). Thus, the fact that we can hear the difference between major and minor melodies and that we are affected by each in a characteristic way indicates that each note in a melody is, in a very real sense, heard in the context of a tonic.

Comparison of speech formants and musical ratios. The frequency ratios of the first two formants in excited and subdued speech were compared with the frequency ratios of the intervals that specifically distinguish major and minor music. To index F1 and F2 we used the harmonic nearest the peak formant value given by the linear predictive coding (LPC) algorithm utilized by PRAAT'S "to formant" function, i.e., the harmonic with the greatest local amplitude (in a normal voiced speech sound there may or may not be a harmonic power maximum at the LPC peak itself; Fant 1960; Press *et al.*, 1992). The frequency values of the harmonics closest to the LPC peaks were determined by multiplying the fundamental frequency by sequentially increasing integers until the difference between the result and the LPC peak was minimized. The ratios of the first two formants were then calculated as F2/F1 and counted as chromatic if they were within 1% of the just intonation ratios in Table IA. Differences in the prevalence of specific formant ratios in excited and subdued speech were evaluated for statistical significance using chi-squared tests for independence. The analysis focused on F1 and F2 because they are the most powerful resonances of the vocal tract and because they are necessary and sufficient for the discrimination of vowel sounds (Delattre, 1952; Pickett *et al.*, 1957; Rosner and Pickering, 1994). Other formants (e.g., F3 and F4) are also important in speech perception, but are typically lower in amplitude and not critical for the discrimination of vowels (op cit.).

III. RESULTS

A. Intervals in major and minor music

The occurrence of different chromatic intervals among the tonic and melodic intervals in major and minor classical and folk music is shown in Table IV. As expected from the

TABLE IV. Frequency of occurrence of chromatic intervals in major and minor Western classical and Finnish folk music. (A) Tonic intervals; defined as the number of semitones between a melody note and its tonic. (B) Melodic intervals; defined as the number of semitones between adjacent melody notes. The preponderance of small intervals in (B) is in agreement with previous studies (Vos and Troost, 1989). The intervals that distinguish major and minor music are underlined (dashed-lines indicate intervals with less marked contributions).

Major melodies			Minor melodies		
Intervals	Classical (%)	Folk (%)	Intervals	Classical (%)	Folk (%)
(A) Tonic intervals					
Unison/octave	19.9	20.4	Unison/octave	19.4	19.1
Minor second	0.4	0.1	Minor second	0.6	0.2
Major second	12.8	15.9	Major second	13.0	19.6
Minor third	0.8	0.0	Minor third	15.8	15.6
Major third	18.2	16.8	Major third	0.7	0.2
Perfect fourth	10.6	9.5	Perfect fourth	10.5	10.1
Tritone	1.1	0.4	Tritone	1.7	0.2
Perfect fifth	19.1	19.6	Perfect fifth	20.3	19.9
Minor sixth	0.4	0.0	Minor sixth	7.9	1.6
Major sixth	8.4	8.9	Major sixth	1.3	2.9
Minor seventh	0.6	0.2	Minor seventh	3.4	7.5
Major seventh	7.7	8.1	Major seventh	5.4	3.1
(B) Melodic intervals					
Unison	10.7	24.9	Unison	11.4	24.0
Minor second	20.8	13.3	Minor second	28.2	19.1
Major second	36.1	29.9	Major second	27.6	27.6
Minor third	9.4	11.6	Minor third	10.4	12.1
Major third	6.9	8.6	Major third	5.4	6.2
Perfect fourth	7.6	7.4	Perfect fourth	7.2	7.1
Tritone	0.4	0.2	Tritone	1.1	0.0
Perfect fifth	2.7	2.0	Perfect fifth	3.2	2.5
Minor sixth	1.1	0.6	Minor sixth	1.8	0.8
Major sixth	1.3	0.9	Major sixth	1.2	0.1
Minor seventh	0.4	0.3	Minor seventh	0.4	0.1
Major seventh	0.1	0.0	Major seventh	0.2	0.0
Octave	1.3	0.3	Octave	1.2	0.2
Larger	1.0	0.0	Larger	0.9	0.0

formal structure of major and minor scales (see Fig. 1) as well as musical practice, the most salient empirical distinction between major and minor music is the tonic interval of the third scale degree. In both the classical and folk databases major thirds made up 16%–18% of the intervals in major melodies and less than 1% of the intervals in minor melodies; this pattern was reversed for minor thirds, which comprised less than 1% of the intervals in major melodies and about 15% of the intervals in minor melodies. Tonic intervals of the sixth and seventh scale degrees also distinguish major and minor music, but less obviously. These intervals are only about half as prevalent in music as thirds, and their distribution in major versus minor music is less differentiated. There were no marked differences between major and minor music among the tonic intervals held in common by major and minor scales (unison/octave, perfect fifth, major second, perfect fourth; see Fig. 1 and Table IVA).

Notable differences between major and minor music were also found for melodic seconds. In the classical database, there were ~8% more melodic major seconds in major melodies than minor melodies, and ~7% more melodic minor seconds in minor melodies than major melodies. In the folk database this same pattern was apparent, with major melodies containing ~2% more melodic major seconds than minor melodies, and minor melodies containing ~6% more melodic minor seconds than major melodies (Table IVB). The prevalence of the other melodic intervals was similar across major and minor music, differing by 1% or less for all except major thirds in folk music, which were 2.4% more prevalent in major melodies than minor melodies.

The average duration of notes in major and minor melodies was comparable, being within 30 ms in the classical database and within 5 ms in the folk database. The overall difference in the mean pitch height of major and minor melo-

dies in the classical and folk databases was also small, with minor melodies being played an average of 10 Hz higher than major melodies.

This empirical analysis of a large number of melodies in both classical and folk genres thus documents the consensus that the primary tonal distinction between major and minor music with respect to tonic intervals is the prevalence of major versus minor thirds, with smaller differences in the prevalence of major and minor sixths and sevenths (see Sec. IV). With respect to melodic intervals, the only salient distinction between major and minor melodies is the prevalence of major versus minor seconds, major music being characterized by an increased prevalence of major seconds and minor music by an increased prevalence of minor seconds.

B. Comparison of fundamental frequencies

Figure 3 shows the distributions of fundamental frequencies for individual speakers uttering speech in an excited compared to a subdued manner. In agreement with previous observations (Banse and Scherer 1996; Juslin and Laukka, 2003; Scherer 2003; Hammerschmidt and Jurgens, 2007), the fundamental frequency of excited speech is higher than that of subdued speech. On average, the mean fundamental frequency of excited speech exceeded that of subdued speech by 150 Hz for females and 139 Hz for males in the single word condition, and by 58 Hz for females and 39 Hz for males in the monologue condition.

Figure 4 shows the distributions of the implied fundamental frequencies for tonic thirds, sixths, and sevenths and melodic seconds in major and minor melodies. As expected from the ratios that define major and minor thirds (5:4 and 6:5, respectively; see Sec. IV) and their different prevalence in major and minor music (see Table IVA), the implied fundamentals of tonic thirds are significantly higher in major music than in minor music for both genres examined. In classical music, the mean implied fundamental of thirds in major melodies is higher than that in minor melodies by 21 Hz; in folk music, the mean implied fundamental of thirds in major melodies is higher than that in minor melodies by 15 Hz. The pattern for tonic sixths is similar, with the mean implied fundamental of sixths in major melodies being higher than that in minor melodies by 46 Hz in classical music, and by 13 Hz in folk music, despite the occurrence of more major than minor sixths in minor folk music. The pattern for sevenths in major and minor music, however, is different than that of thirds and sixths. Whereas in folk music the mean implied fundamental of sevenths in major melodies is a little higher than that in minor music (by ~2 Hz), in classical music the mean implied fundamental of sevenths in major melodies is actually lower than that in minor melodies (by ~5 Hz; see Sec. IV). The differences between the implied fundamental frequencies of melodic seconds in major versus minor melodies follow the same pattern as tonic thirds and sixths with a higher mean implied fundamental in major music than in minor music in both musical databases. However, as with tonic sevenths the differences between the mean implied fundamentals of seconds in major and minor melodies were relatively small (~3 and ~2 Hz for classical and folk music, respectively; all of these differences are highly significant with p-values <0.0001 or less).

An additional consideration in the comparison of fundamental frequencies in speech and the implied fundamental frequencies of musical intervals is whether they are within the same range. In the speech database, ~97% of fundamental frequencies were between 75 and 300 Hz. In the music database, the percentage of implied fundamentals within this range was different for each of the intervals considered. For tonic thirds and sixths, ~70% and ~56% of implied fundamentals were in this range, respectively. However, only ~1% of the implied fundamentals of tonic sevenths and ~7% of those of melodic seconds were within this range (see Fig. 3). Thus the implied fundamental frequencies of the spectra of tonic thirds and sixths, but not tonic sevenths and melodic seconds, conform to the frequency range of the fundamental frequencies in speech.

These results show that the implied fundamentals of tonic thirds and sixths but not other intervals that empirically distinguish major and minor music parallel to the differences in the fundamental frequencies of excited and subdued speech.

C. Comparison of formant and musical ratios

The distributions of F2/F1 ratios in excited and subdued speech spectra are shown in Fig. 5. In both single words and monologs, excited speech is characterized by a greater number of major interval ratios and relatively few minor interval ratios, whereas subdued speech is characterized by relatively fewer major interval ratios and more minor interval ratios. Thus in the single word database, F2/F1 ratios corresponding to major seconds, thirds, sixths, and sevenths made up ~36% of the ratios in excited speech, whereas ratios corresponding to minor seconds, thirds, sixths, and sevenths were entirely absent. In subdued speech only ~20% of the formant ratios corresponded to major seconds, thirds, sixths, and sevenths, whereas ~10% of the ratios corresponded to minor seconds, thirds, sixths, and sevenths. The same trend is evident in the monolog data; however, because the overlap of the distributions of excited and subdued fundamentals is greater, the differences are less pronounced (see Fig. 3 and Sec. IV).

These parallel differences between the occurrence of formant ratios in excited and subdued speech and the ratios of the musical intervals that distinguish major and minor melodies provide a further basis for associating the spectra of speech in different emotional states with the spectra of intervals that distinguish major and minor music.

IV. DISCUSSION

Although the tonal relationships in music are only one determinant of its affective impact—the emotional influence of tonality in melodies competes with the effects of intensity, tempo, and rhythm among other factors—they are clearly consequential, as indicated by the distinct affective qualities of major and minor music. Despite the use of major and minor tonalities for affective purposes in Western music for at least the last 400 years (Zarlino, 1558), the reason for their

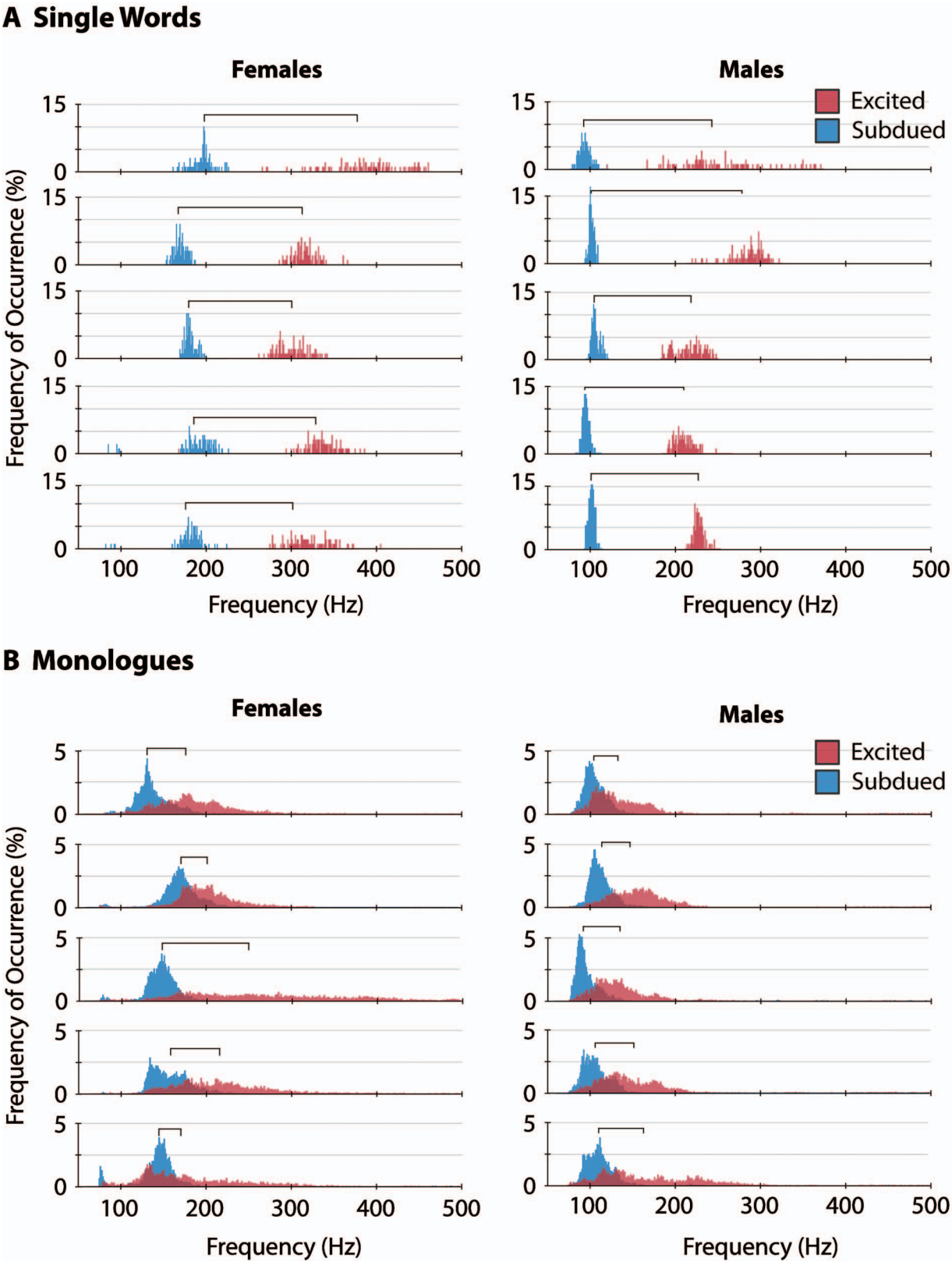


FIG. 3. The fundamental frequencies of excited (red) and subdued (blue) speech segments for individual male and female speakers derived from (A) single word utterances and (B) monolog recordings; brackets indicate means. The differences between the two distributions are significant for each speaker ($p < 0.0001$ or less in independent two-sample t-tests). The difference between the mean fundamentals of the excited and subdued distributions is also significant across speakers ($p < 0.0001$ for single words, and < 0.01 for monologs in dependent t-tests for paired samples).

different emotional quality is not known. We thus asked whether the affective character of major versus minor music might be a consequence of associations made between spectral qualities of the intervals that specifically distinguish major and minor compositions and spectral qualities of speech uttered in different emotional states.

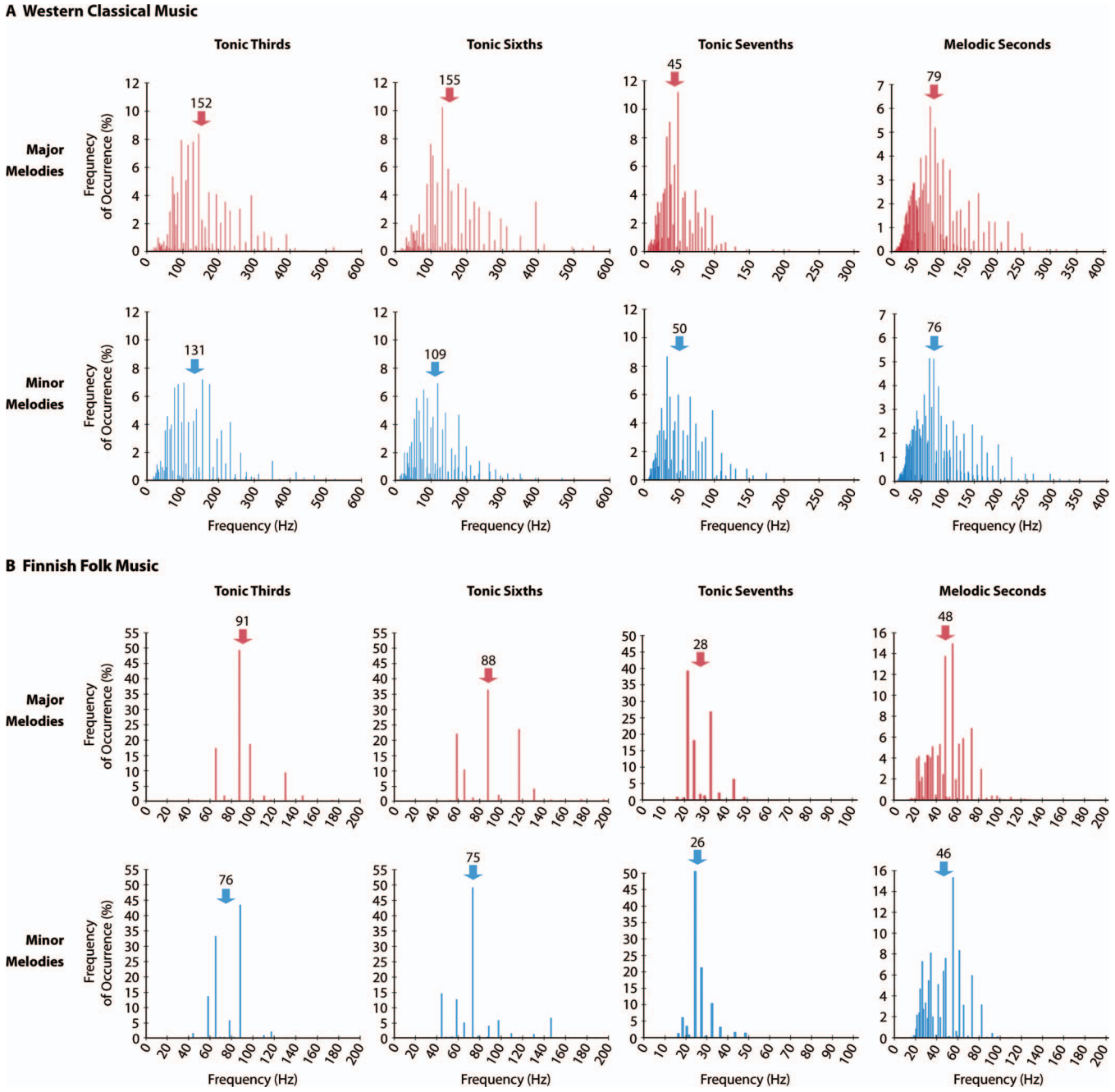
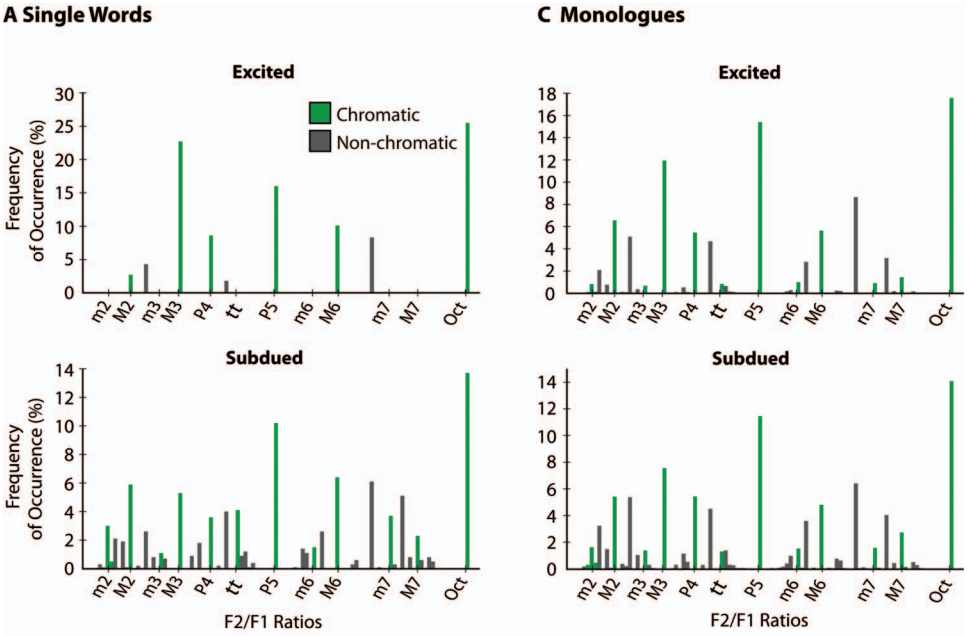


FIG. 4. The implied fundamental frequencies of tonic thirds, sixths, and sevenths and melodic seconds in major and minor melodies from Western classical music (A) and Finnish folk music (B). Arrows indicate the mean implied fundamental frequency values for each distribution. The sparseness of the tonic interval data in (B) is a consequence of fewer different key signatures in our folk music sample compared to our sample of classical music (see Table II), as well as the fact that the classical melodies span more octaves than the folk melodies (~6 octaves vs ~3 octaves). Differences between the distributions of implied fundamental frequencies for major and minor melodies are statistically significant for each of the intervals compared ($p < 0.0075$ or less in independent two-sample t-tests).

A. Empirical differences between major and minor music

Analysis of a large number of classical and folk melodies shows that the principal empirical distinction between major and minor music is the frequency of occurrence of tonic major and minor thirds: nearly all tonic thirds in major melodies are major, whereas this pattern is reversed in minor melodies (see Table IVA). The different distributions of tonic major and minor sixths and sevenths also contribute to the tonal distinction of major and minor music, but less prominently. Although major melodies almost never contain tonic minor sixths and sevenths, melodies in a minor key often include tonic major sixths and sevenths, presumably because the harmonic and melodic versions of the minor scales are often used (see Fig. 1). An additional distinction between major and minor music is the prevalence of melodic major and minor seconds. Compared to minor melodies, major melodies contain more melodic major seconds, and, compared to major melodies, minor melodies contain more melodic minor seconds (see Table IVB).



B Single Words

Intervals	Excited	Subdued	p-value
Unison/Octave	25.5 %	13.7 %	< .0001
Minor Second	0.0 %	3.5 %	< .0001
Major Second	2.7 %	5.8 %	< .001
Minor Third	0.0 %	1.1 %	< .001
Major Third	22.7 %	5.3 %	< .0001
Perfect Fourth	8.6 %	3.6 %	< .0001
Tritone	0.0 %	4.1 %	< .0001
Perfect Fifth	16.0 %	10.2 %	< .0005
Minor Sixth	0.0 %	1.5 %	< .0001
Major Sixth	10.1 %	6.4 %	< .01
Minor Seventh	0.0 %	3.7 %	< .0001
Major Seventh	0.0 %	2.3 %	< .0001
Non-Chromatic	14.4 %	38.7 %	< .0001

D Monologues

Intervals	Excited	Sudbued	p-value
Unison/Octave	17.6 %	14.1 %	< .0001
Minor Second	1.0%	2.1 %	< .0001
Major Second	6.6 %	5.4 %	< .0001
Minor Third	0.7 %	1.4 %	< .0001
Major Third	12.0 %	7.6 %	< .0001
Perfect Fourth	5.5 %	5.4 %	= .82
Tritone	0.8 %	1.3 %	<. 0001
Perfect Fifth	15.4 %	11.5 %	< .0001
Minor Sixth	1.0 %	1.5 %	< .0001
Major Sixth	5.6 %	4.8 %	< .0001
Minor Seventh	0.9 %	1.6 %	< .0001
Major Seventh	1.4 %	2.7 %	< .0001
Non-Chromatic	31.6 %	40.7 %	< .0001

FIG. 5. Comparison of the ratios of the first two formants in excited and subdued speech derived from analyses of the single word and monologue databases. Ratios have been collapsed into a single octave such that they range from 1 to 2. (A) The distribution of formant ratios in excited and subdued speech from the single word database; green bars indicate ratios within 1% of chromatic interval ratios (see Table IA); gray bars indicate ratios that did not meet this criterion. (B) The percentage of formant ratios corresponding to each chromatic interval in (A) for excited and subdued speech. (C) The same as (A), but for the monolog data. (D) The same as (B), but for the monolog data in (C). *p*-values for each interval were calculated using the chi-squared test for independence, with expected values equal to the mean number of occurrences of an interval ratio across excited and subdued speech. Intervals empirically determined to distinguish major and minor music are underlined (dashed-lines indicate intervals with less marked contributions).

B. Comparison of speech and music spectra

Speech uttered in an excited compared to a subdued manner differs in pitch, intensity, tempo, and a number of other ways (Bense and Scherer 1996; Scherer 2003; Hammerschmidt and Jurgens, 2007). With respect to tonality, however, the comparison of interest lies in the spectral features of voiced speech and musical intervals (see Introduction).

The results we describe show that differences in both fundamental frequency and formant relationships in excited versus subdued speech spectra parallel spectral features that distinguish major and minor music. When a speaker is excited, the generally increased tension of the vocal folds raises the fundamental frequency of vocalization; conversely when speakers are in a subdued state decreased muscular tension lowers the fundamental frequency of vocalization (see Fig. 3). In music two factors determine the frequency of a musical interval's implied fundamental: (1) the ratio that defines the interval and (2) the pitch height at which the interval is played. The defining ratios of minor seconds, thirds, and sixths (16:15, 6:5 and 8:5, respectively) yield smaller greatest common devisors than the defining ratios of major seconds, thirds, and sixths (9:8, 5:4, and 5:3); thus minor seconds, thirds, and sixths played at the same pitch height as their major counterparts will always have lower implied fundamentals. In our musical databases, the average pitch height of minor melodies was somewhat higher than major melodies, thus reducing the difference between the implied fundamentals of these major and minor intervals (see Sec. III). Despite these effects, the mean implied fundamentals of tonic thirds and sixths and melodic seconds in major music are still higher than those in minor music. However, for melodic seconds, the relatively small size of this difference (~2–3 Hz) and the fact that most of their implied fundamentals (~93%) fall below the range of fundamental frequencies in speech make associations with the spectra of excited and subdued speech on these grounds less likely. In contrast to the defining ratios of seconds, thirds, and sixths, the defining ratio of the minor seventh (9:5) yields a *larger* greatest common devisor than the defining ratio of the major seventh (16:9), making its implied fundamental at a given pitch height higher than its major counterpart. This reasoning indicates why the pattern observed for tonic thirds and sixths and melodic seconds is not also apparent for tonic sevenths. Furthermore, as with melodic seconds, nearly all of the implied fundamentals of tonic sevenths (~99%) fall below the range of the fundamentals in speech.

The difference between the fundamental frequencies of

excited and subdued speech also affects the prevalence of specific formant ratios. Given the same voiced speech sound, the positions of the first and second formants are relatively stable between excited and subdued speech (as they must be to allow vowel phonemes to be heard correctly). The higher fundamental frequencies of excited speech, however, increase the frequency distance between harmonics, causing lower harmonics to underlie the first and second formants. As a result the F2/F1 ratios in excited speech tend to comprise smaller numbers and thus more often represent musical intervals defined by smaller number ratios. Conversely, the lower fundamental frequencies in subdued speech decrease the distance between harmonics, causing higher harmonics to underlie the formants. Thus the F2/F1 ratios in subdued speech tend to comprise larger numbers, which more often represent musical intervals defined by larger number ratios. Intervals whose defining ratios contain only the numbers one through five (octaves, perfect fifths, perfect fourths, major thirds, and major sixths) were typically more prevalent in the F2/F1 ratios of excited speech, whereas intervals with defining ratios containing larger numbers (all other chromatic intervals) were more prevalent in the F2/F1 ratios of subdued speech (see Table I and Fig. 5). The only exceptions to this rule were the prevalences of major seconds and perfect fourths in the F2/F1 ratios of the monolog recordings, which were slightly higher in excited speech than in subdued speech or not significantly different, respectively. Presumably, these exceptions result from the greater overlap of fundamental frequencies between excited and subdued speech in the monologues (see Fig. 3).

Thus, based on these comparisons, differences in the spectra of excited and subdued speech parallel differences in the spectra of major and minor tonic thirds and sixths, but not of major and minor melodic seconds and tonic sevenths.

C. Melodic seconds and tonic sevenths

Although the empirical prevalence of melodic seconds and tonic sevenths also distinguishes major and minor music (see Table IV), as described in Sec. IV B the spectral characteristics of these intervals do not parallel those in excited and subdued speech. The mean implied fundamental frequency of tonic sevenths in major music is lower than in minor music, and, although the mean implied fundamental frequency of melodic seconds in major music is higher than in minor music the difference is small. Moreover, the majority of implied fundamentals for these intervals are below the range of fundamental frequencies in speech. With respect to formant ratios, there are fewer minor seconds and sevenths in excited speech than in subdued speech, but this was also true of major seconds and sevenths.

These results accord with music theory. Unlike tonic thirds and sixths, melodic seconds and tonic sevenths are not taken to play a significant role in distinguishing major and minor music (Aldwell and Schachter, 2003). Rather these intervals are generally described as serving other purposes, such as varying the intensity of melodic motion in the case of seconds and creating a sense of tension that calls for resolution to the tonic in the case of sevenths (op cit.). This interpretation is supported by the fact that major and minor melodic seconds and tonic sevenths are commonplace in both major and minor music (see Table IV).

D. Just intonation vs equal temperament

Most of the music analyzed here will have been performed and heard in equal tempered tuning, as is almost all popular music today; given that our hypothesis is based on associations between speech and music, our decision to use just intoned rather than equally tempered ratios to define the musical intervals requires justification.

Just intonation is based on ratios of small integers and thus is readily apparent in the early harmonics of any harmonic series, voiced speech included (Rossing, 1990). To relate the spectral structure of voiced speech to the spectra of musical intervals, it follows that just intoned ratios are the appropriate comparison. Equal tempered tuning is a compromise that allows musicians to modulate between different key signatures without retuning their instruments while retaining as many of the perceptual qualities of just intoned intervals as possible. The fact that the differences introduced by equal tempered tuning are acceptable to most listeners implies that either system is capable of associating the harmonic characteristics of music and speech.

Furthermore, the spectral comparisons we made do not strictly depend on the use of just intoned ratios. With respect to implied fundamental frequencies, the phenomenon of virtual pitch is robust and does not depend on precise frequency relations (virtual pitches can be demonstrated with equally tempered instruments such as the piano; Terhardt *et al.*, 1982). With respect to formant ratios, if equal tempered tunings had been used to define the intervals instead of just intonation tunings, the same results would have been obtained by increasing the window for a match from 1% to 2%. Finally, retuning the major and minor thirds and sixths for the 7497 melodies we examined, resulted in a mean absolute frequency difference of only 3.5 Hz.

E. Why thirds are preeminent in distinguishing major and minor music

A final question is why the musical and emotional distinction between major and minor melodies depends primarily on tonic thirds, and how this fact aligns with the hypothesis that associations made between the spectral characteristics of music and speech are the basis for the affective impact of major versus minor music. Although our analysis demonstrates similarities between the spectra of major and minor thirds and the spectra of excited and subdued speech, it does not indicate why thirds, in particular, provide the critical affective difference arising from the tonality of major and minor music. One plausible suggestion is that among the intervals that differentiate major and minor tone collections, thirds entail the lowest and thus the most powerful harmonics. Accordingly, of the distinguishing intervals, thirds are likely to be the most salient in the spectra of both voiced speech sounds and musical tones.

V. CONCLUSION

In most aspects of music—e.g., intensity, tempo, and rhythm—the emotional quality of a melody is conveyed at least in part by physical imitation in music of the characteristics of the way a given emotion is expressed in human behavior. Here we asked whether the same principle might apply to the affective impact of the different tone collections used in musical compositions. The comparisons of speech and music we report show that the spectral characteristics of excited speech more closely reflect the spectral characteristics of intervals in major music, whereas the spectral characteristics of subdued speech more closely reflect the spectral characteristics of intervals that distinguish minor music. Routine associations made between the spectra of speech uttered in different emotional states and the spectra of thirds and sixths in major and minor music thus provide a plausible basis for the different emotional effects of these different tone collections in music.

ACKNOWLEDGMENTS

We are grateful to Sheena Baratono, Nigel Barella, Brent Chancellor, Henry Greenside, Yizhang He, Dewey Lawson, Kevin LaBar, Rich Mooney, Deborah Ross, David Schwartz, and Jim Voyvodic for useful comments and suggestions. We especially wish to thank Dewey Lawson for his help measuring the acoustical properties of the recording chamber, to Deborah Ross help selecting and recording the monologs, to Shin Chang, Lap-Ching Keung, and Stephan Lotfi who helped organize and validate the musical databases, and to Tuomas Eerola for facilitating our access to the Finnish Folk Song Database.

Aldwell, E., and Schachter, C. (**2003**). *Harmony & Voice Leading*, 3rd ed. (Wadsworth Group/Thomson Learning, Belmont, CA).

Banse, R., and Scherer, K. R. (**1996**). "Acoustic profiles in vocal emotion expression," J. Pers. Soc. Psychol. **70**, 614–636.

Barlow, H., and Morgenstern, S. (**1974**). *A Dictionary of Musical Themes* (Crown, New York).

Bernstein, L. (**1976**). *The Unanswered Question: Six Talks at Harvard* (Harvard University Press, Cambridge, MA).

Boersma, P. (**1993**). "Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound," Proceedings of the Institute of Phonetic Sciences, **17**, pp. 97–110, University of Amsterdam.

Boersma, P., and Weenik, D. (**2008**). PRAAT: Doing phonetics by computer (Version 5.0.43), available from http://www.fon.hum.uva.nl/praat/ (Last viewed November, 2008).

Burkholder, J. P., Grout, D., and Palisca, C. (**2005**). *A History of Western Music*, 7th ed. (Norton, New York).

Burns, E. M. (**1999**). "Intervals, scales and tuning," in *The Psychology of Music*, 2nd ed., edited by D. Deutsch (Academic, New York).

Carterette, E. C., and Kendall, R. A. (**1999**). "Comparative music perception and cognition," in *The Psychology of Music*, 2nd ed., edited by D. Deutsch (Academic, New York).

Cohen, D. (**1971**). "Palestrina counterpoint: A musical expression of unexcited speech," Journal of Music Theory **15**, 85–111.

Cooke, D. (**1959**). *The Language of Music* (Oxford University Press, Oxford, UK).

Crowder, R. G. (**1984**). "Perception of the major/minor distinction: Hedonic, musical, and affective discriminations," Bull. Psychon. Soc. **23**, 314–316.

Crystal, D. (**1997**). *The Cambridge Encyclopedia of Language*, 2nd ed. (Cambridge University Press, New York).

Delattre, P., Liberman, A. M., Cooper, F. S., and Gerstman, L. J. (**1952**). "An experimental study of the acoustic determinants of vowel color: Observation of one- and two-formant vowels synthesized from spectrographic patterns," Word **8**, 195–210.

Eerola, T., and Tovianien, P. (**2004a**). Suomen Kasan eSävelmät (Finnish Folk Song Database), available from http://www.jyu.fi/musica/sks/ (Last viewed November, 2008).

Eerola, T., and Toiviainen, P. (**2004b**). MIDI Toolbox: MATLAB Tools for Music Research (Version 1.0.1), available from http://www.jyu.fi/hum/laitokset/musiikki/en/research/coe/materials/miditoolbox/ (Last viewed November, 2008).

Fant, G. (**1960**). *Acoustic Theory of Speech Production* (Mouton, The Hague).

Gabrielsson, A., and Lindström, E. (**2001**). "The influence of musical structure on emotional expression," in *Music and Emotion: Theory and Research*, 1st ed., edited by P. N. Juslin and J. A. Sloboda (Oxford University Press, New York).

Gregory, A. H., and Varney, N. (**1996**). "Cross-cultural Comparisons in the affective response to music," Psychol. Music **24**, 47–52.

Grose, J. H., Hall, J. W., III, and Buss, E. (**2002**). "Virtual pitch integration for asynchronous harmonics," J. Acoust. Soc. Am. **112**, 2956–2961.

Hall, J. W., andPeters, R. W. (**1981**). "Pitch for non-simultaneous harmonics in quiet and noise," J. Acoust. Soc. Am. **69**, 509–513.

Hammerschmidt, K., and Jurgens, U. (**2007**). "Acoustical correlates of affective prosody," J. Voice **21**, 531–540.

Harrington, J., Palethorpe, S., and Watson, C. I. (**2007**). "Age-related changes in fundamental frequency and formants: a longitudinal study of four speakers," in Proceedings of Interspeech, 2007, Antwerp.

Heinlein, C. P. (**1928**). "The affective characters of the major and minor modes in music," J. Comp. Psychol. **8**, 101–142.

Helmholtz, H. (**1885**). *Lehre von den tonempfindungen (On the Sensations of Tone)*, 4th ed., translated by A. J. Ellis (Dover, New York).

Hevner, K. (**1935**). "The affective character of the major and minor modes in music," Am. J. Psychol. **47**, 103–118.

Hillenbrand, J. M., and Clark, M. J. (**2001**). "Effects of consonant environment on vowel formant patterns," J. Acoust. Soc. Am. **109**, 748–763.

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

Hollien, H. (**1960**). "Some laryngeal correlates of vocal pitch," J. Speech Hear. Res. **3**, 52–58.

Huron, D. (**2006**). *Sweet Anticipation: Music and the Psychology of Expectation* (MIT, Cambridge, MA).

Johnston, I. (**2002**). *Measured Tones: The Interplay of Physics and Music* (Taylor & Francis, New York).

Johnstone, T., and Scherer, K. R. (**2000**). "Vocal communication of emotion," in *Handbook of Emotions*, 2nd ed., edited by M. Lewis and J. M. Haviland-Jones (Guilford, New York).

Juslin, P. N., and Laukka, P. (**2003**). "Communication of emotions in vocal expression and music performance: Different channels, same code?," Psychol. Bull. **129**, 770–814.

Krumhansl, C. L. (**1990**). *Cognitive Foundations for Musical Pitch* (Oxford University Press, New York).

MathWorks Inc. (**2007**). MATLAB (Version R2007a) (The MathWorks Inc., Natick, MA).

Nettl, B. (**1956**). *Music in Primitive Culture* (Harvard University Press, Cambridge, MA).

Peretz, I., Gagnon, L., and Bouchard, B. (**1998**). "Music and emotion: Perceptual determinants, immediacy, and isolation after brain damage," Cognition **68**, 111–141.

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

Pickett, J. M. (**1957**). "Perception of vowels heard in noises of various spectra," J. Acoust. Soc. Am. **29**, 613–620.

Pierce, J. R. (**1962**). *The Science of Musical Sound*, revised ed. (Freeman, New York).

Press, W. H., Teukolsky, S. A., Vetterling, W. T., and Flannery, B. P. (**1992**). *Numerical Recipes in C: The Art of Scientific Computing*, 2nd ed. (Cambridge University Press, New York).

Protopapas, A., and Lieberman, P. (**1997**). "Fundamental frequency of phonation and perceived emotional stress," J. Acoust. Soc. Am. **101**, 2267–2277.

Randel, D. M. (**1986**). *The New Harvard Dictionary of Music*, revised 2nd ed. (Belknap, Cambridge, MA).

Rosner, B. S., and Pickering, J. B. (**1994**). *Vowel Perception and Production* (Oxford University Press, New York).
Rossing, T. D. (**1990**). *The Science of Sound*, 2nd ed. (Addison-Wesley, New York).
Schellenberg, E. G., Krysciak, A. M., and Campbell, R. J. (**2000**). "Perceiving emotion in melody: Interactive effects of pitch and rhythm," Music Percept. **18**, 155–171.
Scherer, K. R. (**2003**). "Vocal communication of emotion: A review of research paradigms," Speech Commun. **40**, 227–256.
Scherer, K. R., Banse, R., and Wallbott, H. G. (**2001**). "Emotional inferences from vocal expression correlate across languages and cultures," Journal of Cross-Cultural Psychology **32**, 76–92.
Schwartz, D. A., and Purves, D. (**2004**). "Pitch is determined by naturally occurring periodic sounds," Hear. Res. **194**, 31–46.
Schwartz, D. A., Howe, C. Q., and Purves, D. (**2003**). "The statistical structure of human speech sounds predicts musical universals," J. Neurosci. **23**, 7160–7168.
Spencer, H. (**1868**). *Essays: Scientific, Political, and Speculative: Volume 2* (Williams & Norgate, London).
Stevens, K. N. (**1998**). *Acoustic phonetics.* (MIT, Cambridge, MA).
Terhardt, E. (**1974**). "Pitch, consonance, and harmony," J. Acoust. Soc. Am. **55**, 1061–1069.
Terhardt, E., Stoll, G., and Seewann, M. (**1982**). "Pitch of complex signals according to virtual-pitch theory: tests, examples, and predictions," J. Acoust. Soc. Am. **71**, 671–678.
Thompson, W. F., and Balkwill, L. L. (**2006**). "Decoding speech prosody in five languages," Semiotica **158**, 407–424.
Vos, P. G., and Troost, J. M. (**1989**). "Ascending and descending melodic intervals: Statistical findings and their perceptual relevance," Music Percept. **6**, 383–396.
Zarlino, G. (**1558**). *Le Institutioni hamoniche*, translated by G. Marco and C. Palisca, (Yale University Press, New Haven, CT) Book 3.

An acoustical study of vocal pitch matching in congenital amusia

Sean Hutchins[a)]
International Laboratory for Brain, Music, and Sound Research (BRAMS), 1430 Boulevard Mont-Royal, Montréal, Quebec H2V 4P3, Canada and Department of Psychology, Université de Montréal, 1430 Boulevard Mont-Royal, Montréal, Quebec H2V 4P3, Canada

Jean Mary Zarate and Robert J. Zatorre
International Laboratory for Brain, Music, and Sound Research (BRAMS), 1430 Boulevard Mont-Royal, Montréal, Quebec H2V 4P3, Canada and Cognitive Neuroscience Unit, Montreal Neurological Institute, McGill University, 3801 Rue University, Montréal, Quebec H3A 2B4, Canada

Isabelle Peretz
International Laboratory for Brain, Music, and Sound Research (BRAMS), 1430 Boulevard Mont-Royal, Montréal, Quebec H2V 4P3, Canada and Department of Psychology, Université de Montréal, 1430 Boulevard Mont-Royal, Montréal, Quebec H2V 4P3, Canada

(Received 12 May 2009; revised 5 November 2009; accepted 10 November 2009)

Vocal pitch matching is a foundational skill for singing and is an interesting place to study the relationship between pitch perception and production. To better understand this relationship, we assessed pitch-matching abilities in congenital amusics, who have documented disabilities in pitch perception, and in matched controls under normal, masked, and guided feedback conditions. Their vocal productions were analyzed for fundamental frequency and showed that amusics were significantly less accurate at pitch matching than the controls. However, five of the six amusics showed a significant correlation between their produced pitches and the target pitch. Feedback condition had no effect on pitch-matching accuracy. These results show impaired vocal pitch-matching abilities in amusics but also show a relationship between perceived and produced pitches. © *2010 Acoustical Society of America.* [DOI: 10.1121/1.3270391]

PACS number(s): 43.75.Cd, 43.75.Rs [DD] Pages: 504–512

I. INTRODUCTION

Congenital amusia, commonly referred to as "tone-deafness," is characterized by the inability to consciously detect small pitch changes, recognize familiar tunes, or detect out-of-key notes in melodies (Ayotte *et al.*, 2002; Peretz, 2001; Peretz *et al.*, 2003). Prior studies have also linked amusia to worse-than-normal singing. Compared with normals, who are relatively proficient singers (Dalla Bella *et al.*, 2007; Pfordresher and Brown, 2007), congenital amusics are judged as having impaired singing abilities (Ayotte *et al.*, 2002). Acoustic analyses show that amusics are, on average, substantially worse at singing familiar melodies than normals (Dalla Bella *et al.*, 2009). These differences are amplified when the amusics are asked to sing using a neutral syllable (/la/) in place of lyrics, suggesting that most of the (relatively slight) proficiency the amusics do show in singing production may be due to long-term memory cues from pitches associated with lyrics.

However, this same study (Dalla Bella *et al.*, 2009) also provided evidence that auditory perception and action streams may be distinct in some ways. This study identified two congenital amusics who showed preserved singing abilities, despite their severe pitch perception deficits. In addition, Loui *et al.* (2008) showed that amusics can vocally mimic the direction of intervals played to them. In their study, amusics could produce an interval in the same direction as a perceived interval, although their productions were at different absolute pitches and interval magnitudes; however, they were less accurate at naming the direction of the same interval (by saying "up" or "down"). This dissociation between conscious perceptual and production abilities is further supported by reports that quarter-tone differences in pitch can evoke electrophysiological responses without perceptual awareness in amusics (Peretz *et al.*, 2009). These reports suggest that sensory information about pitch is entering the system in amusics, and that some pitch discrimination abilities may be preserved (perhaps as an auditory analog to the visual perception phenomenon of blindsight, specifically the action-blindsight subtype, as categorized by Danckert and Rossetti (2005). These residual abilities may be too weak to support conscious perception but may be best expressed in production responses. Thus, although amusics are often reported to lack singing proficiency, this may be neither absolute nor universal. It is, therefore, important to more fully understand the pitch production abilities of amusics, both as a group and at the individual level.

Here, we tested amusics' ability to vocally match heard pitches. Pitch matching is a fundamental musical skill, considered to underlie many core musical abilities, and to be the most important factor in judging singing talent (Watts *et al.*,

[a)]Author to whom correspondence should be addressed. Electronic mail: sean.michael.hutchins@umontreal.ca

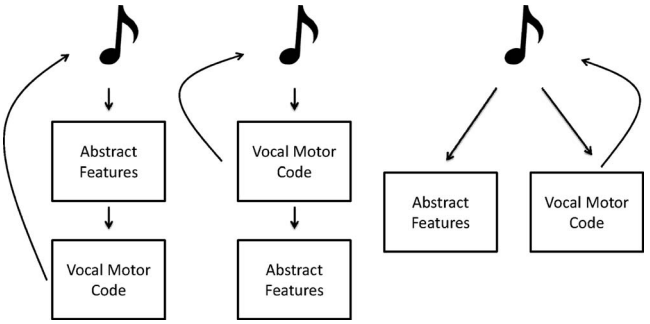


FIG. 1. Three possible models describing the relationship between perceived sounds, abstract feature categorization, the vocal-motor code, and auditory feedback.

2003). Despite the relative importance of this factor in singing (one of the most common means that people engage in music production), pitch matching is relatively unstudied as a phenomenon. The vast majority of the literature concerning pitch matching has been undertaken from a pedagogical standpoint and focuses on young children or adolescents. These studies have suggested several factors that may lead to poor pitch-matching abilities, including poor pitch discrimination (Pederson and Pederson, 1970), vocal control (Jones, 1979), and problems with normal developmental maturation (Goetze *et al.*, 1990). In addition, pitch matching seems to be aided by use of a limited and appropriate range for the singer as well as by the timbral similarity of the model (Goetze *et al.*, 1990). Pfordresher and Brown (2007) found that 87% of non-musicians could vocalize within one semitone of a target pitch, and that the instances of poor-pitch singing could not be attributed to poor perceptual abilities.

Pitch matching is an interesting place to study the relationship between perception and production because it represents the conceptually simplest interface between the two areas in the domain of pitch. Figure 1 shows three models outlining possible relationships between perceived notes in the environment, our brains' ability to abstract conceptual features of the heard note, and the motor codes underlying our ability to produce the pitch vocally. In each of the models, tones are present in the environment and are picked up by our sensory apparatuses. Two types of processes may subsequently ensue, based on the task demands. In the case of many standard perceptual tasks, the tone may need to be categorized, classified, and compared. This requires accessing abstract features of the tone, including a consciously accessible, symbolic representation of its (relative) pitch. A vocal pitch-matching task, however, does not necessarily require access to the abstract features of a tone but does require access to the vocal-motor code specific to that pitch. The leftmost diagram in Fig. 1 shows a traditional explanation of the relationship between these two types of processes. This model presumes that tones are first processed for their abstract features, which are then mapped to vocal-motor codes for production. Pitch matching in this model is mediated by the abstract representations of pitch. The middle diagram outlines a motor theory of pitch production, in which tones are first processed in terms of the motor codes necessary to produce the pitches, and then categorized into their abstract features, including a symbolic representation of pitch (as in Liberman and Mattingly, 1985, for speech), but only if necessary for the task. Finally, the rightmost diagram in Fig. 1 shows a dual-route model, in which the vocal-motor code and abstract feature representation are accomplished by separate subsystems and need not be intrinsically related to each other. Each model can also receive auditory feedback from hearing one's own singing. To be clear, in all of these models, both abstract and motor codes include a representation of pitch. However, the abstract representation yields a categorical representation necessary for decision-making, and the motor code represents pitch solely in terms of motor activity. Furthermore, these models do not dispute that other motor code representations of pitch may be present, for example, in trained instrumentalists; however, the vocal-motor code should be present in all normal individuals.

The rightmost model shown in Fig. 1 bears a resemblance to other dual-route processing models, especially that of Hickok and Poeppel (2004). In their model, speech processing is handled by two separate subsystems, one of which maps speech sounds onto conceptual representations, while the other maps speech sounds onto motor representations. These two subsystems are neurally instantiated in two processing streams, i.e., the ventral and dorsal pathways, respectively. The ventral stream is used for conceptually oriented tasks, and the dorsal is used for production-oriented tasks. Other studies have also shown production and perception task dissociations. Goodale and Milner (1992) and Milner and Goodale (1995) proposed a similar division of labor within the visual system, with the ventral stream responsible for conceptual processing and the dorsal stream responsible for vision-for-action. Danckert and Rossetti (2005) described a dissociation between types of residual abilities found in patients with blindsight. Patients with residual ventral stream pathways show residual abilities in conceptual tasks, including discrimination and decision tasks, whereas patients with residual dorsal stream pathways show better performance in action tasks, such as grasping, pointing, or saccades.

In general audition, dual-pathway models have also been gaining some support. Some neuroscientists have suggested that ventral and dorsal pathways process auditory signals for information concerning "what" and "where," respectively (Rauschecker, 1998; Warren and Griffiths, 2003). However, others have proposed that the ventral and dorsal pathways may be better thought of as conceptual and action pathways, similar to the case of vision (Warren *et al.*, 2005). Hickok and Poeppel's (2004) dual-route model of speech processing also falls under this organization schema of a ventral comprehension pathway and a dorsal production pathway. Indeed, recent findings on music processing also show evidence of a separable action stream (Griffiths, 2008). For example, Hafke (2008) showed vocal adjustments in trained singers to pitch-shifted feedback even when these shifts were not overtly perceivable. The work of Loui *et al.* (2008) also pointed toward separate streams of processing, as the amusic subjects were better able to process pitch information using sung responses than same-different judgments.

In theory, the pattern of errors we see in amusia for pitch-matching tasks can help us rule out particular models.

TABLE I. Amusic and control participant characteristics. Asterisks indicate amusic participants. MBEA global score refers to the average score across the six tests of the Montreal Battery of Evaluation of Amusia (Peretz *et al.*, 2003). Pitch discrimination threshold refers to minimum pitch difference between tones needed for participants to be able to discriminate them 75% of the time (Hyde and Peretz, 2004). Trials completed refers to the number of vocal pitch-matching recordings obtained.

Subject	Age	Gender	Years of education	Years of music education	MBEA global score	Pitch discrimination threshold (in cents)	Trials completed
IC*	65	M	19	1	51.1	100	119
GC*	61	F	20	1	72.7	100	116
FA*	67	F	15	2	69.4	50	89
MB*	65	F	21	4	71.1	50	118
AS*	67	F	14	2	62.7	50	57
JL*	70	M	15	1	73.8	50	120
CL	57	F	16	3	90	25	107
GB	62	F	19	2	90.5	25	120
FB	57	F	13	2	91.6	25	119
RS	63	F	18	2	93.3	25	120

For example, if vocal pitch-matching is mediated by abstract pitch categorization of the perceived target note, performance on the former task should be as bad or worse than on decision tasks only requiring abstract pitch categorization. However, if amusics show better performance on pitch-matching tasks than on decision tasks, this points toward a motor model or dual-route model. Based on Loui *et al.*'s (2008) findings, we expect that amusics should be better at this pitch-matching task than their perceptual deficits would suggest, arguing against a vocal-motor code mediated by abstract categorization of pitch.

Another important way in which the relationship between perception and production has been studied is in terms of feedback. It is a well-known phenomenon that delayed auditory feedback can severely disrupt speech and music production (Finney, 1997; MacKay, 1987). In both speech and singing, shifting the pitch of the feedback a subject receives from her own voice can result in an automatic correction in the other direction (Burnett *et al.*, 1998; Burnett and Larson, 2002; Natke *et al.*, 2003). Altered feedback can also disrupt instrumental (piano) performance, regardless of skill level or experience (Pfordresher, 2005). Additionally, previous studies have shown that masking auditory feedback reduces pitch-matching accuracy (Anstis and Cavanagh, 1979; Elliott and Niemoeller, 1970; Mürbe *et al.*, 2002; Ternström *et al.*, 1988). These studies show that, under normal conditions, people are sensitive to the pitch, timing, and amplitude of the feedback they receive from their own voice or instrument. Pfordresher and Brown (2007) examined singing performance under three difference feedback conditions. They found that augmenting feedback (by playing the correct tones along with the sung responses) could help intervallic precision somewhat but had a small negative effect on pitch-matching abilities of poor singers. Masking the feedback had little effect on vocal performance. Wise and Sloboda (2008), on the other hand, found that accompaniment could aid pitch matching in a singing task. Manipulating the feedback to amusics during a pitch-matching task will help us understand how they use perceptual information to guide their productions. If amusics rely on feedback to guide their production, this would show evidence for a link between perception and production.

We hypothesized that performance would be affected by manipulations of the acoustic feedback. Compared with no feedback manipulations, we hypothesized that amusics and controls should match pitches more accurately when the target was re-presented as the participants began to sing (similar to Wise and Sloboda, 2008). In contrast, eliminating most of their auditory feedback through masking with noise should cause a decrease in pitch-matching accuracy (Anstis and Cavanagh, 1979; Elliott and Niemoeller, 1970; Mürbe *et al.*, 2002; Ternström *et al.*, 1988). This would indicate that amusics are sensitive to their own feedback and could point toward a way to improve amusics' vocal performance. We also expect to find a range of vocal pitch-matching abilities among amusics, who should be less accurate than controls. Because any inaccurate production can reasonably be attributed to perceptual difficulties, their average error should be related to the degree of perceptual deficit.

II. METHODS

A. Participants

Participants included six amusics (four female) and four non-amusic controls (all female). Amusics were determined by the scores on the Montreal Battery of Evaluation of Amusia (MBEA) (Peretz *et al.*, 2003). This battery assesses the components of music processing with a total of six sub-tests concerning melody, rhythm, meter, and memory. Amusics obtained a global score over two standard deviations below controls (amusic mean score=67.6% and control mean score=91.4%) but had normal hearing. Controls were matched for age (amusic mean age=65.6 and control mean age=59.8). Table I shows individual participant information for controls and amusics, including ages, education levels, MBEA global scores (the average score on all subtests of the MBEA, Peretz *et al.*, 2003), and pitch discrimination thresholds (the minimum pitch difference between two tones needed for participants to be able to discriminate them over 75% of the time).

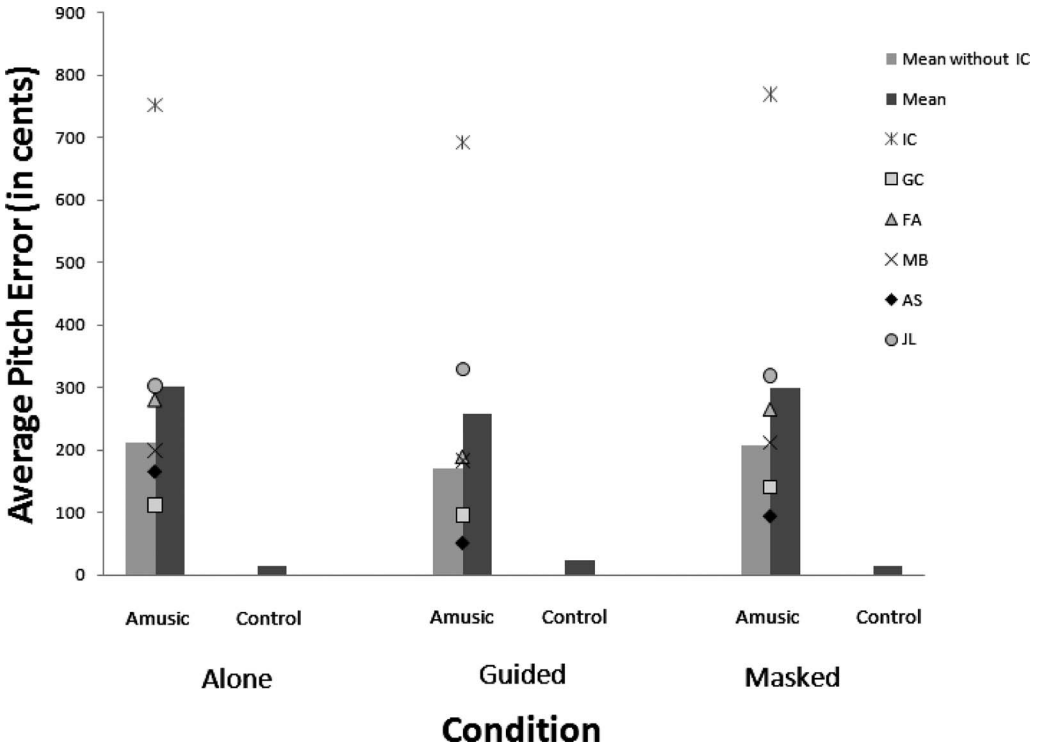


FIG. 2. Average absolute pitch errors of each amusic and control across feedback conditions. Mean for amusics is shown, both including and excluding IC.

B. Stimuli and equipment

Stimuli consisted of vocal tones with a duration of 2000 ms presented over Beyerdynamic DT770 pro headphones (Beyerdynamic GmbH & Co. KG, Heilbronn, Germany). Tones were presented at five pitches for each participant; males heard tones at D#3 (156 Hz), F3 (175 Hz), G#3 (208 Hz), B3 (247 Hz), and C#4 (277 Hz) in a male voice, and females heard female vocal tones 1 octave above those presented to males. Participants were visually cued on a computer screen to sing back the note for 4 s. All vocalizations were captured using a Shure 565D microphone (Shure Inc., Niles, IL) and recorded onto a Marantz PMD-670 digital recorder (Marantz Professional, Itasca, IL). The vocal signal was routed through a VoiceOne digital signal processor (TC Helicon Vocal Technologies, Westlake Village, CA) before it was presented to the participant as auditory feedback; for one particular task, the digital signal processor was programed to turn off the auditory feedback (see Sec. II C). We also presented pink noise to ensure that the main source of auditory feedback would stem from the digital signal processor. All auditory stimuli, including the participant's own vocal feedback, were delivered to the headphones through a mixer, where all volume levels were adjusted to comfortable levels for each participant. Pink noise was delivered at an average of 73.3 dB sound pressure level (SPL) A, and the vocal tones were presented an average of 9.7 dB above the pink noise. The experiment was run using MEDIA CONTROL FUNCTIONS (MCF) software (DigiVox, Montreal, Canada).

C. Procedure and design

Three response conditions were presented to each participant: alone, guided, and masked. In the alone conditions, participants heard the stimulus tone and attempted to vocally match it after the target tone finished sounding. In the guided condition, participants heard the stimulus tone and attempted to match it while the tone was presented again for 4 s. The masked condition had participants match the target tone while not receiving any vocal feedback; pink noise was the only auditory stimulus delivered to the headphones as participants sang. The experiment was composed of 120 trials; however, due to fatigue, some participants completed fewer than this (see Table I).

III. ANALYSIS AND RESULTS

A. Pitch analysis

Pitch analyses were conducted using YIN, a pitch analysis program available in MATLAB, which outputs fundamental frequency, aperiodicity, and signal power every 32 samples or a rate of 1378 Hz (de Cheveigné and Kawahara, 2002). These fundamental frequency outputs were converted into cents, using the originally presented vocal tones as the target frequencies, in order to allow for meaningful comparisons between trials on different pitches. For each sung tone, we analyzed the mean pitch, excluding the first and last 100 ms of each vocalization. The absolute errors were used in analyses concerning pitch errors to avoid sharp and flat errors canceling each other out, but pure pitch heights were used for comparisons between mean produced pitch heights and targets.

B. Data analysis

Figure 2 shows mean pitch errors for each amusic and control across alone, guided, and masked conditions. Results were computed with the Greenhouse–Geisser corrections when the assumptions of sphericity were violated. A 3×5

TABLE II. Correlation for each amusic and controls between target and produced pitches (as measured in cents), as well as average standard error for each subject at each target pitch. Asterisks denote that the correlation is significant. The numbers in parentheses next to the note names represent the relative distance in cents from C3 in males and C4 in females.

		Std. error				
	Correlation	D# (300)	F (500)	G# (800)	B (1100)	C# (1300)
IC	0.13	15.54	13.12	17.86	17.64	13.62
GC	0.92*	23.57	25.95	19.58	22.50	21.33
FA	0.81*	103.10	118.91	137.19	75.58	101.18
MB	0.82*	55.17	67.06	39.79	37.37	29.48
AS	0.89*	20.01	43.29	54.71	44.74	14.46
JL	0.85*	24.71	37.12	48.78	49.80	25.50
Controls	0.99*	11.45	7.89	6.57	9.33	6.91

$\times 2$ mixed design analysis of variance (ANOVA) on the mean pitch errors, comparing the three feedback conditions (alone, guided, and masked), five pitch heights, and two participant groups (amusic and control) revealed the main effect of participant group to be approaching significance ($F(1,8)=4.99$, $p=0.056$). A non-parametric Mann–Whitney U test was subsequently run comparing amusics and controls, due to the small sample size, which showed a significant difference of participant group [$U(4,6)=0$, $p=0.01$]. Amusics had significantly larger errors in pitch matching than did controls. No other main effects or interactions reached significance. Further inspection of the data revealed that one amusic participant (IC) did not substantially change the pitch of his vocalizations across the different target pitches. An ANOVA comparing the pitches of each of IC's vocalizations (in fixed terms rather than relative to the target pitch) showed no differences between the productions for each target pitch $F(4,114)=0.53$, n.s. Notably, this participant also showed the lowest global MBEA score, by a wide margin (see Table I). When IC was removed from the analysis, the pattern of results remained the same as presented above, with the main effect of participant group rising just above significance in the parametric test [$F(1,7)=6.07$, $p=0.043$].

We also examined correlations between the pitch height of the sung tones and of the targets. Using each of the six amusics' average pitch heights (in fixed terms relative to C3 in males and C4 in females) for each of the five target pitches, this correlation was $r(29)=0.72$, $p<0.01$. When IC was removed from the analysis, this correlation rose to $r(24)=0.94$, $p<0.01$. The correlations for each amusic and control participants are presented in Table II, as well as the standard errors for each participant, as a measure of precision. Figure 3 shows the individual pitch-matching performances for each target note collapsed across feedback conditions, as well as regression and R^2 values. With the exception of IC, amusics reliably produced higher pitches in response to higher targets. Amusics produced responses to target pitches with standard errors less than half a semitone, which is small enough to make their responses distinguishable from responses to other target pitches, and shows that there was not much overlap between their attempted pitch matches to different target pitches. These relationships between target and produced pitches did not differ between feedback conditions.

IV. DISCUSSION

These data demonstrate that amusics are indeed significantly worse than controls at matching pitches vocally. Although there was considerable variation between the amusics' performances, the most accurate amusic singer (GC) was still much less accurate than any matched control subject. The perceptual deficits shown by amusics thus seem to be accompanied by production deficits. However, despite their inaccurate performance, five of the six amusics showed a very clear relationship between the target pitch and their produced pitch, with R^2 values ranging from 0.66 to 0.85. For each target pitch, these amusics typically had a constant degree of pitch error, with average pitch responses ranging from very close to the target pitch, as in the case of GC, to averaging nearly three semitones flat, in the case of JL. The amusics typically had standard error measurements averaging around 50 cents. This shows a good deal of vocal precision and provides evidence that they are producing a different response to each of the five pitch heights; however, none reached the level of precision attained by the controls. The vocal precision shown by these five amusic subjects is surprising, given their lack of ability to perform perceptual decision tasks, such as detecting an anomalous note in a melody. It should be noted that the five target pitches presented spanned a range of ten semitones, covering a large vocal range. Because each target tone differed from its nearest neighbors by 200 or 300 cents, the targets were above the discrimination threshold of the amusics. Their ability to discriminate between the different target tones may explain the pattern of results. However, this is unlikely to be the sole reason for their ability to produce different responses to each tone, given that, in other experimental paradigms, amusics have shown the ability to produce vocal responses, which discriminate between targets differing by less than a semitone (Loui *et al.*, 2008).

The different response range for each target pitch and high correlation between targets and produced pitches indicates a fairly well-preserved pitch production ability. This shows evidence that the link between the input and vocal output is not completely severed. However, even among the

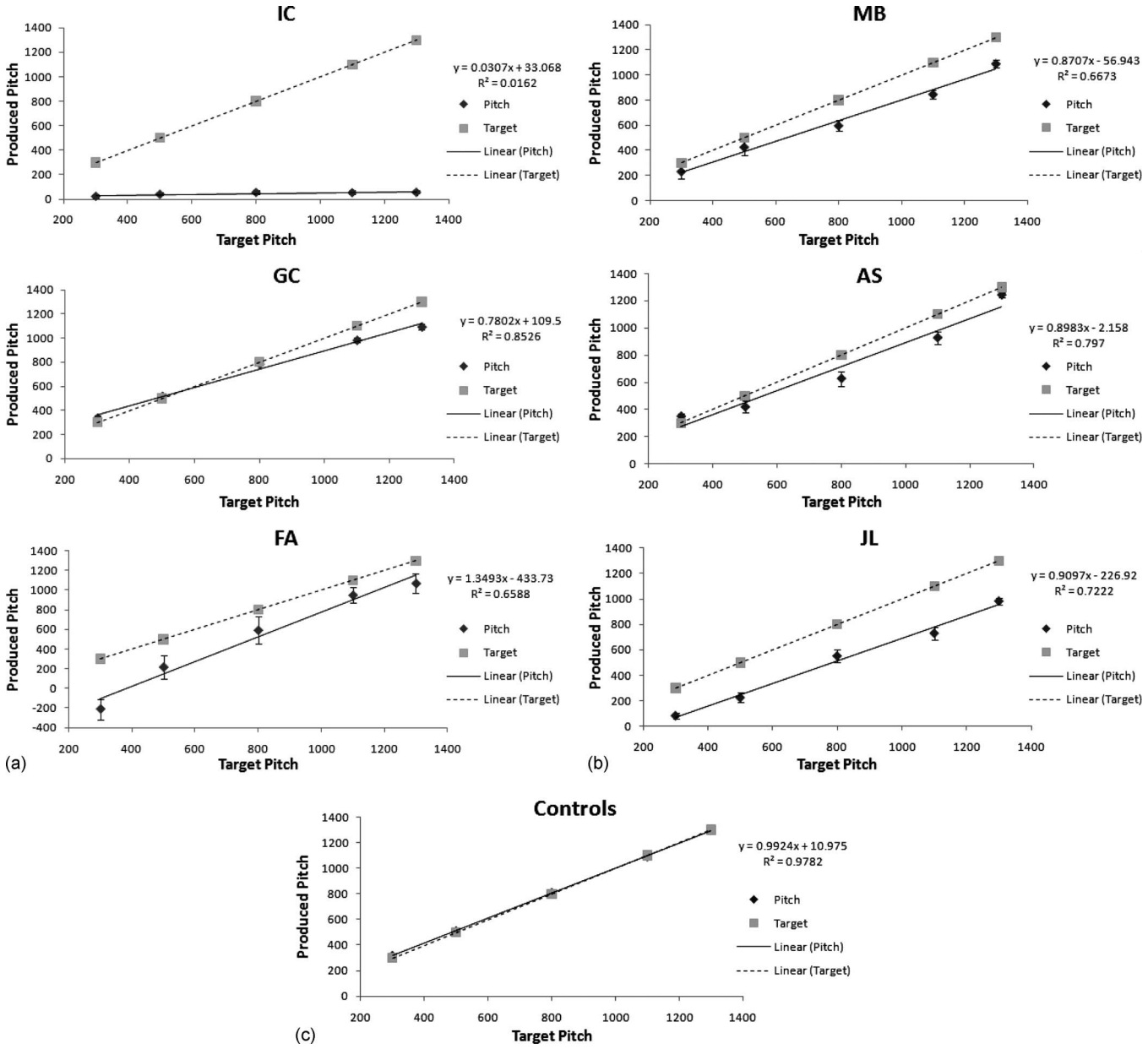


FIG. 3. The mean pitch (in cents) produced by each amusic and all controls for each of the five pitch heights with standard error bars and regressions comparing their performance to the targets and R^2 values. The data are collapsed across octaves between genders. 0 cents here refers to C3 for males and C4 for females.

five amusics who showed this pattern, there was a fairly constant downward shift from target pitches to produced pitches. Each amusic had a different degree of shift present in their responses, but all shifts were more or less constant across target pitches (with a slight tendency for greater downward shifts at higher target pitches). It is unlikely that this shift is a result of vocal range limitations, as it occurs nearly equally at all target pitches. Such a shift may be systematic, meaning that these amusics have distinct representations of vocal-motor codes for each pitch, which are linearly related to the auditory input. In this case, a constant shift may arise from their perceptual deficiencies, perhaps from an inadequate comparison between the vocal output and the target, which would be consistent with the lack of effect found for feedback condition in this experiment as well.

An alternative account for this shift is that each pitch may be mapped onto vocal productions in a manner that is based on the target range and is non-linear, but still preserving order. In this account, target pitches are categorized relative to the range of targets in a manner similar to Braida and Durlach's (1988) and Durlach and Braida's (1969) context coding mode in intensity perception. However, this account relies on positing that target pitches are constantly being compared to each other in memory, which is not necessary for a straight pitch-matching task. However, both of these accounts hold that the unique mapping between targets and vocal productions is no coincidence. Most amusics can perceive these pitch differences and possess the vocal control to produce different responses to each pitch.

In contrast to these patterns, IC showed no relationship between target pitches and his productions. This subject is

one of the most severe cases of amusia and showed no indication of distinguishing between different pitch heights. This trend provides a sharp contrast to that of the other five amusics, who did show different responses to each pitch. Other studies with IC have indicated that he is able to produce pitches at different heights (Dalla Bella *et al.*, 2009), demonstrating that this pattern of results is not due to total vocal monotonicity. IC may represent a subpopulation of amusics who have severely impaired vocal production abilities in addition to their impaired perceptual abilities. A task for future research will be to determine how many different subtypes of amusia may exist, and what their relative prevalence may be.

The considerable degree of pitch-matching ability showed by most amusics demonstrates a partially preserved, but consciously inaccessible, perception ability. The fact that some perceptual abilities can be shown among amusics in this and other production-based experiments (Dalla Bella *et al.*, 2009; Loui *et al.*, 2008) provides evidence that vocal production does not necessarily require a conscious perceptual representation. Amusics' relatively high level of precision in the vocal pitch-matching task (less than half a semitone standard error) suggests that a motoric representation may exist apart from conscious perception. This study highlights an association between pitch input and vocal production output as well as a dissociation between vocal production and discrimination judgments as output measurements. A model that assumes first a coding of pitch based on its perceptual attributes, which is subsequently translated into a motor representation for production, would have a difficult time explaining this pattern. While we do not wish to oversell the vocal abilities of amusics, which are still considerably worse than controls, the fact that motoric types of responses tend to display better underlying knowledge of pitch than perceptual types of responses in amusics argues against this viewpoint, and for a direct link between perception and a motor representation of pitch. Although this pattern of results cannot directly discriminate between a motor theory and a dual-route theory, the fact that other individuals whose pitch discrimination and categorization abilities outstrip their singing abilities exist (see, for example, Pfordresher and Brown, 2007; Dalla Bella *et al.*, 2007) suggests a dual-route theory.

A dual-route model is also supported by neuroscience evidence. Specifically, several authors have suggested that auditory-motor integration, in general, likely involves interactions between posterior-dorsal auditory cortex and premotor regions (Hickok *et al.*, 2003; Zatorre *et al.*, 2007). In the case of auditory-vocal control for singing, a network has been identified that includes the auditory cortex, insula, supplementary motor area, and anterior cingulate (Brown *et al.*, 2004; Kleber *et al.*, 2007; Perry *et al.*, 1999; Özdemir *et al.*, 2006), together with parietal and premotor cortical regions (Zarate and Zatorre, 2008; Zarate *et al.*, 2009), which are specifically important for auditory-vocal integration. These cortical regions are primarily found within the dorsal stream of auditory processing, implying that it is this system that is responsible for transforming auditory representations of pitch information into vocal production of pitch, as is the case for speech, but with more of a right-hemisphere bias for pitch. Critically, recent studies of anatomical anomalies in amusia have identified anteroventral frontal regions (Brodmann area 47) and anteroventral temporal auditory cortices (BA 22) as sites of likely dysfunction (Hyde *et al.*, 2006; Hyde *et al.*, 2007; Mandell *et al.*, 2007). Putting together these anatomical studies which suggest that the ventral pathway is compromised, with the functional studies in normal singing showing that the dorsal pathway is recruited for pitch production, leads to the conclusion that the relatively spared singing capacity of amusics uncovered in the present study may well be the result of a partly functional dorsal pathway, whereas the major impairment in conscious pitch perception may be due to disconnection in a ventral route. Recent diffusion tractography data (Loui *et al.*, 2009) showed reduced connectivity in the dorsal pathway of amusics, however. This finding may be related to the fact that pitch matching is still impaired in amusics, even if it is relatively more preserved than pitch discrimination or may reflect differences in the populations studied. Nonetheless, such data will have to be reconciled with models proposing that auditory-motor interactions in speech take place primarily via a dorsal pathway (Hickok and Poeppel, 2004; Rauschecker and Scott, 2009), as they also do in vision (Milner and Goodale, 2008); this type of model would be in keeping with the proposal of a relative sparing of this system in amusia to the extent that the present data, and the data of Loui *et al.* (2008), demonstrate a greater degree of spared pitch production than perception.

The fact that the amusics' performance was not affected by feedback condition may indicate that they are behaving similarly to musically normal participants in this respect. The matched controls were neither aided nor hindered by providing more target pitch information (guided condition) or by masking the feedback through noise (masked condition), compared with the unaltered feedback condition (alone condition). This is similar to the effect found by Pfordresher and Brown (2007) for the monotonic (no pitch change) sequences. Their study used similar feedback conditions as the ones used here and found no effect of masking on singing performance. However, their guided condition led to different results for different skill groups. Among the accurate-singing non-musicians, guiding aided the production of the more complex stimuli (with multiple pitch changes). However, guiding tended to engender even less accurate singing among the 13% of participants who were poor-pitch singers. This inaccurate singing could not be attributed to perceptual difficulties. Our current data show a similar lack of feedback effect in all groups, although we do not see a hindrance from the guided condition. The lack of effect of feedback on performance may indicate that perception does not influence production control of single sung tones, and that perception and production constitute separable streams. However, this non-effect of feedback may have different root causes in amusics and normals. Whereas controls may not need to rely on external feedback to control their voices, amusics' perceptual difficulties may render them unable to recognize potentially useful sources of pitch feedback.

It also should be noted that all congenital amusics in this study were a great deal worse at the music perception tasks than non-amusics (by definition). Although there were not enough data to provide reliable correlations, there was gen-

erally better performance on the production task among amusics who scored higher on the MBEA. The performance of individual amusics also seems to be consistent between this pitch-matching task and the melodic singing data reported by Dalla Bella *et al.* (2009), which included the data of five of our six amusic participants. IC, who was categorized as particularly error-prone in that study, had a very high mean error in this study as well, and GC, who Dalla Bella *et al.* (2009) classified as a particularly good amusic singer, was also good, although still impaired, in the present pitch-matching study. This suggests consistency between pitch-matching and melodic singing tasks.

In conclusion, these data confirm that amusics do show impaired pitch-matching abilities compared to controls, across different pitch heights and feedback conditions. However, among amusics, there is a range of production abilities. The deficit in pitch matching can be attributed to their core deficit in pitch perception, and constitutes one of the most noticeable effects of amusia in daily life. As more research becomes available on the pitch-matching abilities of neurologically normal adults, finer comparisons between amusics and non-amusics on this subject will be possible.

ACKNOWLEDGMENTS

The authors would like to thank Nathalie Gosselin for her assistance in testing all participants, and Olivier Piché, Miranda Tradewell, and Sean Wood for their technical assistance. The research was supported by grants from Natural Sciences and Engineering Research Council of Canada, the Canadian Institutes of Health Research, and from a Canada Research Chair to Isabelle Peretz, a grant from the Canadian Institutes of Health Research to Robert Zatorre, and an Eileen Peters McGill Majors fellowship and a Centre for Interdisciplinary Research in Music Media and Technology (CIRMMT) student award to Jean Mary Zarate. S.H. and J.M.Z. are equally contributing authors.

Anstis, S. M., and Cavanagh, P. (**1979**). "Adaptation to frequency-shifted auditory feedback," Percept. Psychophys. **26**, 449–458.

Ayotte, J., Peretz, I., and Hyde, K. (**2002**). "Congenital amusia: A group study of adults afflicted with a music-specific disorder," Brain **125**, 238–251.

Braida, L. D., and Durlach, N. I. (**1988**), "Peripheral and central factors in intensity perception," in *Auditory Function*, edited by G. Edelman, W. Gall, and W. Cowan (Wiley, New York), pp. 559–583.

Brown, S., Martinez, M. J., Hodges, D. A., Fox, P. T., and Parsons, L. M. (**2004**). "The song system of the human brain," Brain Res. Cognit. Brain Res. **20**, 363–375.

Burnett, T. A., Freedland, M. B., Larson, C. R., and Hain, T. C. (**1998**). "Voice F0 responses to manipulations in pitch feedback," J. Acoust. Soc. Am. **103**, 3153–3161.

Burnett, T. A., and Larson, C. R. (**2002**). "Early pitch-shift response is active in both steady and dynamic voice pitch control," J. Acoust. Soc. Am. **112**, 1058–1063.

Dalla Bella, S., Giguère, J.-F., and Peretz, I. (**2007**). "Singing proficiency in the general population," J. Acoust. Soc. Am. **121**, 1182–1189.

Dalla Bella, S., Giguère, J.-F., and Peretz, I. (**2009**). "Singing in congenital amusia: An acoustical approach," J. Acoust. Soc. Am. **126**, 414–424.

Danckert, J., and Rossetti, Y. (**2005**). "Blindsight in action: What can the different sub-types of blindsight tell us about the control of visually guided actions?," Neurosci. Biobehav Rev. **29**, 1035–1046.

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

Durlach, N. I., and Braida, L. D. (**1969**). "Intensity perception. I. Preliminary theory of intensity resolution," J. Acoust. Soc. Am. **46**, 372–383.

Elliott, L. L., and Niemoeller, A. F. (**1970**). "The role of hearing in controlling voice fundamental frequency," Int. J. Audiol. **9**, 47–52.

Finney, S. A. (**1997**). "Auditory feedback and musical keyboard performance," Music Percept. **15**, 153–174.

Goetze, M., Cooper, N., and Brown, C. J. (**1990**). "Recent research on singing in the general music classroom," Bull. Council Research in Music Education **104**, 16–37.

Goodale, M. A., and Milner, A. D. (**1992**). "Separate visual pathways for perception and action," Trends Neurosci. **15**, 20–25.

Griffiths, T. D. (**2008**). "Sensory systems: Auditory action streams?," Curr. Biol. **18**, R387–R388.

Hafke, H. Z. (**2008**). "Nonconscious control of fundamental voice frequency," J. Acoust. Soc. Am. **123**, 273–278.

Hickok, G., Buchsbaum, B., Humphries, C., and Muftuler, T. (**2003**). "Auditory-motor interaction revealed by fMRI: Speech, music, and working memory in area SPT," J. Cogn Neurosci. **15**, 673–682.

Hickok, G., and Poeppel, D. (**2004**). "Dorsal and ventral streams: A framework for understanding aspect of the functional anatomy of language," Cognition **92**, 67–99.

Hyde, K. L., Lerch, J. P., Zatorre, R. J., Griffiths, T. D., Evans, A. C., and Peretz, I. (**2007**). "Cortical thickness in congenital amusia: When less is better than more," J. Neurosci. **27**, 13028–13032.

Hyde, K. L., and Peretz, I. (**2004**). "Brains that are out of tune but in time," Psychol. Sci. **15**, 356–360.

Hyde, K. L., Zatorre, R. J., Griffiths, T. D., Lerch, J. P., and Peretz, I. (**2006**). "Morphometry of the amusic brain: A two-site study," Brain **129**, 2562–2570.

Jones, M. (**1979**). "Using a vertical keyboard instrument with the uncertain singer," J. Res. Music Educ. **27**, 173–184.

Kleber, B., Birbaumer, N., Veit, R., Trevorrow, T., and Lotze, M. (**2007**). "Overt and imagined singing of an Italian aria," Neuroimage **36**, 889–900.

Liberman, A. M., and Mattingly, I. G. (**1985**). "The motor theory of speech perception revised," Cognition **21**, 1–36.

Loui, P., Alsop, D., and Schlaug, G. (**2009**). "Tone deafness: A new disconnection syndrome?," J. Neurosci. **29**, 10215–10220.

Loui, P., Guenther, F., Mathys, C., and Schlaug, G. (**2008**). "Action-perception mismatch in tone-deafness," Curr. Biol. **18**, R331–R332.

MacKay, D. G. (**1987**). *The Organization of Perception and Action* (Springer, New York).

Mandell, J., Schulze, K., and Schlaug, G. (**2007**). "Congenital amusia: An auditory-motor feedback disorder?," Restor. Neurol. Neurosci. **25**, 323–334.

Milner, A. D., and Goodale, M. A. (**1995**) *The Visual Brain in Action* (Oxford University Press, Oxford).

Milner, A. D., and Goodale, M. A. (**2008**). "Two visual systems re-viewed," Neuropsychologia **46**, 774–785.

Mürbe, D., Pabst, F., Hofmann, G., and Sundberg, J. (**2002**). "Significance of auditory and kinesthetic feedback to singers' pitch control," J. Voice **16**, 44–51.

Natke, U., Donath, T. M., and Kalveram, K. T. (**2003**). "Control of voice fundamental frequency in speaking versus singing," J. Acoust. Soc. Am. **113**, 1587–1593.

Özdemir, E., Norton, A., and Schlaug, G. (**2006**). "Shared and distinct neural correlates of singing and speaking," Neuroimage **33**, 628–635.

Pederson, D. M., and Pederson, N. O. (**1970**). "The relationship between pitch recognition and vocal pitch production in sixth-grade students," J. Res. Music Educ. **18**, 22–33.

Peretz, I. (**2001**). "Brain specialization for music," Ann. N.Y. Acad. Sci. **930**, 153–165.

Peretz, I., Brattico, E., Järvenpää, M., and Tervaniemi, M. (**2009**). "The amusic brain: In tune but unaware," Brain **132**, 1277–1286.

Peretz, I., Champod, A. S., and Hyde, K. (**2003**). "Varieties of musical disorders: The Montreal battery of evaluation of amusia," Ann. N.Y. Acad. Sci. **999**, 58–75.

Perry, D. W., Zatorre, R. J., Petrides, M., Alivisatos, B., Meyer, E., and Evans, A. C. (**1999**). "Localization of cerebral activity during simple singing," NeuroReport **10**, 3979–3984.

Pfordresher, P. Q. (**2005**). "Auditory feedback in music performance: The role of melodic structure and musical skill," J. Exp. Psychol. Hum. Percept. Perform. **31**, 1331–1345.

Pfordresher, P. Q., and Brown, S. (**2007**). "Poor-pitch singing in the absence of "tone deafness"," Music Percept. **25**, 95–115.

Rauschecker, J. P. (**1998**). "Cortical processing of complex sounds," Curr.

Opin. Neurobiol. **8**, 516–521.
Rauschecker, J. P., and Scott, S. K. (**2009**). "Maps and streams in the auditory cortex: Nonhuman primates illuminate human speech processing," Nat. Neurosci. **12**, 718–724.
Ternström, S., Sundberg, J., and Colldén, A. (**1988**). "Articulatory F0 perturbations and auditory feedback," J. Speech Lang. Hear. Res. **31**, 187–192.
Warren, J. D., and Griffiths, T. D. (**2003**). "Distinct mechanisms for processing spatial sequences and pitch sequences in the human auditory brain," J. Neurosci. **23**, 5799–5804.
Warren, J. E., Wise, R. J., and Warren, J. D. (**2005**). "Sounds do-able: Auditory-motor transformations and the posterior temporal plane," Trends Neurosci. **28**, 636–643.
Watts, C., Barnes-Burroughs, K., Adrianopolous, M., and Carr, M. (**2003**). "Potential factors related to untrained singing talent: A survey of singing pedagogues," J. Voice **17**, 298–307.
Wise, K. J., and Sloboda, J. A. (**2008**). "Establishing an empirical profile of self-defined "tone deafness:" Perception, singing performance and self-assessment," Music. Sci. **12**, 3–26.
Zarate, J. M., Wood, S., and Zatorre, R. J. (**2009**). "Neural networks involved in voluntary and involuntary vocal pitch regulation in experienced singers," Neuropsychologia In press.
Zarate, J. M., and Zatorre, R. J. (**2008**). "Experience-dependent neural substrates involved in vocal pitch regulation during singing," Neuroimage **40**, 1871–1887.
Zatorre, R. J., Chen, J. L., and Penhune, V. B. (**2007**). "When the brain plays music: Auditory-motor interactions in music perception and production," Nat. Rev. Neurosci. **8**, 547–558.

Perceptual studies of violin body damping and vibrato

Claudia Fritz[a)]
Institut Jean Le Rond d'Alembert, Université Pierre et Marie Curie, UMR CNRS 7190, 4 place Jussieu, 75005 Paris, France

Jim Woodhouse and Felicia P.-H. Cheng
Department of Engineering, University of Cambridge, Trumpington Street, Cambridge CB2 1PZ, United Kingdom

Ian Cross
Centre for Music and Science, Music Faculty, University of Cambridge, West Road, Cambridge CB3 9DP, United Kingdom

Alan F. Blackwell
Computer Laboratory, University of Cambridge, 15 JJ Thomson Avenue, Cambridge CB3 OFD, United Kingdom

Brian C. J. Moore
Department of Experimental Psychology, University of Cambridge, Downing Street, Cambridge CB2 3EB, United Kingdom

(Received 10 July 2009; revised 2 October 2009; accepted 16 October 2009)

This work explored how the perception of violin notes is influenced by the magnitude of the applied vibrato and by the level of damping of the violin resonance modes. Damping influences the "peakiness" of the frequency response, and vibrato interacts with this peakiness by producing fluctuations in spectral content as well as in frequency and amplitude. Initially, it was shown that thresholds for detecting a change in vibrato amplitude were independent of body damping, and thresholds for detecting a change in body damping were independent of vibrato amplitude. A study of perceptual similarity using triadic comparison showed that vibrato amplitude and damping were largely perceived as independent dimensions. A series of listening tests was conducted employing synthesized, recorded, or live performance to probe perceptual responses in terms of "liveliness" and preference. The results do not support the conclusion that liveliness results from the combination of the use of vibrato and a "peaky" violin response. Judgments based on listening to single notes showed inconsistent patterns for liveliness, while preferences were highest for damping that was slightly less than for a reference (real) violin. In contrast, judgments by players based on many notes showed preference for damping close to the reference value.
 [DOI: 10.1121/1.3266684]

PACS number(s): 43.75.De, 43.66.Lj, 43.75.Cd, 43.66.Jh [DD] Pages: 513–524

I. INTRODUCTION

Vibrato is the intentional modulation of the frequency and amplitude of a musical tone. It is an attribute of many sounds in the contemporary western musical world and is generally linked to what might be termed "expressive performance" (Brown, 1988). It is employed in a wide range of musical contexts as one of the resources in the palette of expressive strategies available to a performer. It is typically applied at rates around six cyclic fluctuations per second, in line with the rates found for other expressive ornaments such as tremolos and trills (Moelants, 2004). Typical frequency excursions in vibrato are about 2% of the fundamental frequency, but they can range up to 4% (Prame, 1994, 1997). These frequency excursions have been shown to affect the rapidity with which listeners can make judgments of relative pitch (Yoo *et al.*, 1998), but they have a minimal effect on perceived pitch (van Besouw *et al.*, 1996). Vibrato has become an integral constituent of contemporary technique on several string and wind instruments (Moens-Haenen, 2009), and many performers employ it as part of the process of imparting desirable timbral qualities to the musical sounds.

Vibrato may help to define the timbre of a musical instrument since the frequency modulation of the components causes them to move relative to the resonant frequencies that are characteristic of the instrument or voice, defining the center frequencies of the resonances more precisely (McAdams, 1989). McAdams and Giordano (2009) noted that "vibrato may increase our ability to extract information relative to the resonance structure of the instrument." The question of whether or not the use of vibrato has consistent effects that relate to the perceived timbre of musical sounds remains open.

For the particular case of the violin, Fletcher and Sanders (1967) were probably the first to suggest that fluctuations

[a)]Author to whom correspondence should be addressed. Electronic mail: claudia.fritz@upmc.fr

of spectral content due to vibrato are important for the perception of timbre. The frequencies of some harmonics will fall on positive slopes of the frequency response curve of the violin body, while others fall on negative slopes. The simple frequency modulation from the player's finger movement is then converted by the resonant body response "filter" of the instrument into a complex spectral modulation in which the amplitudes of different harmonics change in different ways.

A few years later Mathews and Kohut (1973) and Gorrill (1975) pioneered the methodology that will be exploited in this paper. Mathews and Kohut (1973) studied the effect of damping of the violin body resonances by creating violin sounds (a G major scale, performed with vibrato) using a near-silent electric violin and a set of analog filters: 20 resonant circuits were tuned to major resonance frequencies measured from a Stradivarius violin. Four values of damping were chosen, but the authors do not give clear details: one corresponded to a completely flat response (infinite damping), the second to a 10-dB peak-to-valley ratio in the response curve, and the remaining two corresponded to bigger ratios. They used their filtered stimuli to conduct informal listening tests using preference judgments. Their second value of damping, corresponding to a 10-dB peak-to-valley ratio, was preferred by their jury. They also reported that when the damping was too low the sound took on a "hollow" quality, which was found to be unpleasant.

McIntyre and Woodhouse (1974) suggested that this hollow quality might arise from transient excitation of the body resonances. The response to vibrato in the bowed-string input waveform does not consist simply of harmonics which track up and down the slopes of the response curve, as they would if the vibrato rate were extremely slow, because the typical rate of vibrato has a similar timescale to the free decay times of the body resonances. They found that simulations involving high-Q resonances but in which this transient effect was artificially omitted did not sound hollow, whereas with the transient effect the hollowness was heard.

More recent work has examined the extent to which the perceived qualities of vibrato tones derive from modulations of frequency or amplitude. Mellody and Wakefield (2000) conducted analyses of real violin sounds produced with vibrato and resynthesized approximations to those tones, either co-varying both frequency and amplitude, as in the original sounds, or varying these parameters independently. They found that the absence of frequency modulation had little effect on perceptual judgments, while the absence of amplitude modulation resulted in large perceptual changes.

Alongside this scientific perspective there is strong anecdotal evidence based on the intuitions of violinists about their strategies in using vibrato and the timbral effects they expect it to achieve. The response of a violin to the use of vibrato is widely considered to be one of the key factors in the performer's perception of the "responsiveness" of a particular instrument (Gough, 2005). Mathews and Kohut (1973), in the study described earlier, reported that when the frequency response was flat, the instrument seemed "unresponsive." As a starting point for this study, we hypothesized that the desirable quality of "liveliness" or responsiveness in a violin may be connected with the interaction of vibrato with the "peaky" frequency response of the violin.

In another study by the present authors (Fritz *et al.*, 2008), experienced violinists were asked to arrange a collection of 61 words that were found to be commonly used to describe violin timbre on a two-dimensional grid so that words which were similar in meaning were close together and words with very different meanings were far apart. The results were analyzed using multidimensional scaling and led to a three-dimensional map, which showed that the word "lively" was considered similar to "alive," "resonant," "ringing," and "responsive," and as opposite to "dead" and "dull." Therefore, lively and responsive will be used interchangeably in this study.

The present study used a method analogous to that of Mathews and Kohut (1973) to explore the perceived quality and discriminability of violin timbre when varying vibrato amplitude and the resonance damping of the violin within the same experimental framework and to probe the perceptual consequences in a series of studies. The authors were surprised by the results from every stage of the study. A succession of tests was designed, in each case with fairly clear expectations based on the intuitions of players, and those expectations were regularly not supported. This may indicate that the anecdotal evidence is wrong, but at least in some cases we feel that it points more strongly to the difficulty of designing tests which are focused enough to be quantitatively convincing without throwing out the "baby" of musical relevance with the "bathwater" of experimental control.

This is an important issue in its own right, underlying any study of musical psychoacoustics. The phrase "musical relevance" here covers two main aspects. First, very short sound samples tend to lose any musical quality to the listener, and repeated listening to similar sounds in a typical test erodes it still further. Second, there is the question of the realism or naturalness of the sounds used: if the test sounds are not close enough to the "training set" that a musical listener will have experienced from hearing real violin performances, it may be that finely honed perceptions will not be able to operate in the way that is intended. Indeed, if some perceptions are of a categorical nature they may not be evoked at all by sounds of insufficient naturalness.

II. GENERAL SOUND SYNTHESIS METHOD

The methodology is based on creating "virtual violins," as described in a previous study (Fritz *et al.*, 2007). The frequency response function of the violin is mimicked using a digital filter, and the output signal for listening tests is generated by applying this filter to an input signal representing the force exerted by the bowed string on the bridge of the violin. The main advantage of this approach lies in the fact that, once the violin response is represented in digital filter form, it becomes easy to make controlled variations of a kind which would be impossible to achieve by physical changes to a violin.

The input signal can be generated in three different ways, all used in different parts of this study. First, the bridge force may be recorded from a player using vibrato, on a

violin whose bridge has been instrumented with piezoelectric force sensors. The same recording can then be used with many different digital filters, thus removing the influence of the player and their adaptation to the instrument. Second, the bridge-force signal from an instrumented violin can be used in real time, passing the force signal to a digital filter system that generates the sounds the player hears. This approach works best if the violin body is essentially mute so that the only sound reaching the player's ears comes from the filtered signal. Third, for some tests the bridge-force signal can be synthesized.

In the case of purely synthesized force signals, the approach exploits the fact that the usual Helmholtz motion of a bowed string produces a bridge force in the form of a sawtooth waveform, at least to a first approximation (e.g., Cremer, 1985). This waveform is easy to synthesize. To take account of the frequency modulation associated with vibrato, each successive period requires a slightly different length. The following formula is used to determine the kth period length:

$$T_k = T_0(1 + a_{\text{vib}} \sin(2\pi k f_v T_0) + r_k), \quad (1)$$

where T_0 is the period corresponding to the nominal frequency of the note, f_v is the vibrato frequency in hertz (typically 5 Hz), a_{vib} is the vibrato amplitude (typically 0.02), and r_k is a random number drawn from a uniform distribution between $-R$ and $+R$, where R determines the amount of randomness. In what follows, a_{vib} and R are both multiplied by 100 to express them as percentages. The random element was included in an attempt to increase the naturalness of the synthesized sounds. In the course of the series of experiments reported here, several further small changes were tried in the quest to improve naturalness, as will be described below.

A window function was applied to the entire synthesized input signal to give a smooth envelope resembling a détaché legato bowed note. For the first test, described in Sec. III A, this was a Hanning window, but in later tests this choice was changed in an effort to improve the naturalness: following Gough (2005) the first 600 ms of the signal was multiplied by $(1-\exp(-t/T))$ and the last 100 ms by $\exp(-(t-600)/T)$ with $T=30$ ms. In these later tests, the whole signal was further processed to round off the "Helmholtz corners" of the ideal sawtooth waveform. This corner rounding was achieved by convolution with a Gaussian function whose width was initially chosen to be 3.5% of the period, a typical value from measured waveforms (see, for example, McIntyre *et al.*, 1981).

Filtering to represent the violin body was based on the bridge admittance frequency response of a good-quality modern violin made by David Rubio in 1992. The admittance was measured using a small impulse hammer and a laser vibrometer (see Fritz *et al.*, 2007 for details). The amplitude of the measured admittance is plotted as the solid line in Fig. 1. The dash-dot line in this figure shows a typical modification made to the response for the tests to be described: the modal damping factors have all been doubled. For comparison, the dashed line shows an approximation to one frequency response used by Mathews and Kohut (1973), the one with 10-dB peak-to-valley fluctuation, as preferred by their listening jury.

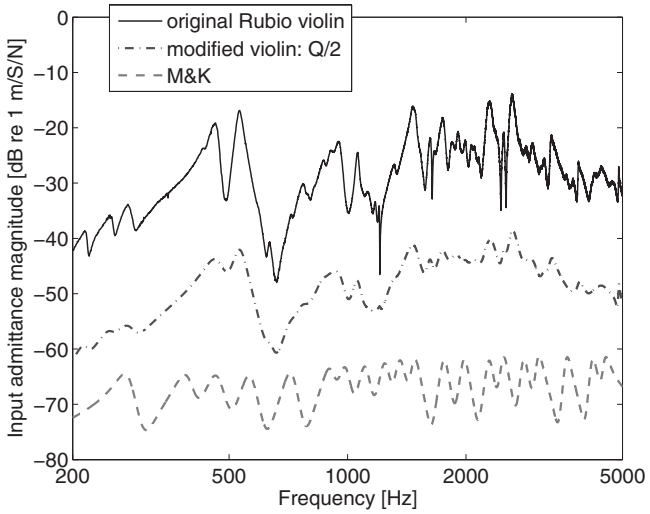


FIG. 1. The input admittance of the original Rubio violin (solid line), a modified version of it (dash-dot line), where all the Q values have been divided by 2, and a resynthesized version of the frequency response function of Mathews and Kohut (1973, Fig. 4b) (dashed line). The original curve is shown as a calibrated measurement (re 1 m/s/N), and the other curves have been shifted down for clarity.

In order to make changes such as the damping modification shown in Fig. 1, the measured frequency response was first analyzed into modal contributions using standard pole-residue fitting procedures (e.g., Ewins, 2000), and then resynthesized from these parameters (with or without some modification being made first). The fitting procedure covered the frequency range up to 7000 Hz and required 54 modes. The resynthesized response was used to construct a causal finite impulse response filter which was then applied to the chosen input signal. Damping is quantified throughout this work by the modal Q values: the Q value is the inverse of the damping factor, so high Q corresponds to low damping, and vice versa. The values of the fitted Q values for the Rubio violin are of some interest and are plotted in Fig. 2. The

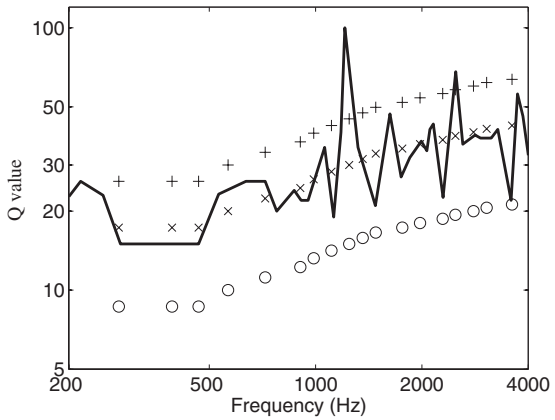


FIG. 2. Comparison of the Q values used here (deduced by modal fitting of the response of the Rubio violin; solid line) and those used by Mathews and Kohut (1973). These correspond to a 10-dB peak-to-valley ratio (○) and two higher ratios (× and +).

estimated values for the study of Mathews and Kohut (1973) are also shown.

III. THRESHOLD AND INTERACTION STUDIES

A. Experiment 1: Discrimination of vibrato amplitude

In a preliminary study (Cheng, 2006), thresholds were measured for detecting a change in vibrato amplitude (a_{vib}) using reference sounds with or without vibrato ($a_{vib}=0$). The sounds used were either the raw synthetic string signals or the same string signals filtered by the acoustical response of the Rubio violin, as described above. This gave a total of four different conditions.

1. Stimuli

All tests were based on the note G3 (fundamental frequency of 196 Hz, the lowest note of the violin). String signals were synthesized using the method described above, using a sampling rate of 44 100 Hz and 16-bit resolution. The vibrato rate (f_v) was 5 Hz and the randomness in amplitude R was 1%. The stimuli were 2 s in duration. They were presented diotically in a relatively quiet environment, via Sennheiser HD580 headphones, chosen because of their diffuse-field response and low distortion.

2. Procedure

Thresholds were estimated using a three-alternative forced-choice (3AFC) procedure. Three sounds, two the same (the reference violin sound with a fixed vibrato amplitude of either 0% or 2%) and one different (with a greater vibrato amplitude), were played in a random order, and the participant was asked to choose which was different. The vibrato amplitude was increased after a single incorrect response and decreased after three successive correct responses. The step size of these changes was initially a factor of 1.414 (relatively large for fast convergence toward the threshold region). After two turnpoints (changes from decreasing to increasing vibrato amplitude of the test sound and vice versa), the step size was reduced to a factor of 1.189. Eight turnpoints were obtained and threshold was taken as the geometric mean of the values of the vibrato amplitude at the last six turnpoints. Participants were given visual feedback during the experiment and were given some practice by performing the test twice. The thresholds shown here are those obtained for the second run.

3. Participants

The participants in this study were four experienced string players and seven other musicians (UK Grade 8). All subjects reported having normal hearing, although this was not checked. The two tests based on a reference sound with no vibrato used all 11 participants, while the two involving a reference sound with 2% vibrato amplitude used the four string players and five of the other musicians.

4. Results

The results did not differ for the two types of musicians, and so the results were averaged over all participants. When the reference sound had no vibrato, the threshold for detecting the vibrato was 0.5% for the raw string signal and 0.8% for the "Rubio"-filtered signal. When the reference sound had a vibrato amplitude of 2%, the thresholds for detection of a change were 2.9% and 3.2% for the unfiltered and filtered cases, respectively. Contrary to our expectations based on earlier findings (Mathews and Kohut, 1973; Meyer, 1992; Gough, 2005) thresholds were not lower for the filtered than for the unfiltered signals. Thus, the fluctuations in amplitude of individual harmonics produced by passing the vibrato string signal through the synthesized violin body did not lead to enhanced detection or discrimination of the vibrato. However, listeners may still be sensitive to fluctuations in amplitude induced by vibrato. Indeed, as described earlier, such fluctuations might be used to infer properties of violins, such as the degree of damping. To explore this, it was decided to carry out the inverse test to that described above: the threshold was measured for detecting a change in the Q values of the violin response, using an input signal with no vibrato and with two fixed amplitudes of vibrato.

B. Experiment 2: Effect of vibrato amplitude on the discrimination of damping

1. Stimuli

The stimuli (again the bottom violin note G3) were synthesized as before. Some details were changed from those used in experiment 1 in an attempt to improve the naturalness of the synthesized sound. The randomness in amplitude was decreased to 0.2% and the vibrato rate was increased to 6 Hz. The whole signal was "corner-rounded," as described in Sec. II. The duration of the sounds was shortened to 700 ms to enhance the effectiveness of echoic memory (Darwin *et al.*, 1972). Three amplitudes of vibrato were used: 0%, 1%, and 2%. Each raw signal was filtered, either using the input admittance of the Rubio violin (reference sound) or with a modified version of that admittance (test sound) resulting from a multiplication of all Q values by a factor whose value was varied to determine the threshold for discrimination.

2. Procedure

Thresholds for discriminating the reference and test sounds were estimated using the 3AFC procedure described earlier. This time, the quantity being varied was the scaling factor applied to all the modal Q values (an upward shift in all cases). The initial scaling factor was 2.8. The step size in the factor was 1.189 until two turnpoints had occurred and 1.091 thereafter.

3. Participants

In this and all subsequent experiments the participants were experienced violinists (UK grade 8), who practiced regularly. They were paid for their participation. For this experiment there were 14 participants whose hearing was checked to be normal (defined here and below as audiometric thresholds below 15-dB hearing level at the standard audiometric frequencies).

4. Results

A one-way within-subjects analysis of variance (ANOVA) showed no significant difference in threshold for the three vibrato amplitudes, including the case with no vibrato: in all cases the threshold Q scaling factor was 1.4. In other words, a 40% reduction in damping was required for "threshold." Again, the result seems surprising. One might have expected that an input signal with vibrato would lead to enhanced sensitivity to changes in damping since the frequency modulation produced by the vibrato would cause fluctuations in spectrum which might provide a cue related to the Q values. A possible explanation for the lack of effect of vibrato amplitude involves informational masking (Neff and Green, 1987): the auditory stimulus is more complex in the presence of vibrato, with more "irrelevant" information, and this may make the task of discriminating a change in damping more difficult, even though more information is being presented to the auditory system.

C. Experiment 3: The effect of vibrato amplitude and level of damping on perceptual dissimilarity

Experiments 1 and 2 showed, somewhat surprisingly, that the detection and discrimination of vibrato amplitude were not affected by the presence or absence of resonances in the frequency response, and that the detection of changes in Q value was not influenced by the amount of vibrato in the input signals. However, both of these experiments involved discrimination of stimuli varying along a single dimension. It could be argued that the results are not relevant to the type of judgments made when assessing differences in quality between virtual "instruments" with different amounts of vibrato. Violinists often describe such quality differences in terms of liveliness or responsiveness. Accordingly, an experiment was designed to investigate how both vibrato amplitude and damping influence listeners' judgment of differences between synthesized violin sounds.

The hypothesis to be tested was that perceptual similarity or dissimilarity may correlate with the richness or complexity of spectral fluctuations induced by the resonant body response when vibrato is applied. An intuitive aspect of a more responsive instrument is that it does not require a very large vibrato amplitude to evoke the level of spectral fluctuations for the desired richness. Many players might therefore expect to be able (to a degree) to compensate for inadequate responsiveness of an instrument by increasing the vibrato amplitude. Following this line of reasoning, perhaps the perceived degree of richness depends on the interaction of damping and vibrato amplitude. This idea was tested by obtaining ratings of dissimilarity for sounds which differed along two dimensions, amount of damping and vibrato amplitude.

1. Stimuli

Sound files were synthesized corresponding to all combinations of two different parameter variations: three values of the vibrato amplitude (1%, 2%, and 3%) and three sets of Q values (original, divided by 2, and multiplied by 2); for a detailed rationale of these values, see the account of experiment 7 below. The synthesis details were the same as for Sec. III B, except that this time some reverberation was added to further simulate natural listening conditions, as typically experienced with recorded sound. We used an industry-standard digital audio processing package (DIGIDESIGN PRO TOOLS), using a "small room" option, with a decay time of 1.42 s and diffusion of 87%.

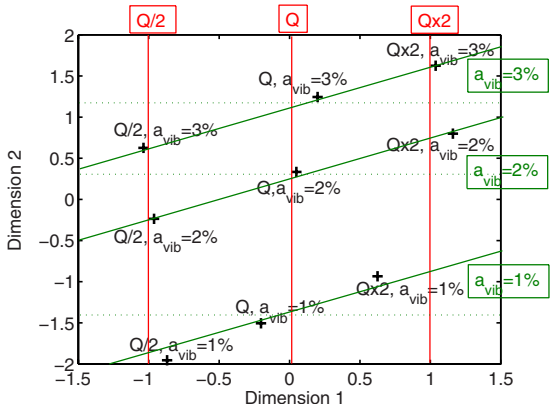


FIG. 3. (Color online) Perceptual space obtained using MDS for stimuli that were varied in two ways: (1) by changing the Q values of the violin resonances [Q values divided by 2 ($Q/2$), original Q values (Q) and Q values multiplied by 2 ($Q\times2$)] and (2) by changing the vibrato amplitude (a_{vib} equal to 1%, 2% and 3%).

2. Procedure and participants

The method of triadic comparisons was used (Wickelmaier and Ellermeier, 2007). Participants were presented with triads of sounds, each of which could be heard individually as often as desired, by clicking a button on the computer screen. For each triad of sounds, participants had to specify the most similar pair and the most dissimilar pair. Each participant listened to 84 different triads (all possible triadic combinations of the nine pairs of parameter values) plus the repetition of 13 of them to check consistency. The 14 participants reported having normal hearing (which was not checked).

3. Results

Responses were generally consistent across repeated stimuli; chance responses would have resulted in 16% of second responses being the same as initial responses, whereas our subjects performed at 56.5%.

A simple rating scale for dissimilarity was used to process the results: 2 points were allocated to the most dissimilar pair, 1 to the intermediate pair, and 0 to the most similar pair. By adding all participants' dissimilarity points for all triads, a dissimilarity matrix was constructed. This was then analyzed with the multi-dimensional scaling (MDS) algorithm ALSCAL available in SPSS. The two-dimensional map shown in Fig. 3 was obtained with an S-stress value of 0.03.

The two dimensions correspond approximately to the two physical parameters which were varied: vibrato increases along dimension 2 and the Q value along dimension 1. Q also maps partly onto dimension 2, but the overall shape indicates that the two physical parameters have largely independent perceptual effects. High Q values combined with a

low amplitude of vibrato (e.g., $Q\times 2$, $a_{vib}=1\%$) cannot be perceptually substituted for low Q values combined with a large amplitude of vibrato (e.g., $Q/2$, $a_{vib}=3\%$); these combinations of parameters are widely separated in the perceptual space. Contrary to our initial hypothesis, the interaction of vibrato and Q in determining perceived similarity is not large. There is some interaction, however, as evidenced by the parallelogram form in Fig. 3: the top-left and bottom-right corners are closer than the top-right and bottom-left corners.

Dimension 2 appears to be mainly related to the vibrato amplitude while dimension 1 appears to be mainly related to the Q values and may reflect a perceptual dimension that is related to spectral smoothness versus spectral unevenness in the acoustic properties of the violin sound. The perceptual effect of a modification of the Q values was approximately uniform on a logarithmic scale, with the map location for Q lying in the middle of the locations for $Q/2$ and $Q\times 2$ along dimension 1. The perceptual effect of an increase in the vibrato amplitude did not fit well to either a linear or a logarithmic interpretation: the distance between 1% and 2% is approximately twice the distance between 2% and 3%.

IV. PERCEPTUAL CORRELATES OF VIOLIN VIBRATO

A. Experiments 4a and 4b: Effects of damping and vibrato amplitude on judgments of liveliness

1. Stimuli

Experiment 3 revealed the perceptual dimensions associated with change in damping and vibrato amplitude, but did not establish whether these were associated with changes in the quality of liveliness often described by violinists; one might expect that a decrease in the vibrato amplitude and/or the Q values would reduce the perceived liveliness. A linked pair of experiments was performed to probe this question.

For both experiments, sounds were synthesized exactly as for experiment 3. To investigate fully the influence of vibrato, cases were included with a very small amount of vibrato and with no vibrato at all. For experiment 4a, the range of the vibrato amplitudes was therefore increased relative to those used for experiment 3: the values 0%, 0.5%, 1%, 2%, and 3% were used. The Q values were the same as for experiment 3, giving a corpus of 15 sounds. The sounds were equalized in loudness, using the methodology described in Sec. II D of Fritz *et al.* (2007), to minimize the effect of loudness on perceptual differences.

Experiment 4b was similar, except that the damping now had five values, with the Q values being either normal or multiplied/divided by 2 and 4, and the number of values of vibrato amplitude was reduced to 3 (1%, 2%, and 3%) to make the duration of the test reasonable.

2. Procedure and participants

Participants first listened to the whole corpus of sounds and were then asked to rate the liveliness of each sound—presented in random order—on a scale from 0 to 10. Each sound was presented twice to check consistency. As described below, consistency was very good. Data were therefore averaged for each participant.

TABLE I. Average liveliness ratings for the five values of vibrato amplitude and the three values of damping for experiment 4a.

	Vibrato amplitude				
Q value	0%	0.5%	1%	2%	3%
$Q/2$	2.6	5.1	6.7	6.6	6.8
Q	2.2	4.6	6.8	6.0	5.1
$Q\times 2$	1.3	4.5	6.5	5.3	4.6

There were 17 participants for the first experiment and 11 for the second. Their hearing was checked to be normal.

3. Results

The results of experiment 4a were as follows. First, a reliability analysis was performed with SPSS (using a two-way mixed-effect model), which gave an intra-class correlation coefficient (ICC) of 0.9 with $p<0.001$ for average measures. The correlation being very good, we can consider the averaged liveliness ratings to be meaningful. They are given in Table I.

A within-subjects ANOVA was performed, with Q and vibrato amplitude as factors. The influence of Q was significant [$F(1.3,19.9)=4.3$, $\epsilon=0.7$, $p=0.02$], as was that of vibrato amplitude [$F(2.3,33.8)=34.7$, $\epsilon=0.6$, $p<0.001$]. The interaction was also significant [$F(5.4,80.9)=2.5$, $\epsilon=0.7$, $p=0.03$]. While the larger values of vibrato amplitude were rated the liveliest for $Q/2$, this was not the case for Q or $Q\times 2$, for which ratings were highest for the middle vibrato amplitude. This effect was most pronounced for $Q\times 2$. A significant quadratic trend was found for vibrato amplitude [$F(1,15)=44.5$, $p<0.001$]: liveliness increased when the amplitude increased from 0% to 1% and then stayed constant or decreased for higher values. This pattern of results can be explained by subjective reports of the participants that a large amount of vibrato makes the sound appear artificial and unpleasant, especially when combined with low damping. Thus, if liveliness is considered as a positive quality, such sounds may be rated as less lively.

There was a significant linear trend for the Q value [$F(1,15)=5.0$, $p=0.04$]: liveliness decreased when the Q value increased, although this effect was small for intermediate values of vibrato amplitude. This effect is the opposite of what would be expected from the claim of Mathews and Kohut (1973) referred to in the Introduction.

For experiment 4b, a reliability analysis was again performed. The ICC was equal to 0.5, with $p<0.02$. Since the ICC was much lower than for experiment 4a, we calculated Pearson's bivariate correlations to assess the extent to which the pattern of results was similar across participants and the extent to which each participant's results were consistent across repetitions. 5 of the 11 participants showed no significant correlations with ratings of the other subjects, nor within their own ratings across repetitions. These five participants were considered as unreliable. The ratings for the remaining six showed positive correlations with each other and within their own ratings.

When the results for the five unreliable participants were removed from the analysis, the ICC increased to 0.9, with

TABLE II. Average liveliness ratings for the three values of vibrato amplitude and the five values of damping for experiment 4b for the six participants whose results were consistent.

	Vibrato amplitude		
Q value	1%	2%	3%
$Q/4$	9.4	8.7	6.7
$Q/2$	8.7	6.3	5.5
Q	7.3	5.1	2.7
$Q\times 2$	6.3	3.0	1.8
$Q\times 4$	5.3	3.0	2.3

$p<0.001$. The mean liveliness ratings for the remaining six participants are given in Table II. The mean ratings decreased with increasing vibrato amplitude and with increasing Q value.

A within-subjects ANOVA was performed on the data for the six reliable participants, with factors vibrato amplitude and Q value. The effects of both vibrato amplitude $[F(2,10)=36.1,\ p<0.001]$ and Q $[F(4,20)=50.4,\ p<0.001]$ were significant. The surprising linear trend with damping that was found for experiment 4a was also found here $[F(1,5)=100,\ p<0.001]$, liveliness again decreasing as Q increased. The decrease in liveliness with increasing vibrato amplitude was confirmed by a significant linear trend $[F(1,5)=38.2,\ p=0.002]$, which is consistent with results of experiment 4a for vibrato amplitudes of 1% or more. However, it should be noted that some rating values in identical conditions differed markedly between experiments 4a and 4b. For example, for the "standard" (middle) amount of damping, and 3% vibrato amplitude, the mean rating was 5.1 for experiment 4a and 2.7 for experiment 4b. This may reflect individual differences across participants and/or an influence of the range of conditions presented on judgments for any specific condition.

The fact that 5 of 11 participants in experiment 4b gave unreliable results is an indication that the liveliness scale may not be entirely appropriate for characterizing perceptual differences among this particular set of sounds. The fact that the synthesized sounds were sometimes unrealistic (especially for large vibrato amplitudes) may have contributed to the unreliability of listener judgments. It was therefore decided to repeat part of the experiment, but using a recorded string signal instead of a synthesized sawtooth signal and using only a single representative value of vibrato amplitude.

B. Experiment 5: Liveliness judgments using recorded rather than synthesized string signal

1. Stimuli and procedure

A recording was made of a C sharp (277 Hz) played on the G string by an experienced violinist, instructed to use vibrato as in normal expressive playing. This was then taken as a typical signal. Vibrato amplitude was measured to be 2%. The Q values of the violin acoustical response were varied over the same five levels as in experiment 4b. The experiment was conducted in the same way as experiment 4b, except that participants were additionally asked to assess how much they liked each sound on a 0–10 scale. There were 12 participants, whose hearing was checked to be normal.

TABLE III. Average liveliness (* not reliable because of low ICC) and preference ratings for the five values of damping using a recorded string signal as input.

	Q value				
	$Q/4$	$Q/2$	Q	$Q\times 2$	$Q\times 4$
Liveliness*	5.5	5.2	5.2	4.9	4.5
Preference	2.8	3.5	5.3	6.5	5.3

2. Results

The reliability analyses gave an ICC of −0.3 for liveliness ($p=0.56$) and of 0.88 for preference ($p<0.001$). Therefore, averaging the liveliness ratings across participants would not be meaningful, as there was no consistent pattern of the results across participants. With the recorded string signal as input, but no variation in vibrato amplitude, liveliness seems even less appropriate as a quality for listeners to use for rating variations in damping, at least in the context of this kind of single-note test. Lack of anything resembling musical context may be important here, and this is explored later.

In contrast to the liveliness ratings, the ratings for preference proved to be similar across participants, and the averaged ratings are presented in Table III.

A one-way within-subjects ANOVA on the preference ratings showed a significant effect of the amount of damping $[F(1.9,21.3)=8.2,\ \epsilon=0.5,\ p=0.002]$, with a preference for somewhat higher Q than for the measured violin used as the baseline in this study (which is typical of conventional violins in this regard). The trend was cubic $[F(1,11)=11.4,\ p=0.006]$: too much damping $(Q/4)$ was definitely not liked but the extreme case with very low damping $(Q\times 4)$ was reported as making the sound hollow and metallic [echoing comments noted by Mathews and Kohut (1973)]. This trend makes more intuitive sense than the linear trend for liveliness obtained in experiments 4a and 4b.

The inconsistent use of the liveliness scale in this test prompted the question of whether liveliness is interpreted by participants as a quality of each individual sound, or whether listeners interpret it as a more global property of an instrument, in terms of how it responds to changes in, for example, the amount of vibrato applied. A first attempt to address this possibility led to another pair of experiments.

C. Experiments 6 and 7: Liveliness and preference for sounds grouped by "violin"

1. Experiment 6: Stimuli, procedure, and participants

The same synthesized stimuli as in experiment 4b were used (Sec. IV A), but this time they were not presented in an inter-mixed fashion but grouped by amount of damping, resulting in three sounds, differing in the amount of vibrato, for each of five violins. Participants were asked to listen to the three sounds for each violin, and then to assess the liveliness

TABLE IV. Average liveliness (* not reliable because of low ICC) and preference rating for the five "violins," each corresponding to a different amount of damping.

	Q value				
	$Q/4$	$Q/2$	Q	$Q\times2$	$Q\times4$
Liveliness*	4.4	4.3	4.5	5.1	4.6
Preference	2.3	3.0	5.9	6.3	4.4

of that violin and indicate how much they liked it. There were seven participants, whose hearing was checked to be normal.

2. Results for experiment 6

The degree of agreement among the participants was extremely low for the liveliness ratings (ICC=−6.1, $p=0.97$), whereas it was high (ICC=0.92, $p<0.001$) for the preferences. The mean ratings are shown in Table IV. Liveliness ratings did not show a clear dependence on damping. The results for preference are very similar to those of experiment 5. Again, there was a significant effect of damping on preference [$F(4,24)=12.8$, $p<0.001$], with a significant cubic trend [$F(1,6)=24.3$, $p=0.03$]: the highest mean rating was given for half of the damping ($Q\times2$) of the original violin.

3. Experiment 7: Stimuli, procedure, and participants

The results so far suggest that liveliness is not a suitable word for characterizing the perceptual effects of a change in the modal damping, at least within the constraints of the experiments reported here. Hence we designed an experiment which assessed preferences for vibrato amplitude and body damping without using any specific verbal description of the quality that was being judged.

It is also fair to say that none of the synthesized sounds used in the experiments so far really achieved satisfactory realism. Experiment 7 used a synthesized input signal, but it incorporated some further small adjustments to the synthesis details in a continuing effort to improve naturalness. The randomness in amplitude was increased to 0.6% and the percentage of the period used in the Gaussian filtering was increased to 4.5%. Furthermore, the vibrato amplitude was not constant throughout the duration of the note (Schoonderwaldt and Friberg, 2001): after a period of 100 ms where the vibrato amplitude was small and constant (called the delay), the vibrato amplitude was linearly increased over 400 ms (called the attack), then linearly increased with a lower slope or remained constant during the sustained part (400 ms), and finally decreased linearly to zero over 500 ms. The amounts of vibrato at the end of the delay, attack, and sustain phases are denoted a_d, a_a, and a_s, respectively.

The note was a C4 sharp (273 Hz), chosen because a measured string signal was available for this note. The spectral envelope of the synthesized input signal was adjusted from the very regular form of an ideal sawtooth to approximately match the rather more irregular spectral envelope of the recorded sound, using fast Fourier transform-based equalization.

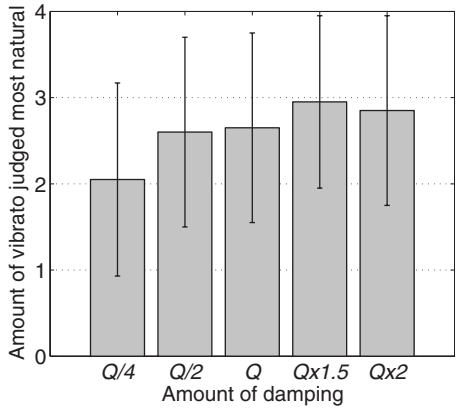


FIG. 4. Average number of the preferred amount of vibrato for the five "violins," each corresponding to a different amount of damping. The error bars represent ±1 standard deviation.

Four vibrato envelopes were used:

(1) $a_d=a_a=a_s=0.02\%$ (i.e., almost no vibrato),
(2) $a_d=0.2\%$, $a_a=a_s=0.5\%$,
(3) $a_d=0.2\%$, $a_a=a_s=1\%$, and
(4) $a_d=0.2\%$, $a_a=a_s=2\%$.

The numbers (1)–(4) are used as labels for the amount of vibrato. There were five violins, corresponding to the original Q values and Q values divided by 4 and 2, and multiplied by 1.5 and 2. These asymmetric ranges were motivated by the constraints imposed by real materials: damping could feasibly be increased by a factor of 4 by the violin maker or even by the player (by the way that they hold the violin against the chin and body, or with the left hand grip), but the damping cannot be reduced by the same factor, given the conventional wooden construction of the violin body. The larger factor was used in earlier experiments to create large effects, but for this test, in the interests of naturalness of the sounds, the range of Q values was restricted.

The participants were asked to choose, for each violin, the amount of vibrato which was the most natural for them. The five selected sounds were then re-presented in a second stage, and the participants were asked to rate how much they liked them on a 0–10 scale. The ten participants reported normal hearing.

4. Results for experiment 7

The degree of agreement among the participants for the first stage of the experiment was moderately good (ICC =0.63, $p=0.046$), so average results are shown in Fig. 4. A one-way within-subjects ANOVA showed no significant effect of the amount of damping [$F(1.2,11.2)=2.7$, $\epsilon=0.31$, $p=0.12$] indicating that, when the frequency response of the violin is relatively flat (low Q), naturalness is not increased by the use of more vibrato. This is consistent with the conclusion drawn from the results of experiment 3, that damping and amount of vibrato are largely independent in their perceptual effects.

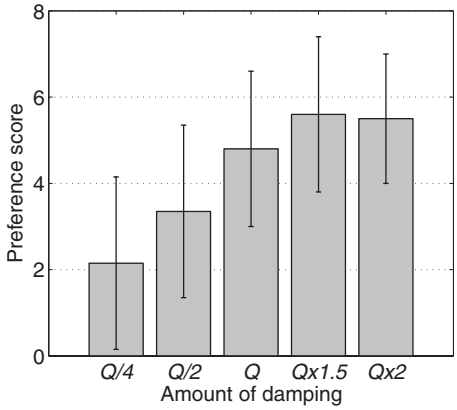


FIG. 5. Average preference score (on a scale from 0 to 10) for each violin (each corresponding to a different amount of damping) obtained using the preferred amount of vibrato. The error bars represent ±1 standard deviation.

There is evidence that the discrepancy between judgments could represent a difference in strategy. In particular, when high Q values are considered, half of the participants judged $a_{vib}=0.5\%$ to be most natural and half judged $a_{vib}=2\%$ as most natural. But very few judged $a_{vib}=0\%$ or $a_{vib}=1\%$ as most natural.

For the second stage of the experiment, which involved judgments of preference, the agreement of results across participants was much higher (ICC=0.88, $p<0.001$). The average preference score for each amount of damping is shown in Fig. 5. A one-way within-subjects ANOVA showed a significant effect of the amount of damping [$F(4,32)=7.7$, $p<0.001$] with a significant linear trend [$F(1,8)=11.3$, $p=0.01$], showing that the participants preferred higher Q values.

This preference for somewhat higher Q values than for the original response is consistent with the results of experiments 5 and 6. However, it seems not to be consistent with the results of Mathews and Kohut (1973). The discrepancy may derive from the fact that their study involved more extensive playing rather than single-note comparisons. Perhaps the reason for their participants not liking a peak-to-valley ratio larger than 10 dB was partly associated with unevenness of the violin notes across the scale, some being very loud and some being very soft. This was not a factor in the experiments described so far, as only a single note was used in each experiment. A further experiment was therefore carried out that involved real performance.

D. Experiment 8: Real-time playing on an electric violin

Violinists may need to play a violin to judge it reliably. Tests were therefore conducted with an electric violin, which allowed players to assess each virtual violin under fairly natural conditions.

1. Methodology

Instead of doing the filtering off-line using pre-synthesized/measured string input signals, the filtering was done in real time, using the output signal from a Harley Benton electric violin. The filtering was realized with Signal Wizard hardware developed by Gaydecki (2009). The filtered sound was played through a stereo pair of high-quality loudspeakers (ATC SCM100ASL) in a recording studio. The player was facing the loudspeakers and 2 m away from both. The sound level was adjusted for each subject in order to provide a comfortable level, while being sufficient to mask the direct sound from the electric violin. The experiment took place in a recording studio.

This methodology allows a player to test different violin sounds using the same physical violin. Also, it allows violinists to play a wide variety of effects, which increases the validity of the results with respect to violin performance, in contrast to the passive listening tests reported above.

TABLE V. Average liveliness and preference ratings for the six virtual violins played through the electric violin, each corresponding to a different amount of damping.

	Q value					
	$Q/4$	$Q/2$	Q	$Q\times1.5$	$Q\times2$	$Q\times4$
Liveliness	4.6	5.4	5.5	5.9	6.4	6.8
Preference	5.4	6.1	5.8	6.0	4.4	2.5

2. Test procedure and participants

There were six synthetic violins with a range of Q values; the original, divided by 4 and 2, multiplied by 2, and either by 1.5 or by 4. This experiment was, in fact, carried out simultaneously with the listening experiments, so the first six subjects were tested using $Q\times4$. Later, $Q\times4$ was dropped and the remaining nine subjects were tested using $Q\times1.5$ instead. The 15 violinists reported normal hearing. Two of them were also violin makers of a high standard. The test was divided into two stages. In the first stage, violinists were instructed to play whatever they wanted, but without vibrato. In the second stage, they were encouraged to play with vibrato. In each stage, they were presented with the six violins in order of ascending Q value. After a first experience with each violin, to give an idea of the range of variation, they had to give a score to each violin (on a 0–10 scale) for both their preference (i.e., how much they liked it) and liveliness/responsiveness (i.e., how the violin responded to what they wanted to get from it).

3. Results

The agreement between participants was very good. The ICC was 0.76 ($p=0.01$) for preference and 0.91 ($p<0.001$) for liveliness/responsiveness. A within-subjects ANOVA showed no significant difference between ratings for the first stage (playing without vibrato) and the second stage (playing with vibrato) for either preference and liveliness. This can perhaps be explained by the comments of several players: vibrato is used to make the sound musically more interesting, but it does not change the quality of the violin. Average results for the two stages are presented in Table V.

While the liveliness increased significantly and progressively with decreasing damping, preference showed a broad peak around the original damping. It seems that low damping

increases liveliness, but is not preferred overall. This is a different pattern than was seen in Tables III and IV from the single-note tests, where slightly lower than normal damping was preferred. This may mean that there can be "too much liveliness," or more likely that judgments of pleasantness are influenced by perceptual factors such as unevenness, or increased awareness of the body impulse response during the initial transient, especially with vigorous bowing techniques such as martelé (Woodhouse, 1983).

V. DISCUSSION AND CONCLUSIONS

The literature of violin acoustics has to date concentrated predominantly on physics, but most of the key questions have a perceptual dimension. We have attempted, through a series of tests, to shed light on one aspect of why a violinist or listener prefers one instrument over another. There are several threads running through the work: specifically, the perceptual effect of violin vibrato, the perceptual effect of changing body damping, and the interaction (or lack of it) between the two. More generally, we raise the issue of what is needed to make a synthesized sound "natural," and the much bigger issue of how to design controlled empirical tests to access high-level musical perceptions and judgments using terms that are recognizable and meaningful to performers and instrument makers.

Any study which is to be relevant to the concerns of players and makers must take notice of their informal evidence, beliefs, and experience. This study took as its starting point two observations. First, western classical violinists habitually use vibrato, and the interaction of vibrato with the peaky frequency response of the violin undoubtedly produces clear perceptual effects (Gough, 2005; Curtin and Rossing, in press). Second, among the verbal descriptions very commonly used by violinists to describe differences between instruments is a desirable quality of liveliness or responsiveness, two terms which appear to be regarded as closely related (Fritz *et al.*, 2008). It seemed a promising hypothesis that these two things were connected, i.e., that liveliness would result from the combination of the use of vibrato and a peaky violin response, but any direct link between them has proved remarkably elusive to demonstrate.

The overall picture revealed by the tests reported in this paper is complicated. In order to probe the perceptual correlates of changes in vibrato amplitude and of damping of the violin body resonances, several experiments were conducted. The simplest of these measured just-noticeable differences in vibrato amplitude and in violin body damping. Even these first results gave surprises; experiment 1 showed that the threshold for detection of a change in vibrato amplitude was unaffected by the amount of damping, and, conversely, experiment 2 showed that the threshold for detection of a change in damping was unaffected by the presence or amount of vibrato.

Experiment 3 investigated the combined perceptual effects of vibrato amplitude and body damping based on judgments of similarity and dissimilarity. The results showed that, to a large extent, the two parameters are perceived as independent dimensions. However, there was some systematic interaction which seems intuitively plausible: low damping with small vibrato amplitude was judged somewhat similar to high damping with high vibrato amplitude when compared to the opposite pair of extreme combinations.

A subsequent series of listening tests examined how vibrato and body damping affect the perceived liveliness of, and general preference for, the sound. For those tests in which the amplitude of vibrato needed to be varied, the only way to achieve appropriate input without additional uncontrolled factors was by synthesis. However, all synthesized sounds achieved so far have suffered from a lack of naturalness, despite successive attempts to improve this aspect. This artificial quality may have compromised the ability of listeners to make judgments reflecting real musical contexts. Recorded signals taken from an actual performance are much closer to a musical context, but offer less flexibility in test design. Live playing using an electric violin and real-time filter system is better again, but is only suitable for certain kinds of test and also brings a host of new factors into play.

In summary, the results show that, when listening to single notes, participants find it difficult to make judgments of liveliness, and in most tests, the word was not used in a consistent way across participants. This was especially true in the later tests when participants were also asked to judge overall preference for the sounds. Judgments of preference were more consistent than judgments of liveliness. Only for the tests with live playing on the electric violin were participants able to judge both liveliness and preference consistently. The results of this experiment showed a trend which the authors had been expecting at the outset: lower damping produces progressively greater judged liveliness. However, no direct link with vibrato was found. Also, preference was broadly centered around the damping values for the original, reference violin. In the earlier tests with single synthesized notes, the liveliness ratings which were sufficiently consistent produced unexpected patterns. Sounds with no vibrato were judged less lively than sounds with some vibrato, but as vibrato amplitude increased, the liveliness ratings flattened off or even decreased. More surprisingly, liveliness was judged to decrease, not to increase, as the damping was reduced. This all seems to suggest that liveliness is a quality more relevant to a player than to a non-playing listener, or at least that it is given a different interpretation in the two contexts.

Despite the fact that vibrato clearly influences the perception of musical timbre, we have been unable to confirm a clear relationship between the timbral properties of the instrument itself and those produced by use of vibrato. It has previously been suggested that vibrato might either accentuate or compensate for resonance characteristics of stringed instruments. We did find evidence that varying the resonant behavior of an instrument can sometimes improve its musical properties. But the property of liveliness or responsiveness, often used to describe the capacity of an instrument for expressive sound production, appears to be largely independent of the specific expressive technique of vibrato. The popularly described relationship between vibrato and violin resonance may arise indirectly from the fact that both factors influence the perceived timbre in any given musical context.

A consistent trend from the tests involving preference judgments of single notes was a slight preference for damping somewhat lower than for the reference case (i.e., slightly higher Q values). However, this trend was not found in the final experiment with the electric violin. In that experiment, the preferences were roughly symmetrical around the reference case. These different results are not necessarily in contradiction. They probably indicate a genuine tension, whereby some aspects of violin sound (such as timbre of an individual prominent note) benefit from lower damping, while other aspects (such as evenness) benefit from higher damping. This may echo a debate in the violin-making world concerning the selection of wood for instruments. Luthiers frequently express a preference for wood with a "good ring," suggesting low damping, but there is also a persistent belief that some instruments improve with age. In some cases the wood in old instruments has degenerated to a chalky texture, suggesting a high density of micro-cracks and consequently high damping.

What makes for a perception of "naturalness" in a synthesized violin sound? Even for single short notes with vibrato, it has proved surprisingly hard to produce sounds which do not evince an immediate sensation of artificiality. One would guess that the explanation has something to do with irregularities of various kinds on different timescales: cycle by cycle, correlated variations over several cycles, and variations at timescales relevant to the player's moment-to-moment input. There are few published studies on this question: the work of Schumacher (1992) is a notable exception. This question of realistic synthesis is a worthy subject of study in its own right. There is some knowledge and experience within the world of synthesis for the purposes of musical performance, but it is important to note a philosophical difference. If the aim is simply to make a good sound, there are no rules about what can and cannot be tried. But if the aim is to understand key features of actual played musical notes on conventional acoustic instruments, the ingredients of the synthesis should all have some basis in physics or in the physiological limitations on, for example, bow control and vibrato production.

Finally, there are issues of general methodological significance. In this paper the authors have attempted to design tests which have internal empirical and experimental validity but which also retain at least some external validity through relevance to musical practice. This has proved challenging, and some of the results seem contradictory, but they should not be dismissed on those grounds. The experience of observing skilled musicians testing instruments suggests very strongly that violinists can make consistent judgments about subtle qualities of instruments. The fact that simple tests do not readily yield the knowledge of how they do it is a spur to further work, not an indication of a dead end. High-level musical perceptions are presumably of a comparable order of complexity to, for example, face or voice recognition, and they may take a comparable scientific effort to begin to understand.

In conclusion, while there are relationships between perceived timbre, Q values and vibrato, the primary function of vibrato when applied in real performance does not appear to be that of enhancing the perception of the properties of a violin body (damping). In conjunction with the fact that the use of vibrato, even on the violin, appears to have varied significantly over the last 400 years of western musical history (see, e.g., Moens-Haenen, 2009), this finding could be taken to suggest that the principal purpose of the use of vibrato in contemporary musical performance is to enhance expressivity.

ACKNOWLEDGMENTS

The authors are grateful to Professor P. Gaydecki for making the Signal Wizard system available for this study and to the Leverhulme Trust for having funded this research. They thank an anonymous reviewer for helpful comments.

Brown, C. (**1988**). "Bowing styles, vibrato and portamento in nineteenth-century violin playing," J. Roy. Music. Assn. **113**, 97–128.
Cheng, F. P.-H. (**2006**). "Physics and perception of stringed instrument vibrato," MPhil. thesis, University of Cambridge, Cambridge, UK.
Cremer, L. (**1985**). *The Physics of the Violin* (MIT Press, Boston, MA).
Curtin, J., and Rossing, T. (**in press**). "Violins," in *Science of String Instruments*, edited by T. Rossing (Springer, New York).
Darwin, C. J., Turvey, M., and Crowder, R. G. (**1972**). "An auditory analogue of the Sperling partial report procedure: Evidence for brief auditory storage," Cogn. Psychol. **3**, 255–267.
Ewins, D. (**2000**). *Modal Testing: Theory, Practice and Application* (Research Studies Press Ltd., Baldock, England).
Fletcher, H., and Sanders, L. (**1967**). "Quality of violin vibrato tones," J. Acoust. Soc. Am. **41**, 1534–1544.
Fritz, C., Blackwell, A. F., Cross, I., Moore, B. C. J., and Woodhouse, J. (**2008**). "Investigating English violin timbre descriptors," in Proceedings of the 10th International Conference on Music Perception and Cognition, Sapporo, Japan.
Fritz, C., Cross, I., Woodhouse, J., and Moore, B. C. J. (**2007**). "Perceptual thresholds for detecting modifications applied to the acoustical properties of a violin," J. Acoust. Soc. Am. **122**, 3640–3650.
Gaydecki, P. (**2009**). "Signal wizard 2," http://www.signalwizardsystems.com/ (Last viewed July 1, 2009).
Gorrill, S. (**1975**). "A viola with electronically synthesised resonances," Catgut Acoust. Soc. Newsletter **24**, 11–13.
Gough, G. E. (**2005**). "Measurement, modelling and synthesis of violin vibrato sounds," Acust. Acta Acust. **91**, 229–240.
Mathews, M. V., and Kohut, J. (**1973**). "Electronic simulation of violin resonances," J. Acoust. Soc. Am. **53**, 1620–1626.
McAdams, S. (**1989**). "Segregation of concurrent sounds. I: Effects of frequency modulation coherence," J. Acoust. Soc. Am. **86**, 2148–2159.
McAdams, S., and Giordano, B. L. (**2009**). "The perception of musical timbre," in *Oxford Handbook of Music Psychology*, edited by S. Hallam, I. Cross, and M. Thaut (Oxford University Press, Oxford), pp. 72–80.
McIntyre, M. E., Schumacher, R. T., and Woodhouse, J. (**1981**). "Aperiodicity in bowed-string motion," Acustica **49**, 13–32.
McIntyre, M. E., and Woodhouse, J. (**1974**). "Towards a psychoacoustically realistic violin physics," Catgut Acoust. Soc. Newsletter **22**, 18–19.
Mellody, M., and Wakefield, G. (**2000**). "The time-frequency characteristic of violin vibrato: Modal distribution analysis and synthesis," J. Acoust. Soc. Am. **107**, 598–611.
Meyer, J. (**1992**). "Zur klanglichen Wirkung des Streicher-Vibratos (On the tonal effect of string vibrato)," Acustica **76**, 283–291.
Moelants, D. (**2004**). "The timing of tremolo, trills and vibrato by string instrument players," in *Proceedings of the 8th International Conference on Music Perception and Cognition*, edited by S. Lipscomb, R. Ashley, and P. Webster (Northwestern University, Evanston, IL).
Moens-Haenen, G. (**2009**). "Vibrato," in Grove Music Online, Oxford music online, http://www.oxfordmusiconline.com/subscriber/article/grove/music/29287 (Last viewed July 1, 2009).
Neff, D. L., and Green, D. M. (**1987**). "Masking produced by spectral uncertainty with multi-component maskers," Percept. Psychophys. **41**, 409–415.
Prame, E. (**1994**). "Measurements of the vibrato rate of ten singers," J. Acoust. Soc. Am. **96**, 1979–1984.

Prame, E. (**1997**). "Vibrato extent and intonation in professional western lyric singing," J. Acoust. Soc. Am. **102**, 616–621.
Schoonderwaldt, E., and Friberg, A. (**2001**). "Towards a rule-based model for violin vibrato," in Proceedings of the Workshop on Current Research Directions in Computer Music, Barcelona, Spain.
Schumacher, R. (**1992**). "Analysis of aperiodicities in nearly periodic waveforms," J. Acoust. Soc. Am. **91**, 438–451.
van Besouw, R. M., Breteton, J. S., and Howard, D. M. (**1996**). "Range of tuning for tones with and without vibrato," Music Percept. **26**, 145–155.
Wickelmaier, F., and Ellermeier, W. (**2007**). "Deriving auditory features from triadic comparisons," Percept. Psychophys. **69**, 287–297.
Woodhouse, J. (**1983**). "On recognising violins: Starting transients and the precedence effect," Catgut Acoust. Soc. Newsletter **39**, 22–24.
Yoo, L., Sullivan, D., Moore, S., and Fujinaga, I. (**1998**). "The effect of vibrato on the response time in determining the pitch relationship of violin tones," in Proceedings of the 5th International Conference on Music Perception and Cognition, edited by S. W. Yi, Seoul, Korea.

Normal modes of a musical drumhead under non-uniform tension

Randy Worland[a)]
Department of Physics, University of Puget Sound, 1500 North Warner, Tacoma, Washington 98416-1031

(Received 24 July 2009; revised 3 October 2009; accepted 3 November 2009)

The effect of non-uniform tension on the normal modes of musical drumheads is examined. Using the standard (m,n) designation for the number of nodal diameters and circles, ideal circular membrane modes with $m>0$ are doubly degenerate. These degeneracies can be lifted by perturbations to the circular symmetry. Of practical interest to drummers are perturbations caused by non-uniform tension applied at the rim of the drum, leading in some cases to audible frequency splitting. The role of the (1,1) mode in practical drum tuning is analyzed using data obtained using time-averaged electronic speckle-pattern interferometry along with finite element analysis. The resulting model is then generalized to include all modes, using symmetry arguments along with a selection rule taken from group theory. The model compares favorably with both perturbation theory and finite element analysis, and is consistent with experimental observations.
 [DOI: 10.1121/1.3268605]

PACS number(s): 43.75.Hi, 43.40.Dx [NHF] Pages: 525–533

I. INTRODUCTION

While many descriptions of drumhead physics assume uniform tension, drummers know that this is at best an approximation. Tuning real drums is a challenging process that starts with the need to tune a single head to itself (what timpanists refer to as "clearing the head") by adjusting the applied tension. In Sec. II we introduce the standard solution for an ideal drumhead under uniform tension and establish the notation. Section III discusses the manner in which drummers tune a single head, and relates this procedure to the physics of the normal modes on a circular membrane. In particular, the significant role of the lowest degenerate mode (1,1) is developed using both experimental and numerical techniques. Drums typically found in the modern drum set (e.g., snare drums and tom toms) are the focus of the experimental presentation, although single-headed frame drums are also used due to their relatively simple design.

In Sec. IV the practical tuning results are generalized to include the qualitative behavior of all degenerate modes in response to non-uniform tension. The arbitrary tension perturbation is decomposed into a sum of symmetric perturbations. The effects of the individual perturbations are then studied using a result taken from group theory. Results from finite element analysis and perturbation theory compare favorably with the model's predictions regarding frequency splitting. Changes in the mode shapes are also considered briefly.

Although several published discussions[1–3] make reference to the goal of obtaining uniform tension in drumhead experiments, there appears to be little in the way of systematic study of non-uniform drumhead tension and its effect on mode splitting in the musical acoustics literature, either from a practical drum tuning standpoint, or from a more general perspective. Bassett[4] mentioned the beating of doublet modes on a bass drum, due to unevenness in the tension. A systematic approach to the non-uniform tension problem is provided by Mei,[5] who used finite element analysis to calculate frequencies and mode shapes of a circular membrane subject to non-uniform tension, but the issue of frequency splitting is not addressed. Rhaouti *et al.*[6] incorporated non-uniform tension into their finite element timpani model to obtain agreement with experimental frequency splitting results, but only apply the method to a particular case.

The present work extends these ideas by developing a general qualitative model of frequency splitting on a drumhead under non-uniform tension, with an emphasis on effects that are audible during the process of tuning a drum.

II. IDEAL DRUMHEAD

The ideal circular drumhead is assumed to be of uniform mass density σ and uniform tension T, fixed at the perimeter, without air loading or membrane stiffness. The analytical solution for the spatial part of the normal modes yields[7]

$$\psi_{mn}(r,\theta) \sim J_m(k_{mn}r)\begin{Bmatrix}\cos[m(\theta+\phi_{mn})]\\ \sin[m(\theta+\phi_{mn})]\end{Bmatrix}, \quad (1)$$

where $\psi(r,\theta)$ represents the normal displacement of the membrane as a function of the radial and angular coordinates r and θ. The modes are designated by the integers m and n, which represent the number of nodal diameters and concentric circles (including the perimeter) in the standing wave pattern. For $m=0$, only the cosine solution of Eq. (1) is used. The J_m are Bessel functions of the first kind. The k_{mn} are related to the angular frequencies ω_{mn} by $k_{mn}=\omega_{mn}\sqrt{\sigma/T}$, subject to the constraint that the nth circular node of $J_m(k_{mn}R)$ equals zero at the perimeter $(r=R)$ of the membrane. The angles ϕ_{mn} define the orientations of the modes containing nodal diameters. These phase angles can be fixed

a)Electronic mail: worland@pugetsound.edu

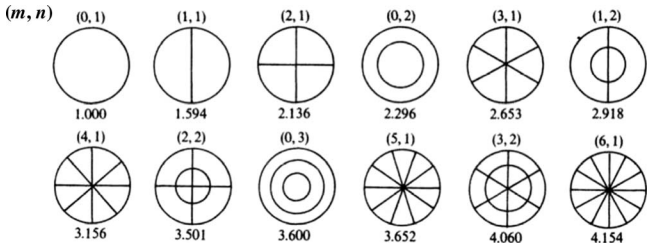


FIG. 1. The first 12 modes of an ideal membrane. The (m,n) designation corresponds to the number of nodal diameters and circles, respectively. The frequency relative to the fundamental (0,1) is shown below each mode.

either by a perturbation to the system or by the initial conditions.

Figure 1 shows the first 12 normal mode shapes, along with their (m,n) designations.[8] Any mode containing at least one nodal diameter $(m>0)$ is doubly degenerate (see Ref. 7, p. 212), as for each set of indices m and n the sine and cosine solutions of Eq. (1) represent two orthogonal modes having the same frequency, ω_{mn}. Thus, each of the degenerate modes of Fig. 1 has a partner (not shown) with the same frequency and shape, rotated by an angle $\pi/2m$ (i.e., with nodes and antinodes interchanged).

The degeneracy of these modes may be lifted by a variety of perturbations, including non-uniformities in the applied tension. The response of these degenerate modes to tension perturbations can result in audible frequency splitting that is of practical concern to the drummer.

III. DRUM TUNING AND THE (1,1) MODE

Each head of a drum such as a tom tom or snare drum consists of a membrane (typically Mylar) that is stretched over the bearing edge of a cylindrical drum shell. The head is held taut by a counterhoop fastened to threaded tension rods that are screwed into symmetrically placed tuning lugs. Tightening the tension rods causes the counterhoop to be pulled down, stretching the membrane and generally raising the pitch of the drum. Although tom toms and snare drums contain two drumheads, only the tuning of a single head to itself is considered here, without regard to the overall pitch of the drum.

Tuning a drumhead "to itself" involves tapping the head near each tuning lug and adjusting the tension rods with a drum key in an attempt to make each tap sound the same. In this manner the drummer strives to minimize the perceived frequency differences and beats that are heard at various perimeter points when the tension is not uniform. The process is iterative, as adjustment at a particular lug changes the tension, and thus the perceived sound, at all points on the head.

Drummers often place the drum on a flat surface and/or mute the drum lightly at the center while tuning. Placing a finger at the center prevents motion of the head at the antinode of the fundamental mode (0,1). Placing the drum on a flat surface also inhibits the (0,1) motion, as for this mode the entire surface moves in phase, necessitating a maximum of air motion on either side. The flat surface also dampens the opposite head, such that the head being tuned is heard in isolation.

The model presented below assumes a circularly symmetric system. The tuning process may, in fact, benefit from non-uniform tension in cases where an existing asymmetry of the membrane is being counteracted, but such cases are not considered here.

A. Simple model of (1,1) mode splitting

The fact that drummers tap near the edge while muting the center of the head provides some insight into which modes are being heard during the tuning process. The fundamental (0,1) is generally quite prominent in the audio spectrum of drums such as tom toms and snare drums, particularly when struck near the center of the head. However, this mode is not degenerate and thus does not split into two frequencies in response to a perturbation. Tightening or loosening a given tension rod will raise or lower the fundamental frequency, but listening to this frequency alone does not provide the drummer with any information regarding the uniformity of the tension.

Tapping near the perimeter of the drum, as opposed to the center, preferentially excites the asymmetric $(m\neq 0)$ modes. The (1,1) mode, in particular, provides a good starting point in the analysis of drum tuning, as it is the lowest frequency degenerate mode and typically the most prominent of the modes that can be split. Under ideal circumstances, the (1,1) mode produces a single frequency and occurs with its nodal diameter oriented in any direction [a linear combination of the two orthogonal states of Eq. (1)]. In practice the location where the drum is struck determines the orientation, as an antinode is forced near that region. However, when the circular symmetry is broken by a tension perturbation that creates orthogonal fast and slow axes on the membrane, the (1,1) mode splits into two modes with different frequencies and orthogonal orientations determined by the perturbation. This frequency splitting must be minimized in order for the drumhead to sound in tune with itself.

The tuning method used in practice suggests the model of (1,1) mode behavior shown in Fig. 2. An increase in tension across a single diameter of the drum represents the simplest tension perturbation. In Fig. 2(a) the increased tension is applied at the top and bottom, causing the wave speed to be greater in the vertical direction than in the horizontal direction. The orthogonal (1,1) modes will align relative to this perturbation, with the vertical nodal line corresponding to the lower frequency f_-. We anticipate the lower frequency with this alignment due to the lower speed for waves traveling across this nodal line. From an energy perspective, this orientation places the antinodes in the region of relatively lower tension, again consistent with a lower frequency. The mode with the horizontal nodal line corresponds to the higher frequency f_+ (higher wave speed across this nodal line, and antinodes in areas of higher tension). This model is in agreement with Rayleigh's principle,[9] which states that in re-

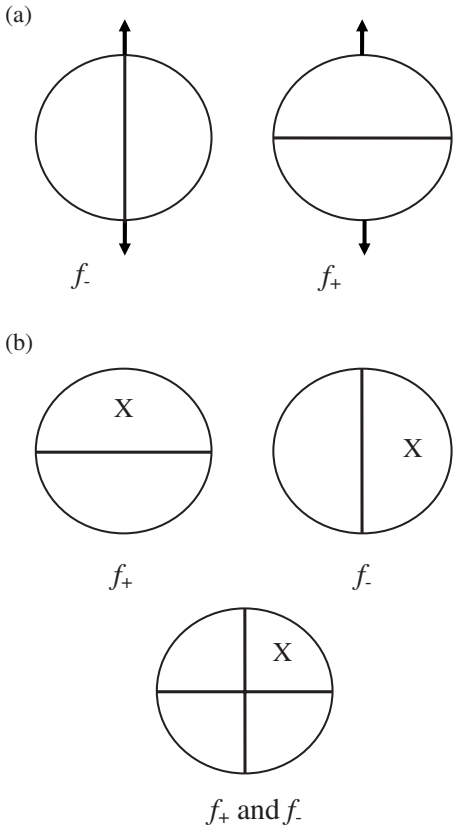


FIG. 2. Model of the (1,1) mode's response to the simplest tension perturbation. (a) A slightly non-uniform tension, with higher tension from top to bottom (indicated by arrows), has lifted the degeneracy of the (1,1) mode, producing two modes of slightly different frequencies: f_- and f_+. The orthogonal nodal lines are vertical and horizontal in response to the orientation of the perturbation. (b) The three images show how one may excite either or both of these (1,1) modes depending on where the drum is struck (location indicated by an "X"), resulting in pitch differences (f_- and f_+) and beats.

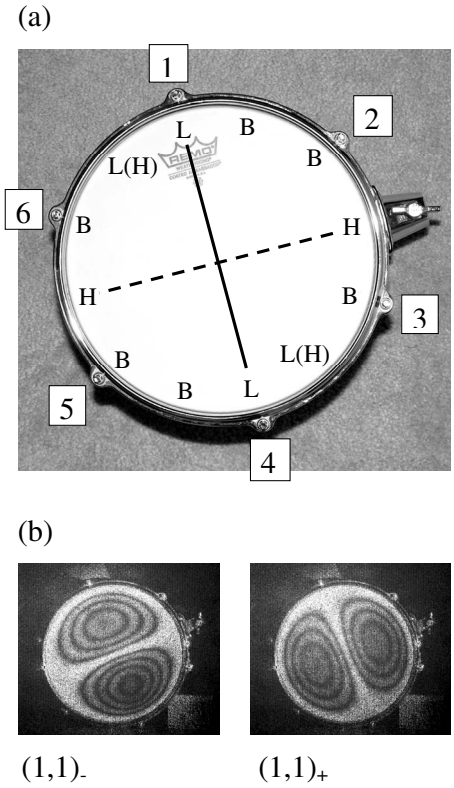


FIG. 3. (a) A 10 in. Yamaha tom tom with six numbered lugs. The head is a single-ply Remo Coated Ambassador. The low and high (1,1) mode frequencies detected with the spectrum analyzer at each tapping location are indicated by L (f_-=226 Hz) and H (f_+=237 Hz). Symbols in parentheses indicate smaller intensities. Locations producing comparable intensities for both frequencies are indicated by B. Dotted line shows where the nodal diameter for the $(1,1)_-$ mode is expected. Solid line indicates where the nodal diameter for $(1,1)_+$ is expected. (b) Electronic speckle-pattern images of the same drum, driven acoustically at f_-=226 Hz and f_+=237 Hz. Nodal regions appear white in these images. Darker lines indicate contours of equal amplitude motion. The nodal lines are oriented approximately as predicted from the data shown in part (a).

sponse to a perturbation the nodal diameters of the mode pair will align so as to produce the maximum and minimum possible frequencies.

Figure 2(b) shows that striking this drumhead at different locations will selectively excite the $(1,1)_+$ and $(1,1)_-$ modes. For example, striking near the top causes the higher frequency f_+ to be heard, while striking to one side causes the lower frequency f_- to be heard. Striking at regions between the two nodal lines will excite both (1,1) modes. Depending on the frequency difference, this may lead to the production of audible beats.

B. Experimental

1. Measurements of the (1,1) mode

To test this model, a drumstick was used to tap near each lug of a drum. A microphone placed about 10 cm above the head detected the resulting acoustic signal near each lug location, which was input into a spectrum analyzer. With a moderately well tuned drum the frequencies f_- and f_+ of the (1,1) mode pair could be identified on the spectrum analyzer.

The (1,1) identification and the alignment of the two modes were verified using an electronic speckle-pattern interferometry (ESPI) system based on the design of Moore[10,11] to image the modes. This optical system produces a time-averaged image displaying contours of equal amplitude motion when the drum is driven acoustically by a speaker connected to a function generator. The sinusoidal driving frequency is varied until a stable image of the drumhead is observed, with nodal regions appearing white.

Results taken with a six-lug tom tom are shown in Fig. 3(a). Frequency data were taken near each of the six lugs and halfway between each lug as well. At each location the f_- and f_+ frequencies were detected, in varying strengths. Largest intensities for the lower frequency (L) are seen across from each other along roughly the lug 1–4 diameter. Largest intensities for the higher frequency (H) are seen near either end of the lug $2\frac{1}{2}$–$5\frac{1}{2}$ diameter. The low and high frequencies both occur (B) with comparable intensities along perimeter points midway between these axes. Higher and lower pitches, as well as beats, were heard at the various tapping locations as expected.

The ESPI images in Fig. 3(b) show that the orientations of the two (1,1) modes are consistent with the predictions of the model. The nodal line of the higher frequency mode is oriented approximately in the 1–4 direction; the lower frequency mode shows the $2\frac{1}{2}$–$5\frac{1}{2}$ orientation.

Tension measurements were also taken at the perimeter

locations using a mechanical device called a "DrumDial."[12] The DrumDial is a commercial product for percussionists that indicates tension based on deflection of a spring-loaded piston when placed on the drumhead. Although the numerical scale is uncalibrated, higher and lower readings can give some indication of tension variations around the perimeter. These measurements were consistent with the interpretation of higher membrane tension along the $2\frac{1}{2}$–$5\frac{1}{2}$ axis and lower tension along the 1–4 axis (i.e., higher readings near regions of higher frequency measurements).

In principle the drumhead can be tuned by adjusting the tension rods until all of the tap locations yield a common frequency for the (1,1) mode (although this does not guarantee that higher modes will be beatless, as discussed in Sec. IV). However, correcting the tuning is not simply a matter of tightening the lugs where the pitch is low. Due to the two-dimensional nature of the membrane, tightening one lug will cause an increase in tension at points other than just the region directly across the diameter. In addition, contact with the membrane is not isolated at the lugs; it is through the counterhoop, which is a continuous ring of metal or wood. As a result, in smaller diameter drums with relatively closely spaced lugs, a very stiff hoop may lead to cases where a single tension rod is completely loosened without appreciable change in tension at that location. A further complication is the high degree of friction between the membrane and the bearing edge of the shell, which prevents the tuning process from being reproducible with a particular sequence of tension rod adjustments.

2. Sequential detuning of a drumhead

To understand how the (1,1) modes change as the drum is tuned, a sequential process for varying the tension was employed. In the laboratory, it proved simplest to start with a tuned drum and then tighten or loosen a single lug in small steps to see gradual changes in the mode shapes and frequencies.

Drums were initially tuned by ear, as described above. In addition, the spectrum analyzer was used to confirm that the tuning procedure did tend to minimize the frequency splitting of the lower degenerate modes, particularly the (1,1) mode. Mode shapes, as seen with ESPI imaging, also appeared generally more symmetric as the tuning was improved. "Perfect" tuning, which presumably would be characterized by fully symmetric mode shapes and the absence of frequency splitting, was never achieved.

Figure 4 shows the first few modes of a representative drum tuned in this manner. Note the complementary orientations (rotated by $\pi/2m$ in each case) of the (1,1), (2,1), and (3,1) mode pairs, and the relatively small frequency splittings of about 2–6 Hz ($\Delta f/f \sim 1\%$). Small variations in the mode shapes from the ideal symmetric shapes of Fig. 1 are also seen.

A set of sequential detuning results is shown in Fig. 5. Figure 5(a) shows the (1,1) modes of a tom tom, tuned such that $\Delta f = 2$ Hz. From this initial state, ESPI observations were made as one lug (see arrow) was loosened in small incremental turns of the drum key. Figure 5(b) shows the first observed change—the modes have rotated slightly such that

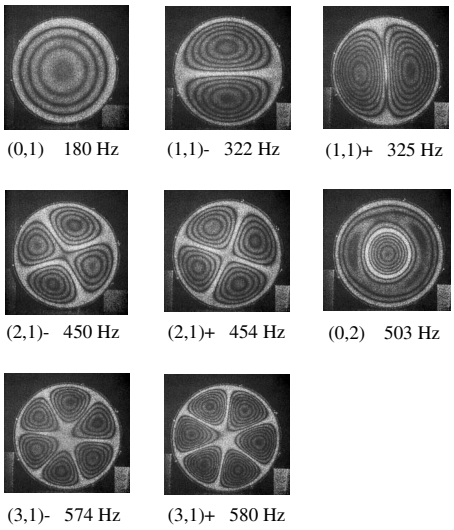


FIG. 4. Electronic speckle-pattern images showing the first several modes of a "tuned" drum (single-headed 12 in. Remo frame drum with eight lugs and a single-ply head). The (1,1), (2,1), and (3,1) modes appear in pairs with complementary orientations and small frequency differences. Slight asymmetries can be seen, particularly in the (0,2) mode.

one nodal line aligns with the perturbation while the other rotates to an angle perpendicular to the perturbation. As the arrow in Fig. 5(b) indicates a *lower* tension, it is the nodal

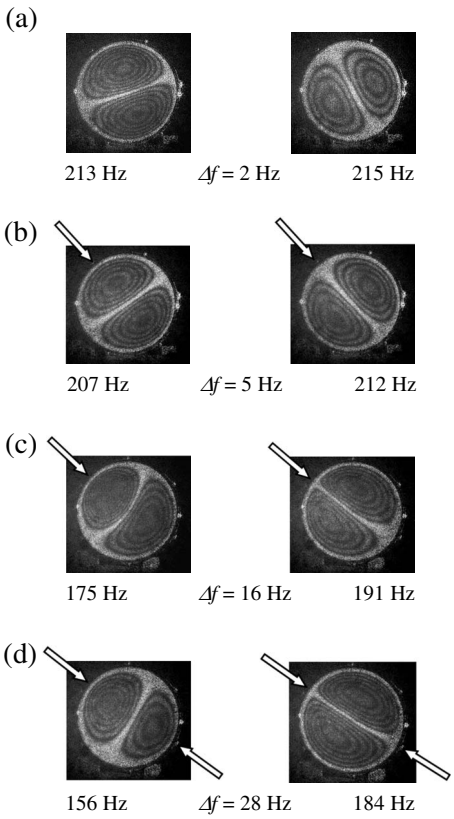


FIG. 5. Electronic speckle-pattern images showing the (1,1) modes of a Yamaha 12 in. tom tom with six tuning lugs. The drum head is a single-ply Remo Coated Ambassador. Arrows indicate tuning lugs that were loosened. Note the alignment, curvature, and frequency splitting of the modes. (a) Approximately tuned. (b) The indicated lug (see arrow) has been loosened $\frac{1}{4}$ turn. (c) The same lug has now been loosened $1\frac{1}{4}$ turns. (d) The opposite lug has now been loosened $\frac{1}{2}$ turn.

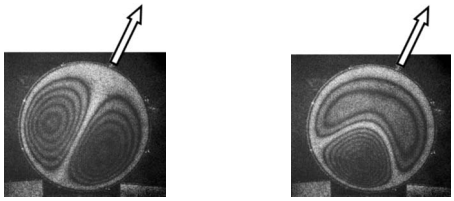


FIG. 6. Electronic speckle-pattern images showing the (1,1) modes of a Remo 10 in. frame drum with eight lugs and a single-ply head. Arrows indicate the tuning lug that was tightened approximately two full turns. Note the high degree of curvature of the higher frequency mode and the fairly modest amount of frequency splitting.

line of the *higher* frequency mode that aligns with the perturbation in this case. The frequency splitting has increased to $\Delta f = 5$ Hz at this stage.

Figure 5(c) shows a new feature after the lug has been loosened by $1\frac{1}{4}$ turns. The nodal line of the lower frequency mode is now curved around the region where the perimeter tension was lowered. In addition, the frequency splitting has increased to $\Delta f = 16$ Hz. The curvature of the lower frequency mode indicates that the lower tension is not carried symmetrically across the diameter of the drumhead from the perturbed lug. If the result of loosening this lug was simply to create slow and fast axes at right angles to each other there could be no tendency for a nodal line to curve in one direction or the other. The curvature shown in Fig. 5(c) is a commonly observed feature when the tension is adjusted in this manner, and it provides a key to understanding the role of symmetry in drum tuning.

To investigate this further, the opposing lug (across the diameter) was loosened $\frac{1}{2}$ turn [see Fig. 5(d)]. Note that this has the effect of straightening out the previously curved nodal line as well as dramatically increasing the frequency splitting to $\Delta f = 28$ Hz. Thus, loosening the opposing lug has made the decrease in tension more symmetric across the diameter of the drum and has also enhanced the difference between the fast and slow axes on the membrane.

As another illustration, Fig. 6 shows a frame drum that was initially tuned (not shown) and then had a single lug *tightened* (see arrow), rather than loosened, by nearly two full turns. This results in alignment of the modes relative to the perturbation as well as a moderate degree of frequency splitting ($\Delta f = 9$ Hz). In this case, the nodal line oriented in the direction of the perturbation represents the *lower* frequency mode, as waves traveling across this line now have antinodes in the lower tension region. The orthogonal mode has its nodal line curve *away* from the lug that was tightened. This case shows a high degree of curvature with a rather modest amount of frequency splitting.

The results of this section show the expected alignment and frequency splitting of the (1,1) mode. Figures 5 and 6 also illustrate that with this type of detuning procedure at least one of the nodal diameters can become curved—although the frequency splitting and curvature effects are not directly correlated with one another. This latter result is explained with the help of a finite element model below.

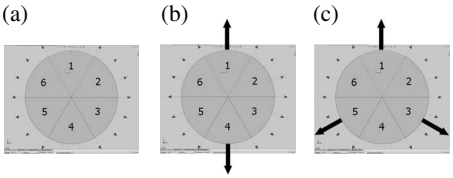


FIG. 7. (a) A screen shot of the finite element model geometry, showing the circular membrane divided into six regions. Small arrows represent the radial forces applied along the six perimeter arcs. (b) Large arrows indicate the simplest tension perturbation: two identical forces added across the diameter at sectors 1 and 4 (twofold perturbation symmetry, $a=2$). (c) Another symmetric perturbation: three identical forces applied at sectors 1, 3, and 5 (threefold perturbation symmetry, $a=3$).

C. Finite element model

Commercial finite element analysis software[13] was used to model the circular drumhead under non-uniform tension. The membrane is broken up into symmetric pie-shaped wedges as shown in Fig. 7(a). Radial forces are applied at the perimeter of each wedge, subject to the constraint that the net applied force equals zero. The software then performs a "prestress modal analysis," with normal mode shapes and frequencies as the output. Air loading was not incorporated into the model, as the effects under study depend only on the symmetry of the system. No attempt was made to reproduce the numerical frequencies or overtone ratios of the drums used in the laboratory; the goal was to make qualitative comparisons of the mode shapes and frequency splittings in response to particular symmetric tension perturbations.

Tightening a single lug on the drum is modeled, in part, by an increased radial force at the corresponding location on the model. Equilibrium must be maintained, but the resulting distribution of tensions at other points on the perimeter is not known. Figure 7(b) shows the simplest tension perturbation that maintains equilibrium, with equal and opposite forces added across the diameter of the membrane. Figure 7(c) shows another acceptable solution—three equal forces separated by 120° (i.e., at sectors 1, 3, and 5). The more general case in which sectors 3, 4, and 5 all act to compensate for the extra force at sector 1 can be treated as a sum of the twofold and the threefold symmetric perturbations depicted in Figs. 7(b) and 7(c). We designate the number of symmetric perturbations (twofold, threefold, etc.) with an integer a, in anticipation of its use in the generalized model developed in Sec. IV.

In Fig. 8 the finite element model is applied to the (1,1) mode of the circular membrane. Figure 8(a) shows a typical example of the model's output for the twofold tension perturbation. Note the alignment of the two perpendicular modes relative to the perturbation and the resulting frequency splitting. The lower frequency mode has its nodal line along the high tension axis, as expected. As required by symmetry, the twofold perturbation does not lead to any curvature of the nodal lines.

Figure 8(b) shows the threefold symmetric perturbation. Here we see curvature of the nodal lines *but no frequency splitting*. The threefold perturbation does not create orthogonal fast and slow axes, which can couple with the (1,1) modes to produce two different frequencies. We infer from

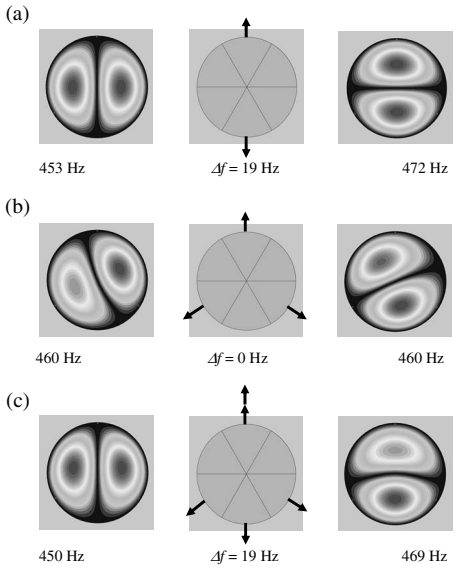


FIG. 8. Finite element results for the (1,1) mode. Nodal regions appear in black. Forces on individual sectors are adjusted such that the total applied force remains constant in all three cases. With uniform tension the (1,1) mode has a frequency of 462 Hz. (a) Twofold perturbation symmetry: tension is added in equal amounts across a diameter, as indicated by the two arrows. The orthogonal (1,1) modes align relative to the perturbation and display frequency splitting. The shapes of the nodal lines are unchanged. (b) Threefold perturbation symmetry: tension is added in equal amounts at 120° intervals as shown by the arrows. The nodal lines become curved and there is a slight decrease in frequency, but no frequency splitting. (c) The twofold and threefold symmetric tension perturbations are combined, as shown by the arrows. Alignment, curvature, and frequency splitting of the (1,1) modes are seen [compare with experimental image in Fig. 5(c)].

these results that the frequency splitting and curvature of the (1,1) modes come from different perturbations. This is consistent with our previous experimental observations (see Figs. 5 and 6) that large frequency differences and large amounts of curvature do not necessarily occur together.

Figure 8(c) shows a more general case—a combination of twofold and threefold symmetric perturbations. The results are similar to what is seen in the laboratory when a single lug is adjusted on an initially tuned drum. As in the experimental data [e.g., Fig. 5(c)], we see the expected alignment of the $(1,1)_-$ and $(1,1)_+$ modes with respect to the perturbation, curvature of one nodal line away from the highest tension wedge, and frequency splitting.

IV. GENERALIZED MODEL

These results for the (1,1) mode are part of a more fundamental principle: The frequency splitting of any degenerate mode depends on both the symmetry of the unperturbed mode and the symmetry of the perturbation. Group theory has been used to investigate these symmetry relations in circular rings,[14] bells,[15,16] and plates[17] with results that are readily modified for the circular drumhead.

A. Frequency splitting

1. Group theory selection rule

The ideal circular drumhead belongs to the symmetry group $C_{\infty v}$ (in the Schoenflies system), where ∞ represents a rotation by any angle about the axis perpendicular to the head through its center and v represents reflection about a vertical plane along any diameter of the head. These operations leave the circular system invariant, as they also do in circular plates, bells, rings, etc. In all such systems group theory predicts that modes containing nodal diameters (or meridians) will be doubly degenerate.[15]

Application of an integer number (a) of symmetrically spaced identical perturbations to the circular drumhead will lower the symmetry group of the system to C_{av}. In general, the perturbed system will exhibit fewer degeneracies than the unperturbed system. Thus, the perturbation will lift some, but not necessarily all, of the degeneracies of the unperturbed system.[16,17]

Previous authors[14–17] have used group theoretical analyses to investigate the splitting of degenerate modes of circular plates, rings, and bells in response to symmetrically placed mass perturbations. In musical acoustics, this approach has been particularly useful in the study of bells, where small deviations from circular symmetry result in "warble" (the equivalent of beats) due to the splitting of degenerate modes. This association of beats with symmetry breaking perturbations of bells was mentioned by Rayleigh (see Ref. 9, Sec. 235).

Group theory is applied to mode splitting of a circular drumhead under particular mass perturbations by Chao[18] and Mathews and Walker.[19] However, the more general discussion of doublet splitting on a flat circular plate described by Murphy *et al.*[17] provides the clearest connection with the current work. For masses placed symmetrically on a circular plate (i.e., equal masses at the vertices of regular polygons inscribed in circles concentric with the plate), the following selection rule has been derived, and verified experimentally by Murphy *et al.*:

$$\frac{2m}{a} = p, \quad p = 1,2,3, \ldots, \tag{2}$$

where m is the number of nodal diameters ($m=1,2,3,\ldots$) and a is the number of symmetric mass perturbations ($a=1,2,3,\ldots$). Frequency splitting of a particular degenerate mode is allowed only if this relation between a and m is satisfied. As pointed out in Ref. 17, the derivation depends only on the symmetry groups $C_{\infty v}$ of the unperturbed system and C_{av} of the perturbed system. Thus, the perturbations need not be small.

To apply this selection rule to tension perturbations on a circular membrane, the only modification is to require that the number of symmetric perturbations must start at two (i.e., $a=2,3,4,\ldots$), so that equilibrium of the drumhead is maintained.

Symmetric tension perturbations of value a are readily expressed by adding terms of the form $T_a \cos(a\theta+\phi_a)$ to the unperturbed tension T_o around the perimeter of the drum. For example, with the orientation angle ϕ taken to be zero, the twofold ($a=2$) perturbation adds a tension $T_2 \cos(2\theta)$ to the uniform tension T_o. This produces a higher tension along the 0-π axis and lower tension along the $\pi/2-3\pi/2$ axis of the membrane. Higher a value perturbations such as the threefold version ($a=3$) discussed in Sec. III are treated in a similar manner. As an arbitrary perturbation may be expressed by

TABLE I. Values of a (number of symmetric perturbations) and m (number of nodal diameters) for which frequency splitting is allowed, based on the selection rule $2m/a$=integer. The $a=2m=a_{\max}$ cases are indicated with an "**X**." Other allowed cases are indicated with an "x."

	a				
m	2	3	4	5	6
1	**X**				
2	x		**X**		
3	x	x			**X**
4	x		x		
5	x			x	
6	x	x	x		x

a sum of symmetric perturbations, $\Delta T(\theta)=\Sigma_{a=2}^{\infty}T_a\cos(a\theta+\phi_a)$, this selection rule provides a powerful tool for analyzing mode splitting due to tension perturbations on the drumhead.

2. Examples of the selection rule

Based on Eq. (2), Table I shows several combinations of a and m values for which frequency splitting is allowed. We can choose a particular mode (characterized by its m value) and determine which symmetric perturbations can lift its degeneracy, or, alternatively, we can choose a particular symmetric perturbation (a value) and determine which modes will be split in response to it.

Note that for a given m value Eq. (2) provides a discrete set of a values, which will allow the mode to split. These perturbation values range from $a=2$ (which splits *all* degenerate modes) up to $a_{\max}=2m$, corresponding to $p=1$ in Eq. (2). We see also that the (1,1) mode can *only* be split by a twofold perturbation. This is consistent with the intuitive notion that that the (1,1) modes should respond to orthogonal fast and slow axes on the membrane. The fact that the threefold perturbation does *not* split the (1,1) mode is consistent with the finite element results of Sec. III. Note that the selection rule does not address changes in mode *shapes*, which will be discussed briefly in Sec. IV B.

Selecting particular perturbation values (columns in Table I), rather than m values, we see that the twofold perturbation ($a=2$) splits all degenerate modes ($m=1,2,3,\ldots$). The threefold perturbation ($a=3$) splits modes with $m=3,6,9,\ldots$, while the fourfold perturbation splits modes with $m=2,4,6,\ldots$. Note that the patterns for even and odd perturbation orders are a bit different. The numerator in Eq. (2) is inherently even: $2m$ is essentially the number of nodal radii, which always occur in pairs. When a takes on an odd integer value, the symmetric tension perturbations do not lie across diameters from one another as they do when a is even. These geometrical relationships of a and m are developed in the remainder of this section.

Although Eq. (2) can be applied to any degenerate mode, in practice, it is the lower frequency modes that ring long enough to be significant in tuning. Furthermore, the number of tuning lugs on the drum (typically 6, 8, or 10) sets a practical limit on one's ability to control the tension. The result is that small values of m and a are of the greatest relevance to the drum tuning process.

A visual representation of the selection rule for the three lowest degenerate modes is shown in Fig. 9. Note that the degeneracy is lifted whenever each symmetric perturbation can be aligned with a nodal radius (or antinodal region in the complementary orientation). The maximum number of perturbations, $a_{\max}=2m$, corresponds to a perturbation at *every* nodal radius. (This is essentially the argument presented in Ref. 17 regarding the symmetric placement of mass perturbations on a circular plate.) Note that only even perturbation values can satisfy the $a_{\max}$ criterion, which turns out to be a particularly significant case.

This $a_{\max}$ case allows all the nodal radii to be aligned with the higher tension in one mode (producing the lower frequency f_-) and all the antinodes to be aligned with the higher tension in the complementary mode (producing the higher frequency f_+). These alignments with respect to the perturbation are again in agreement with Rayleigh's principle that when splitting occurs the orientation is such that one mode obtains the minimum frequency while the other obtains the maximum frequency. Although Rayleigh's principle is obeyed in all cases of allowed splitting, the $a_{\max}$ case is expected to produce the greatest difference between f_+ and f_-.

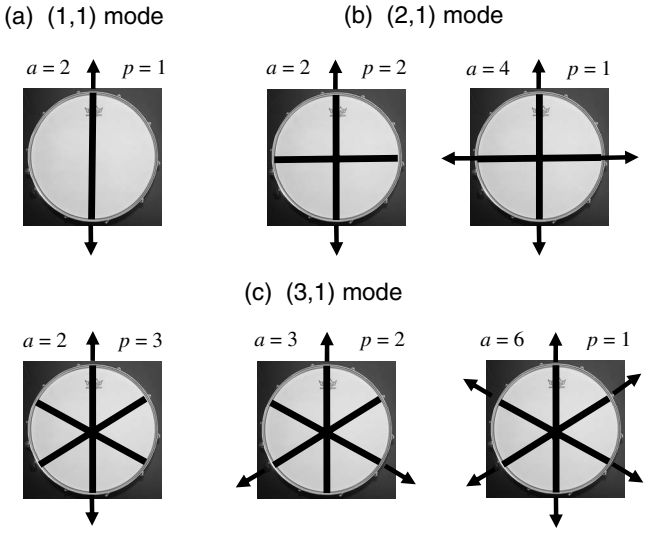


FIG. 9. Diagrams of the three lowest degenerate modes as they relate to the selection rule $2m/a=p$. Arrows represent tension perturbations; solid lines represent nodal diameters. In each case, only one of two orthogonal modes is depicted. (a) The (1,1) mode is only split when $a=2$ ($a_{\max}$). (b) The (2,1) mode is split when $a=2$ or 4 ($a_{\max}$). (c) The (3,1) mode is split when $a=2$, 3, or 6 ($a_{\max}$).

Figure 9 also indicates the integer values of p, based on the selection rule $2m/a=p$. The $p=1$ case corresponds to $a=2m=a_{\max}$, with all nodal radii aligned with a perturbation. For $p=2$ ($a=m$), one half of the nodal radii are aligned with a perturbation. For $p=3$, one-third of the radii align with a perturbation, etc. The pattern terminates at $a=2$, in which case $p=m$ and $1/m$ of the radii are aligned with the perturbation.

3. Perturbation theory

We can also illustrate the significance of the $p=1$ case with a simplified application of first-order perturbation theory.[20] The first-order frequency splitting due to a perturbation H' is described in non-degenerate theory by

$$\Delta f \sim \langle \psi_{mn}(r,\theta)|H'(r,\theta)|\psi_{mn}(r,\theta)\rangle, \tag{3}$$

where the ψ_{mn} represent the unperturbed membrane modes of Eq. (1). We assume that these initially degenerate modes are aligned with a tension perturbation as required by Rayleigh's principle, thus allowing the use of the non-degenerate expression of Eq. (3). Although detailed calculations would require that we include the radial dependence of the resulting stress field, we can gain a qualitative result by simply considering the angular dependence of the tension perturbation, $\Delta T(\theta)\sim\cos(a\theta)$, along with the angular portions of the ψ_{mn}. Thus,

$$\Delta f_{\cos} \sim \langle \cos(m\theta)|\cos(a\theta)|\cos(m\theta)\rangle \sim \begin{cases} 0, & a \neq 2m \\ \dfrac{\pi}{2}, & a = 2m, \end{cases}$$

$$\Delta f_{\sin} \sim \langle \sin(m\theta)|\cos(a\theta)|\sin(m\theta)\rangle \sim \begin{cases} 0, & a \neq 2m \\ -\dfrac{\pi}{2}, & a = 2m. \end{cases} \tag{4}$$

We see that first-order perturbation theory predicts frequency splitting only in the case where $a=2m$, corresponding to $p=1$ in Eq. (2). Although the group theory selection rule provides no information regarding the magnitude of the allowed frequency splitting, both the perturbation calculation and the geometric arguments given above (see discussion of Fig. 9) suggest that the $a_{\max}$ cases will provide the greatest frequency splitting, at least for small values of $\Delta T/T_o$.

In Sec. IV A 4, results from the finite element model are compared with predictions based on the selection rule and the first-order perturbation calculation.

4. Finite element analysis of the frequency splitting

The finite element model was used with symmetric perturbations ranging from $a=2$ to 6. The model was standardized by maintaining a fixed value for the total radial applied tension T_o around the perimeter of the membrane. For each perturbation value a, an amount of tension ΔT was added and subtracted to alternate regions of a drumhead model containing $2a$ sectors. This has the advantage that the fundamental frequency remains constant, nearly independent of both the perturbation value (a) and strength (ΔT). The magnitude of each frequency splitting is then characterized by $\Delta f/f_{\text{ave}}$ for comparison purposes.

Consistency of the finite element results with the selection rule was verified for the first 13 modes, (0,1)–(1,3), with perturbation values $a=2-6$. As suggested by the perturbation calculation, the $a_{\max}$ cases did tend to produce significantly larger frequency splitting than other allowed cases in the numerical model, particularly at lower values of $\Delta T/T_o$.

Among the $a_{\max}$ cases, the lowest (twofold) perturbation acting on modes with a single diameter ($m=1$) resulted in the largest frequency splitting of all. This numerical result is consistent with the drum tuning discussion of Sec. III, which concludes that drummers tune primarily by attempting to eliminate beats caused by the (1,1) mode pair.

B. Normal mode shapes

In our experimental images and in the finite element model, the nodal lines of the mode shapes are often seen to deviate from the concentric circles and diameters of the unperturbed membrane. These distortions are not restricted to the degenerate modes; $m=0$ modes also change shape in response to the perturbations. In practice, the mode shapes are of less importance than the frequency splitting, as the shapes are not heard by the drummer. Experimentally, however, the imaged shapes can serve as "signatures" of specific tension perturbations.

Of particular interest in the laboratory is the (0,2) mode, which can provide a signature of the overall symmetry of the system. The symmetry of the inner nodal circle of the (0,2) mode has been used as an indicator of uniform tension by previous researchers.[1–3] Rayleigh's description (see Ref. 9, p. 346) of Bourget's membrane experiments involving sand patterns includes the following comment: "Any inequality in the tension shows itself by the circles becoming elliptic."

According to the finite element model [see Fig. 10(a)], the shape of the distorted inner circle of the (0,2) mode clearly identifies the perturbation value a. The twofold perturbation distorts the circular node into an elliptical shape, while higher a value perturbations lead to shapes that resemble a-sided polygons. These polygons have been clearly identified in the finite element model in both the (0,2) and (0,3) modes for perturbation values $a=2-6$. In the laboratory, however, *combinations* of perturbations are more typical and the interpretation of the ESPI images is less obvious [see Fig. 10(b)]. Nonetheless, the observed shape of the (0,2) mode can provide some insight as to which regions of the perimeter have too much or too little applied tension.

V. CONCLUSIONS

The qualitative model developed here addresses frequency splitting of the degenerate drumhead modes in response to perturbations in the applied tension. The allowed frequency splitting results are contained in the selection rule $2m/a=$integer, Eq. (2), where the integer a represents the number of symmetric tension perturbations applied around the perimeter and m is the number of nodal diameters of an unperturbed normal mode. For fixed m, the magnitude of the

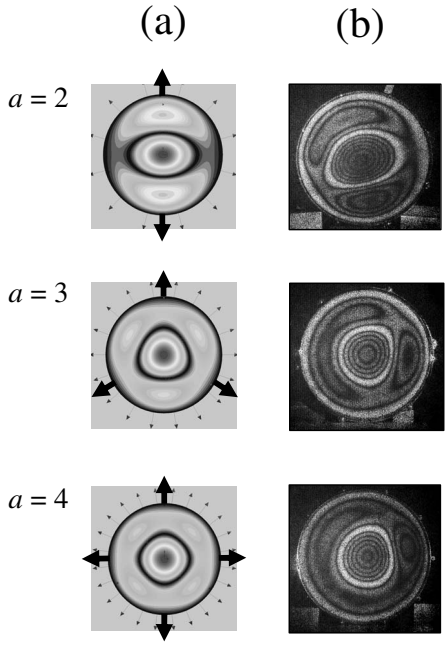


FIG. 10. Examples of the (0,2) mode shape as a perturbation signature. (a) Finite element model results for perturbation values $a=2$, 3, and 4. (b) Experimental speckle-pattern images of a drumhead imply combinations of symmetric perturbations.

frequency splitting is greatest when $a=2m$. Among cases satisfying this criterion, the largest frequency splitting of all occurs with the twofold ($a=2$) perturbation acting on modes with a single nodal diameter ($m=1$).

For musicians, reducing any audible frequency splitting is the goal when tuning a single drumhead to itself. The practical conclusion of this study is that the splitting of the (1,1) mode under the twofold perturbation appears to be the largest contributor to the sound of a drum not being in tune with itself.

This model leaves open the possibility that other $a=2m$ cases [e.g., the fourfold perturbation acting on the (2,1) mode] could produce audible beats even with the twofold perturbation eliminated. These effects have not been tested in our laboratory, but might prove relevant in a more detailed study of drum tuning.

ACKNOWLEDGMENT

Thanks to Puget Sound physics colleague Alan Thorndike for many helpful discussions and suggestions.

[1]T. D. Rossing, C. A. Anderson, and R. I. Mills, "Acoustics of timpani," The Percussionist **19**, 18–31 (1982).
[2]R. S. Christian, R. E. Davis, A. Tubis, C. A. Anderson, R. I. Mills, and T. D. Rossing, "Effects of air loading on timpani membrane vibrations," J. Acoust. Soc. Am. **76**, 1336–1345 (1984).
[3]T. D. Rossing, I. Bork, H. Zhao, and D. O. Fystrom, "Acoustics of snare drums," J. Acoust. Soc. Am. **92**, 84–94 (1992).
[4]I. G. Bassett, "Vibration and sound of the bass drum," The Percussionist **19**, 50–58 (1982).
[5]C. Mei, "Free vibrations of circular membranes under arbitrary tension by the finite-element method," J. Acoust. Soc. Am. **46**, 693–700 (1969).
[6]L. Rhaouti, A. Chaigne, and P. Joly, "Time-domain modeling and numerical simulation of a kettle drum," J. Acoust. Soc. Am. **105**, 3545–3562 (1999).
[7]P. M. Morse and K. U. Ingard, *Theoretical Acoustics* (Princeton University Press, Princeton, NJ, 1968), pp. 209–213.
[8]T. D. Rossing and N. H. Fletcher, *Principles of Vibration and Sound*, 2nd ed. (Springer, New York, 2004), p. 70.
[9]Lord Rayleigh, *The Theory of Sound* (Dover, New York, 1945), Vol. **1**, Chap. 9, pp. 334–339.
[10]T. R. Moore, "A simple design for an electronic speckle pattern interferometer," Am. J. Phys. **72**, 1380–1384 (2004).
[11]T. R. Moore, "Interferometric studies of a piano soundboard," Am. J. Phys. **119**, 1783–1793 (2006).
[12]DrumDial, Inc., www.drumdial.com/ (Last viewed 7/24/2009).
[13]Noran Engineering, NEI FUSION, www.NEiNastran.com (Last viewed 7/24/2009).
[14]R. Perrin, "Selection rules for the splitting of the degenerate pairs of natural frequencies of thin circular rings," Acustica **25**, 69–72 (1971).
[15]R. Perrin and T. Charnley, "Group theory and the bell," J. Sound Vib. **31**, 411–418 (1973).
[16]R. Perrin, "A group theoretical approach to warble in ornamented bells," J. Sound Vib. **52**, 307–313 (1977).
[17]J. P. Murphy, R. Perrin, and T. Charnley, "Doublet splitting in the circular plate," J. Sound Vib. **95**, 389–395 (1984).
[18]B. F. Chao, "Symmetry and terrestrial spectroscopy," Geophys. J. R. Astron. Soc. **66**, 285–312 (1981).
[19]J. Mathews and R. L. Walker, *Mathematical Methods of Physics* (Benjamin/Cummings, Menlo Park, CA, 1970), pp. 443–445.
[20]See, e.g., J. Mathews and R. L. Walker, *Mathematical Methods of Physics* (Ref. 19), Chap. 10.

Detecting overblown flute fingerings from the residual noise spectrum

Vincent Verfaille,[a] Philippe Depalle, and Marcelo M. Wanderley
Sound Processing and Control Laboratory (SPCL), Input Devices and Music Interaction Laboratory (IDMIL), and Centre for Interdisciplinary Research in Music Media and Technology (CIRMMT), Schulich School of Music of McGill University, Montreal, Quebec H3A 1E3, Canada

(Received 19 December 2008; revised 3 September 2009; accepted 27 September 2009)

Producing a tone by increasing the blowing pressure to excite a higher frequency impedance minimum, or overblowing, is widely used in standard flute technique. In this paper, the effect of overblowing a fingering is explored with spectral analysis, and a fingering detector is designed based on acoustical knowledge and pattern classification techniques. The detector performs signal analysis of the strong broadband signal, that is, spectrally shaped by the pipe impedance, and measures the spectral energy during the attack around multiples of the fundamental frequency sub-multiples over the first octave and a half. It is trained and evaluated on sounds recorded with four expert performers. They played six series of tones from overblown and regular fingerings, with frequencies that are octave- and non-octave-related to the playing frequency. The best of the four proposed sound descriptors allows for a detection error below 1.3% for notes with two and three fingerings (C_5, D_5, C_6, and $C\sharp_6$) and below 14% for four (E_6) or five fingerings (G_6). The error is shown to dramatically increase when two fingerings' impedance become too similar (E_6 and A_4 and G_6 and C_5). © *2010 Acoustical Society of America.* [DOI: 10.1121/1.3257205]

PACS number(s): 43.75.Zz, 43.75.Qr [NHF] Pages: 534–541

I. INTRODUCTION AND BACKGROUND

Tone production in the transverse flute involves the interaction of an air jet, in which speed and geometry are controlled by the player, with a resonance of the air column.[1,2] The resonances correspond to minima in the acoustic impedance, $Z(f)$, measured at the embouchure hole. For any combination of open and closed keys (corresponding to what flutists call a fingering), several different minima fall within the playing range. The flutist can usually select which minimum controls the playing regime by adjusting the speed and length of the air jet. A sustained note always consists of a harmonic vibration at the fundamental frequency, plus a variable amount of broadband signal (called residual) due to turbulence produced by the air jet. When emitted, the broadband signal tone is filtered by both the mouth and the pipe resonances. It tends to peak at minima in Z, and therefore has a pronounced tonal quality that is particularly present for high pitch notes, in which the turbulence noise significantly contributes to the sound.[2] The frequency of the impedance minima varies for different fingerings. Frequency minima are not usually harmonically related for high pitched notes emitted with their "standard" fingering. However, frequency minima become almost harmonically related for high pitched notes emitted by increasing the blowing pressure (also called "overblowing") on a low register "regular" fingering, selecting a higher impedance minima. Therefore it may be possible to determine which of a small number of fingerings is used to produce a given note by examining the broadband components of the spectrum especially during the transients.

Interactive music systems such as score-following systems make use of acoustical knowledge to extract performance parameters, for instance, the fundamental frequency, sound intensity, and spectral centroid that relate to pitch, dynamic level, and brightness.[3,4] Their long-term analysis allows for retrieving more performance parameters,[5] such as vibrato, tremolo, trill, etc. On the transverse flute, fingering is another performance parameter of interest. Overblowing of fingerings is widely used in flute playing and often referred to by flutists as "harmonic" or "overtone." In standard technique, $E\flat_5$-$C\sharp_6$ notes are produced using the second vibration mode of $E\flat_4$-$C\sharp_5$ (by convention,[6] the fundamental frequency value is F_0=440 Hz for A_4). Overblowing has also been used since the 19th century to facilitate sound production in the higher register for fast musical phrases, minimizing finger transitions without unduly constraining the air jet control, and in the 20th century as a timbral effect and in pedagogy to learn fine lip control.[7,8]

In this paper, we present a fingering detector that aims to provide a novel tool for interactive music systems. It analyzes the broadband signal spectrum and combines a principal component analysis (PCA) with a linear discriminant analysis (LDA) to classify the fingerings. Section II justifies the applications and principle of a fingering detector based on the sound signal analysis. Section III explains the spectral analysis that is performed on the broadband signal. Section IV describes how classification techniques on the spectral data allow to detect overblown fingerings, and Sec. V concludes and indicates future works.

II. A SOUND SIGNAL-BASED FINGERING DETECTOR

A fingering detector has applications to any application that makes use of performer's gesture, both offline (e.g., ges-

[a]Author to whom correspondence should be addressed. Electronic mail: vincent@music.mcgill.ca

ture analysis from existing recordings, automatic transcription systems that only rely upon pitch, dynamics, and rhythm) and in real time (e.g., interactive systems performing score-following,[4] or using audio effects with acoustic musical instruments).

We could determine which fingering is used either by asking the performer, watching or filming the fingers, using detectors on the keys, or from sound analysis. The first two techniques are discarded as they cannot be automatized. Video analysis is non-intrusive, but it may restrict the flutist's movements or require refinement to be insensitive to movement. The use of sensors to monitor performer's gestures (or direct acquisition[9]) is intrusive and limits expressiveness, as it either requires to use an augmented flute[10,11] (with lower quality standard than the performer's flute) or to add the sensors to the performer's flute (though performers are usually very reluctant to this). Only sound analysis allows for automatic detection, both on existing recordings and in a real-time context.

The fingering-to-pitch relationship is not a one-to-one mapping so proper identification of an overblown fingering is contingent upon the knowledge of both pitch and fingering. Augmented flute such as the Ircam musical instrument digital interface (MIDI)-flute[10] and the Laboratoire de Mécanique et d'Acoustique (LMA) flute[11] are hybrid acquisition systems that remove such ambiguity between $E\flat_4$-$C\sharp_5$ and $E\flat_5$-$C\sharp_6$ by combining key sensors (fingering detection) and sound analysis (fundamental frequency estimation).

In this paper, we aim at detecting fingerings *from the sound only*. Such "indirect acquisition" of gesture[9] requires the identification of its specific *acoustic signature* and its perceptible influence on the sound.[12] It has been used to extract gestural information on guitar,[12,13] violin,[14] multiphonics on wind instruments,[15] clarinet attacks (through timbral changes),[16] and various acoustical parameters of a clarinet from live performance.[5] It is less intrusive and only requires a microphone on the flute headjoint to measure the radiated sound pressure. Performers are more accustomed to the easily removable and light professional microphones than video or sensors.

A previous study investigated the detection of overblown fingerings played by one flute performer[17] using principal component analysis[18] for data reduction and classification after extracting the spectral energies at the first six multiples of $F_0/4$ frequency. The obtained fingering detection was very efficient, but had various limitations, as it only considered octave-related fingerings for a single intermediate level performer, whose non-optimal lip control produced an exaggerated effect that may have made detection easier.

To generalize this technique, we first recorded sounds for a variety of pitches, played by four expert performers, and produced with a larger fingering subset comprising octave- and non-octave-related regular overblown fingerings, together with the regular fingering.

The fingering detection system we developed first requires offline training as follows: (i) from the sound signal, we detect the fundamental frequency F_0, (ii) we infer the corresponding subset of possible regular and overblown regular fingerings, (iii) we detect the fingering from the classification results related to this F_0.

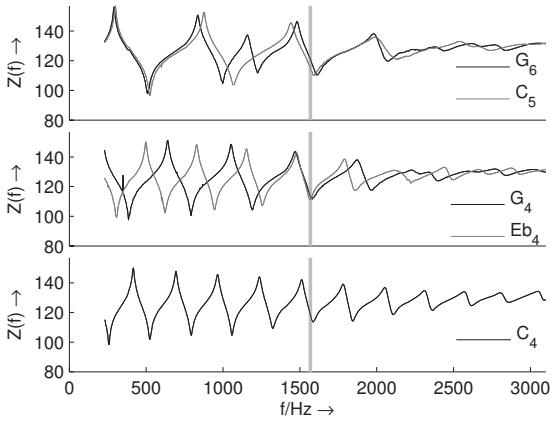


FIG. 1. Acoustical impedances (decibels with respect to 1 MPa s/m^3) measured at the flute embouchure hole for five standard fingerings producing a G_6 tone after Wolfe *et al.* (Refs. 21 and 22). The vertical gray line indicates the fundamental frequency of G_6. Impedance minima correspond to maxima in the residual spectrum of Fig. 2.

III. SPECTRAL ANALYSIS OF THE DATA SET

A. Acoustical signature of overblown fingerings

The various fingerings that can produce a tone of a given pitch differ in terms of the frequency and magnitude of the spectral peaks in the broadband signal. The slight intensity and fundamental frequency differences can be compensated for by experienced flutists adjusting the air jet speed, angle, and lip-to-edge distance.[7,19] This affects the harmonic content[20] but not the frequencies of peaks in the broadband signal (or residual). Moreover, the extent to which a performer controls the relative energy level of harmonics versus residual is limited. We therefore identify the *acoustical signature*[12] *of fingerings* as the *spectral peaks in the broadband signal*, over which the performer has less control.

Figure 1 depicts the impedance measured at the flute input embouchure hole.[21,22] These impedance functions mainly depend on the geometry of the flute they are measured on, in particular, the headcork position in the headjoint and the headjoint position in the pipe.[23] They influence the 250–2300 Hz frequency range where impedance exhibits strong valleys, i.e., where the residual tone spectrum shows discriminant resonances (see Fig. 2). For flutes with similar geometries and tuned to an A_4, these measured impedance functions provide a reliable insight on the residual tone spectrum.

B. Selection of the fingering set

To produce a tone at a given pitch, various playing conditions (which also alter intonation and/or timbre) may be used: the regular fingering, i.e., from the chromatic scale, a regular fingering that is overblown, or an alternate fingering. In this study, we only consider regular fingerings (overblown or not) for two reasons: (i) they are the most obvious and hence standard fingerings that are known by all flutists and (ii) they have a strong enough effect on the sound to hope-

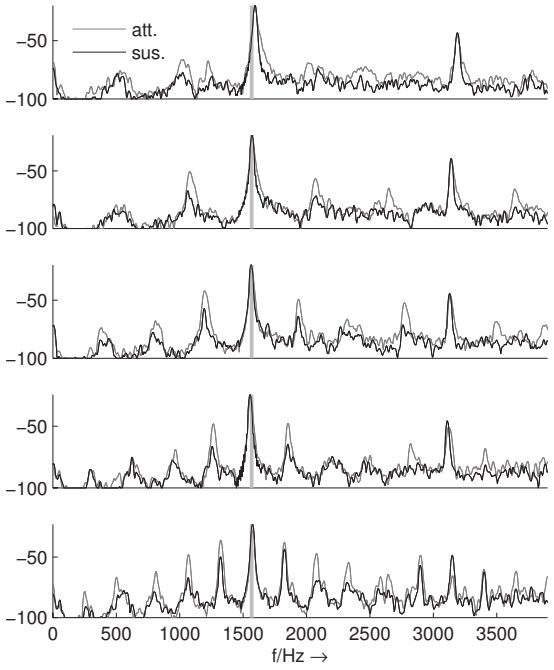


FIG. 2. Mean STFTs (decibel arbitrary) in the first 100 ms (black) versus the next 100 ms (gray) for G_6, $G_6|C_5$ (3), $G_6|G_4$ (4), $G_6|E\flat_4$ (5), and $G_6|C_4$ (6) played *ff* (top to bottom). The attack and sustain clearly share some of the residual resonances.

fully be detected. For these reasons, we select six pitches that can be obtained from one standard fingering and one to four overblown fingerings from the standard chromatic scale fingerings (see Table I). This selection combines sets of octave- and non-octave-related fingerings with few keys changing position (only one key between D_5 and D_4, E_6 and E_4, and G_6 and G_4; only two keys between G_6 and C_5; and only three keys between E_6 and A_4) versus many keys changing position (four to eight for the remaining fingerings). We discard $E\flat_5$ to $B\flat_5$ as their standard fingering is the same as $E\flat_4$ to $B\flat_4$. We denote X_M (alternatively $X_M|X_M$) as a tone played using its regular fingering, $X_M|Y_N$ (R) as a tone of same pitch X_M produced by overblowing the Y_N fingering, where R=round (F_0/F_P) the current tube vibration mode, F_0 the fundamental frequency, and F_P the frequency of the tube first mode. For instance, a C_4 fingering overblown to sound as a G_6 is denoted $G_6|C_4$ (6).

Four expert performers, aged 26–36 and acquainted with extended techniques for 9–16 years, played a tone 20 times with *ff* dynamics and regular articulation for each fingering of each selected pitch. The performers' flutes were similar Boehm flutes with B feet, except the second performer who

TABLE I. Selected fingering set.

Played note	Fingerings (R)
C_5	C_5 (1), C_4 (2)
D_5	D_5 (1), D_4 (2)
C_6	C_6 (1), F_4 (3), C_4 (4)
$C\sharp_6$	$C\sharp_6$ (1), $F\sharp_4$ (3), $C\sharp_4$ (4)
E_6	E_6 (1), A_4 (3), E_4 (4), C_4 (5)
G_6	G_6 (1), C_5 (3), G_4 (4), $E\flat_4$ (5), C_4 (6)

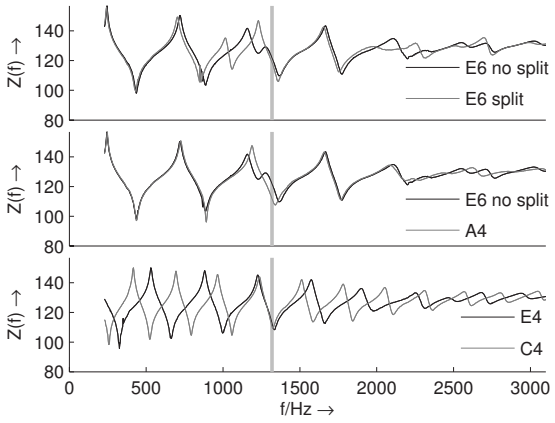


FIG. 3. Acoustical impedances (decibels with respect to 1 MPa s/m^3) measured at the flute embouchure hole (Refs. 21 and 22) for four regular fingerings producing an E_6 tone (20 occurrences of each). Impedance for E_4 and C_4 is far different from that of E_6 (with/without split), whereas both A_4 and E_6 differ little (E_6 split mainly differs around 950 Hz).

had a "split E" mechanism (Powell 2100 with split E and no $C\sharp_5$ key). The three remaining performers had an insert in the G key (performer 1: Powell 2100 with a "crescent" or G-disk, performer 3: Brannen-Cooper Millenium 4582 with a $G\sharp$ donut and a $C\sharp_5$ trill key, and performer 4: Powel Conservatory model 1380) that is used to facilitate the emission of E_6, but also changes the impedance minima peaks below the fundamental frequency (Fig. 3).

C. Spectral sound descriptors

The short-time Fourier transform[24] (STFT) is denoted $X[n,k]$ and computed with a fast Fourier transform (FFT) after apodization with a Hann window, with n the frame index, k the frequency bin for frequency $[(k-1)/L_w]F_s$ Hz, F_c=100 Hz the frame rate, F_s=44 100 Hz the sound sampling rate, and L_w=2048 the FFT size. From the STFT magnitude, the fundamental frequency F_0 is computed and energy measures $E_{\text{meas}}[n,k_{i,l}]$ are derived using the frequency interval

$$I(k_{i,l},\gamma) = k_{i,l}[2^{-\gamma/24};2^{\gamma/24}], \tag{1}$$

which is γ-semitones wide and centered around the ith multiple of the lth sub-multiple (l={2,3,4,5,6}) of the fundamental frequency bin:

$$k_{i,l} = \frac{L_w i F_0}{l F_s}. \tag{2}$$

as depicted in Fig. 4. This is performed on the first octave and a half only (i=1,…,1.5l) since this relative frequency interval corresponds, for the highest pitched note G_6 of our set, to the range [0;2300] Hz within which the most noticeable differences appear (see Sec. III A). Energy measures at multiples of $F_0/4$ allow for detecting octave-related fingerings;[17] energy measures at multiples of F_0/l should allow for detecting non-octave-related fingerings.

A first energy measure is the *local maxima* of energy:

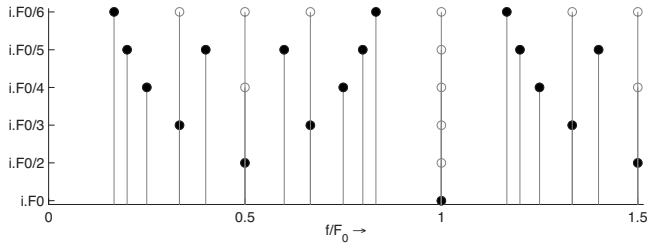


FIG. 4. Analysis frequency ratios (relatively to F_0) for harmonic combs. The combs are centered on F_0/l for $l=\{2,3,4,5,6\}$, where black dots represent the new contribution that each comb adds to previous ones (bottom to up).

$$E_{\max}[n,k_{i,l}] = \max_{k\in I(k_{i,l},\gamma)} |X[n,k]|^2. \tag{3}$$

Depending on interval width γ, a spectral peak may be common to two near frequency intervals, e.g., $I(k_{1,5},\gamma)$ and $I(k_{1,6},\gamma)$. To avoid this issue, we set $\gamma \leq 0.5$, which also corresponds to a good approximation of the bandwidth of impedance minima at a 6 dB loss. However, reducing γ increases the energy measure standard deviation. An alternative is to use a *frequency-integrated* energy measure:

$$E_{\text{win}}[n,k_{i,l}] = \sum_{k=(N_\gamma/2)+1}^{N_\gamma/2} |X[n,k]h_{N_\gamma}[k,k_{i,l}]|^2. \tag{4}$$

The spectrum energy is integrated on the $I(k_{i,l},\gamma)$ interval using a Hann window $h_{N_\gamma}[k,k_{i,l}]$ centered on the $k_{i,l}$ frequency bin and with even size:

$$N_\gamma = 2 \times \text{round}\left[\frac{L_w i F_0}{2lF_s}(2^{\gamma/12}-2^{-\gamma/12})\right]. \tag{5}$$

While 6 measures at $i\ F_0/4$ on the first octave and a half are sufficient to detect octave-related fingerings only,[17] 18 energy measures at $i\ F_0/l$ $(i=1,\ldots,1.5l$ and $l=\{2,3,4,5,6\})$ are required to detect both octave and non-octave-related overblown fingerings with $R\leq 6$. Dimensionality can be reduced by computing the *comb-summed energy measure* that sums the energies of the same harmonic comb (i.e., at harmonically related frequencies $i\ F_0/l$ for $l=\{2,3,4,5,6\}$) normalized by the maximum root mean square energy computed during the attack segment:

$$\mathcal{E}_{l,\text{meas}}[n] = \frac{\sum_{i=1}^{1.5l} E_{\text{meas}}[n,k_{i,l}]}{\max_{m\in\text{attack}} E_{\text{rms}}[m]}, \tag{6}$$

with n the frame index. Note that only five comb-summed energies are used after being converted to decibels. Indeed, the relative energy of the first harmonic $\mathcal{E}_{1,\text{meas}}[n]$ is not used for fingering detection since its contribution is the same for each fingering and it does not bring any useful information to distinguish them.

Knowing that the broadband signal is largely determined from the initial transient and in order to reduce the standard deviation, we use the mean of the energy measures during the attack segment (denoted $\langle E_{\text{meas}}\rangle[k_{i,l}]$ and $\langle \bar{\mathcal{E}}_{\text{meas}}\rangle[l]$) instead of their instantaneous values. The attack segment I_{attack} of an isolated tone is defined as the first 100 ms after the energy envelope exceeds the −50 dB full scale (where 0 dB corresponds to the maximum amplitude) threshold.

IV. DETECTION OF OVERBLOWN FINGERINGS WITH CLASSIFICATION TECHNIQUES

A. Detection probabilities

When training a fingering detector, its performance is evaluated by computing the confusion matrix and then deriving detection error probabilities that depict how well the detector behaves: the lower the error probabilities, the better the detector. A missed detection (or false negative) happens when an $\mathsf{X}_M|\mathsf{Y}_N$ tone being played is not detected; its probability is denoted $P_m(\mathsf{X}_M|\mathsf{Y}_N)$. A false detection (or false positive) happens when an $\mathsf{X}_M|\mathsf{Y}_N$ tone is detected by error; its probability is denoted $P_f(\mathsf{X}_M|\mathsf{Y}_N)$.

B. Simple detection from comb-summed energy maximum

From Fig. 1, we observe an impedance function with three (respectively, 4 and 5) evenly spaced lower peaks for C_5 (respectively, G_4 and $\mathsf{E}\flat_4$). This is a general phenomenon, and "[all overblown fingerings] above a fundamental at a given interval have similar tone qualities."[7] Therefore, the simplest method for detecting an overblown fingering is to look for the l^* value for which the comb-summed energy $\mathcal{E}_{l,\text{meas}}$ is maximum; it should equal the R value. Then, from the fundamental frequency F_0 and vibration regime number R, we can deduce the fingering note name. The left columns of Table II correspond to results with this detection technique. They indicate that, in the best case, the largest proportion of missed and false detection probabilities varies from 6.6% (C_5 with $\langle\mathcal{E}_{\max}\rangle$) to almost 28% ($\mathsf{D}_5$ with $\langle\mathcal{E}_{\max}\rangle$; E_6 and G_6 with both $\langle\mathcal{E}_{\max}\rangle$ and $\langle\mathcal{E}_{\text{win}}\rangle$). Therefore, the maximum of comb-summed energies at different subdivisions of F_0 can give a first impression, but cannot distinguish between the regular fingering and another regular fingering that is overblown. The acoustical signature of fingerings is not extracted accurately enough for the required application. An alternative solution consists of using classification techniques on either all energy measures or all comb-summed energy values.

C. Dimensionality reduction with PCA

Pattern classification techniques such as PCA (Ref. 18) have been successfully used to extract gesture information from the analysis of sound.[13,14,16] PCA decomposes a data set into eigenvectors that define a linear transformation from the original data space to a projected eigenspace, where the eigenvalues reflect the standard deviation conveyed by the data set. When the first few eigenvectors (called principal components) account for a high percentage of the standard deviation in the analyzed data (high eigenvalues), PCA allows us to reduce the data dimensionality with minimal information loss by discarding non-principal components. PCA concentrates information previously spread across several interrelated variables and therefore can be used to interpret the data,

TABLE II. Missed (P_m) and false (P_f) detection probabilities (%) for all subjects during the attack segment (T_{att}=−50 dB) using four energy measures. Probabilities are given for two fingering detectors: using the comb energy maxima (left columns) and using the PCA/LDA cluster classification technique (middle: on all data; right: means and standard deviations with eight-fold cross-validation). The measures with the lowest probabilities are indicated in boldface.

		Comb maxima		Cluster classification			
				All data		Eight-fold cross val.	
	Meas.	P_m	P_f	P_m	P_f	$\bar{P}_m \pm \sigma^2_{P_m}$	$\bar{P}_f \pm \sigma^2_{P_f}$
C_5	$\langle E_{max} \rangle$	⋯	⋯	3.3	3.2	3.3 ± 4.8	3.1 ± 4.6
	$\langle E_{win} \rangle$	⋯	⋯	**1.3**	**1.3**	**0.7** ± 2.0	**0.6** ± 1.8
	$\langle \mathcal{E}_{max} \rangle$	**6.6**	**5.3**	2.0	1.9	**0.7** ± 2.0	**0.6** ± 1.8
	$\langle \mathcal{E}_{win} \rangle$	12.5	6.8	**1.3**	**1.3**	1.0 ± 2.7	1.0 ± 2.7
D_5	$\langle E_{max} \rangle$	⋯	⋯	**0.0**	**0.0**	**0.0** ± 0.0	**0.0** ± 0.0
	$\langle E_{win} \rangle$	⋯	⋯	0.6	0.6	0.7 ± 2.0	0.6 ± 1.8
	$\langle \mathcal{E}_{max} \rangle$	**27.7**	3.7	0.6	0.6	0.7 ± 2.0	0.6 ± 1.8
	$\langle \mathcal{E}_{win} \rangle$	29.7	**0.0**	0.6	0.6	0.7 ± 2.0	0.6 ± 1.8
C_6	$\langle E_{max} \rangle$	⋯	⋯	**0.0**	**0.0**	**0.0** ± 0.0	**0.0** ± 0.0
	$\langle E_{win} \rangle$	⋯	⋯	**0.0**	0.4	2.1 ± 5.1	**0.0** ± 0.0
	$\langle \mathcal{E}_{max} \rangle$	14.2	8.0	5.0	4.3	4.6 ± 3.1	3.9 ± 2.5
	$\langle \mathcal{E}_{win} \rangle$	**8.7**	**3.9**	5.4	5.0	10.0 ± 10.8	11.4 ± 15.9
$C\sharp_6$	$\langle E_{max} \rangle$	⋯	⋯	1.2	**0.0**	**0.8** ± 2.4	**0.0** ± 0.0
	$\langle E_{win} \rangle$	⋯	⋯	**0.4**	**0.0**	3.3 ± 8.4	0.4 ± 1.1
	$\langle \mathcal{E}_{max} \rangle$	14.2	4.5	5.8	5.0	7.1 ± 6.5	3.7 ± 2.1
	$\langle \mathcal{E}_{win} \rangle$	**9.6**	**1.2**	5.4	4.9	10.8 ± 15.2	10.1 ± 15.8
E_6	$\langle E_{max} \rangle$	⋯	⋯	13.7	**13.6**	**13.3** ± 5.3	**13.1** ± 4.8
	$\langle E_{win} \rangle$	⋯	⋯	**13.4**	14.8	13.5 ± 8.8	13.3 ± 6.0
	$\langle \mathcal{E}_{max} \rangle$	6.6	26.7	18.8	19.0	18.9 ± 8.8	18.7 ± 6.4
	$\langle \mathcal{E}_{win} \rangle$	**4.1**	**25.5**	15.9	16.8	16.1 ± 8.0	16.2 ± 5.2
G_6	$\langle E_{max} \rangle$	⋯	⋯	13.0	12.9	13.6 ± 8.9	11.1 ± 7.0
	$\langle E_{win} \rangle$	⋯	⋯	9.7	11.8	**11.1** ± 7.6	**9.2** ± 7.8
	$\langle \mathcal{E}_{max} \rangle$	10.5	29.5	9.7	14.1	15.6 ± 11.8	10.3 ± 8.0
	$\langle \mathcal{E}_{win} \rangle$	**9.7**	**27.8**	**8.2**	**8.5**	21.9 ± 15.8	28.2 ± 10.9

for instance, to analyze guitar timbres.[13] It can also separate the data set dimensions according to standard deviation. It often clusters data projected in the eigenspace and therefore it can be used as a classifier, for instance, to identify the embouchure pressure and attack types on the clarinet.[16]

We form a data matrix $\mathcal{D}$ by combining row vectors, where each row corresponds to an energy measure vector (energies versus comb energies, from the max versus the windowed values, see Sec. III C) measured on the attack segment. A row corresponds to a measure of a given occurrence (up to 20) and fingering (up to 5) used by a subject (4) to obtain a note with the same pitch (i.e., G_6), ***ff*** dynamic and normal articulation. The normalized and reduced matrix $\mathcal{M}$ is obtained by subtracting the mean of each column and dividing by its standard deviation. Then, a singular-value decomposition provides both the eigenvalues and eigenvectors.

For notes produced with a standard fingering and one or two overblown standard fingering (C_5, D_5, C_6, and $C\sharp_6$), the first factorial map (u_1, u_2), which conveys 60%–95% of the standard deviation depending on the energy measure projected, shows clear clusters. The sign of the coordinates in the (u_1, u_2) map is sufficient to separate clusters. In the case of an E_6 tone for which four fingerings can be used (the regular fingering, one octave-related, and two non-octave-related fingerings), clusters for E_6 and $E_6|A_4$ almost entirely overlap (see Fig. 5). This is not surprising as the impedance functions for E_6 and $E_6|A_4$ in Fig. 3 are much more similar than E_6, $E_6|E_4$, or $E_6|C_4$. This is especially true for the three flutes that do not have an E_6 split system. In both cases, the E_6 fingering destabilizes the second impedance minima of A_4 to facilitate the emission of its third regime and help correct the intonation. Only a few values of $\langle E_{meas} \rangle$ highlight this difference in a subtler manner than for the comb-summed measures $\langle \bar{\mathcal{E}}_{meas} \rangle$. Moreover, the short-time spectra are smoother than the measured impedances, and peaks in the residual show maxima that emerge from the background noise by only a few decibels; the differences among reference impedances become even smaller in the residual. For those reasons, the performance decrease in our indirect fingering acquisition method is not surprising in the case of E_6. Similarly for G_6 played with five fingerings—but with a less pronounced effect because impedance functions are more different from one another—none of the four spectral mea-

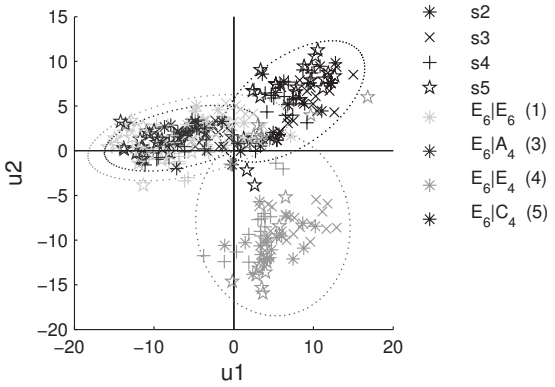


FIG. 5. Spectral energies $\langle E_{\mathrm{win}} \rangle$ projected onto the factorial map (u_1, u_2) for four regular fingerings producing an E_6 tone (20 occurrences of each), where three of the four flutes have no split E mechanism. The factorial map is built from the two eigenvectors with highest eigenvalues. Each marker shape represents a performer, and each color represents a fingering. From the 95% confidence ellipses (doted lines), it appears that E_6 and $\mathsf{E}_6|\mathsf{A}_4$ clusters do almost entirely overlap.

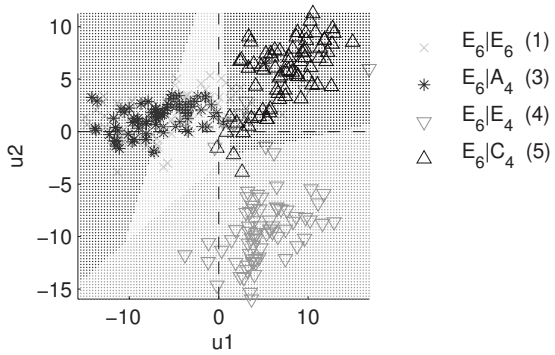


FIG. 6. Cluster areas as detected by LDA applied to spectral energies $\langle E_{\mathrm{win}} \rangle$ projected onto the factorial map (u_1, u_2) for four regular fingerings producing an E_6 tone. Each fingering has its own color and marker shape.

sures provides five perfectly distinct clusters: between two and four clusters are always slightly overlapping.

To some extent, performers were correcting intonation—maybe unconsciously—from one occurrence to another, even though they were asked not to. We observed intonation differences depending on the performer, the flute, and the note. However, this was neither systematic nor reliable: appending F_0 to the energy measure in the matrix data decreased the classification detector performances.

D. Cluster classification with LDA

For both E_6 and G_6, PCA clusters partially overlap, and their relative orientation to one another does not allow for a simple classification using vertical and horizontal lines. This still holds after a varimax cluster rotation procedure that tends to spread clusters while vertically or horizontally orienting them.

Following a method to detect which violin string is bowed,[14] we combine PCA with LDA (Ref. 25) on a PCA factorial map in order to detect each cluster's area. While PCA tends to perform data classification by providing uncorrelated dimensions, LDA tends to perform class separation and classification, but requires data to be tagged according to the cluster to which they belong (which is not the case for PCA). Tagging data requires a training phase during which LDA detects regions that best separate classes in a given space by fitting a multivariate normal density to each class. When applying LDA in a PCA factorial map, PCA clusters directly correspond to the classes that LDA separates, so LDA gives greatest separation to classes corresponding to fingerings. An example of cluster areas as detected by LDA is depicted in Fig. 6.

In practice, the first factorial map (u_1, u_2) carries most of the variance but does not necessarily provide the most distinct and compact clusters. Therefore, we computed the detection probabilities for PCA/LDA on each of the six factorial maps formed by i, j pairs among the four first eigenvalues and selected the (u_i, u_j) factorial map that minimizes $P_m + P_f$. The selected map (u_i, u_j) may then differ among fingerings for the same pitch.

E. Stratified 8-fold cross-validation

To evaluate the PCA/LDA fingering detector performance, we trained the system on a limited data set and computed the detection probabilities on a larger data set; this allows to extrapolate performances before using it for real applications. The N-fold cross-validation consists of (i) partitioning the data set into N folds or disjoint subsets containing $1/N$th of the data set and (ii) for each fold, removing it, training the detector on the remaining $N-1$ folds, and evaluating performance on the whole data set (this is repeated N times). Each fold is proportionally representative of each class: this procedure is called *stratified N-fold cross-validation* (SFCV), and it has been observed that stratification helps to limit the estimator biais, especially for a sufficient number of folds[26] (typically 8). For each stratified fold, the training phase consists of applying dimensionality reduction (PCA) and classification (LDA), and the evaluation phase consists—for fingerings producing notes of a same pitch—of computing the missed and false detection probabilities for each fingering separately, and then for all fingerings. The mean probability values on the $N=8$ training phases are computed, again for each single fingering and for all fingerings of the same pitch.

F. Results and discussion

Table II shows the LDA classification results for the four measures computed on the attack segment and using the PCA factorial map that minimizes the sum $P_m + P_f$ for all selected fingerings of the six pitches in our set up. All tones with two and three fingerings (C_5, D_5, C_6, and $\mathsf{C}\sharp_6$) have missed and false detection probabilities below 1% for the spectral energy measure providing best detection (in boldface in Table II), with the worst cases being $P_m = 0.8 \pm 2.4\%$ for $\mathsf{C}\sharp_6$ with $\langle E_{\max} \rangle$ and $P_f = 0.6 \pm 1.8\%$ for C_5 with $\langle E_{\mathrm{win}} \rangle$ and $\langle \mathcal{E}_{\max} \rangle$. Lowest detection errors do not necessarily correspond to the energy measure with highest variance explained by the two eigenvalues of the current factorial map: comb-summed energies $\langle \bar{\mathcal{E}}_{\mathrm{meas}} \rangle$ better concentrate

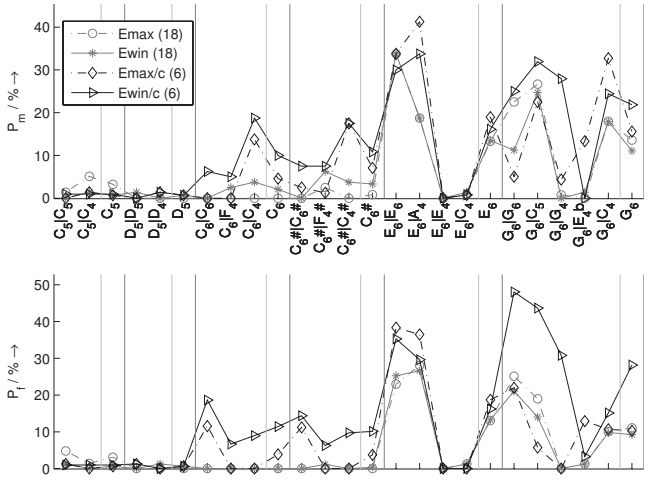


FIG. 7. Missed (upper) and false (lower) detection probabilities using linear discriminant analysis with eight-fold cross-validation for each fingering.

the information (which contains most of the variance) and provide more compact clusters, but clusters are nearer to one another so detection errors also increase.

Conversely for the E_6 and G_6 notes with four or five possible fingerings, detection error probabilities almost reach 14% for the energy measure with best detection: $P_m = 13.3 \pm 5.3\%$ for E_6 with $\langle E_{\max} \rangle$ and $11.1 \pm 7.6\%$ for G_6 with $\langle E_{\text{win}} \rangle$; $P_f = 13.1 \pm 4.8\%$ for E_6 with $\langle E_{\max} \rangle$ and $9.2 \pm 7.8\%$ for G_6 with $\langle E_{\text{win}} \rangle$. Because some fingerings have very similar impedances (E_6 with and without a split E mechanism, and $\mathsf{E}_6|\mathsf{A}_4$ in Fig. 3, and G_6 and $\mathsf{G}_6|\mathsf{C}_5$ in Fig. 1), the related PCA clusters often overlap, resulting in a fingering mismatch and higher missed and false detections (25%–40% for E_6 and 20%–50% for G_6, as depicted by the stratified eight-fold cross-validation results for each fingering in Fig. 7). The slightly better results obtained for G_6 compared to E_6 are explained by the degree of similarity between fingering impedances. Three of the five impedances for G_6 look quite similar (G_6, C_5, and C_4), whereas two of the four impedances for E_6 look *very similar* (E_6 and A_4). The resulting detection errors cannot be prevented with the current spectral sound descriptors.

The comb maxima detection method is outperformed by the PCA/LDA cluster classification with and without stratified eight-fold cross-validation using $\langle \bar{\mathcal{E}}_{\text{meas}} \rangle$: the detection error sum ($P_m + P_f$) is 30%–90% lower. Contrary to the comb maxima detection, the PCA/LDA's missed and false detection probabilities have similar values because LDA limits separate classes according to their variance. Except for G_6 (no SFCV) and C_5 (with SFCV), the most efficient detection for the selected fingering subset happens using the local maxima of frequency-integrated energies but not the comb-summed energies.

V. CONCLUSIONS AND FUTURE WORKS

A detector of overblown flute fingering was designed for interactive music systems based on the spectral analysis of the residual noise at multiples of $F_0/l, l=\{2,3,4,5,6\}$ on the first octave and a half, combining principal component analysis on energy measures, and cluster classification with linear discriminant analysis. A set of tones was produced by four expert performers and contains six pitches with a limited subset of regular and overblown regular fingerings (two to five different octave- and non-octave-related fingerings) that were played with normal articulation and ***ff*** dynamics. The system was trained on these sounds, and its performance was evaluated and extrapolated using stratified eight-fold cross-validation.

For tones with two octave-related fingerings (C_5 and D_5) and three octave- and non-octave-related fingerings (C_6 and $\mathsf{C}\sharp_6$), the maximum proportion of missed and false detection probabilities was below 1.3% for the most efficient classification among the four spectral measures tested. The acoustical correlate of the fingering was reliable for these four pitches and their relative regular fingerings since the system results were stable for one intermediate level performer playing octave-related overblown fingerings[17] and for four expert performers playing non-octave-related overblown fingerings. Further analyses will have to be carried out to better reflect their acoustical signature and enhance the cluster classification, in particular, where tones with four or five possible regular fingerings are concerned. When more than three possible regular fingerings can produce the same high pitch note, at least one overblown fingering may have a similar impedance to the non-overblown regular fingering (e.g., $\mathsf{E}_6|\mathsf{A}_4$ and E_6). In this case, performance dramatically decreases, as illustrated by high missed and false detection probabilities for G_6 (11.1% with $\langle E_{\text{win}} \rangle$) and E_6 (13.3% with $\langle E_{\max} \rangle$).

Our future work will consider complementary sound descriptors to better represent the acoustical signature of fingerings and increase the detector performances. We will explore the generalization of this technique to data sets of musical excerpts with various dynamics (***ff*** and ***pp***) and articulations (normal, *staccato, and legato*).

ACKNOWLEDGMENTS

The authors would like to thank the anonymous reviewers for their valuable comments, the performers who played for them, Cléo Palacio-Quintin for advice during the experiment score design, Joe Wolfe for sharing acoustic impedance measures, and Louise Campbell and Matthew Davies for proof-reading this paper. This research was supported by grants from Quebec's *Ministère du Développement Économique, de l'Innovation et Exportation* (Grant No. PSR-SIIRI-037), the French *Agence Nationale de la Recherche* (Consonnes Project, No. ANR-05-BLAN-0097-01), the *Fond Québécois pour la Recherche en Nature et Technologies*, and the Natural Sciences and Engineering Research Council of Canada.

[1]J. W. Coltman, "Sounding mechanism of the flute and organ pipe," J. Acoust. Soc. Am. **44**, 983 (1968).
[2]N. H. Fletcher and T. D. Rossing, *The Physics of Musical Instruments*, 2nd ed. (Springer-Verlag, Berlin, 1998), pp. 503–551.
[3]R. Rowe, *Interactive Music Systems: Machine Listening and Composing*, 1st ed. (The MIT Press, Cambridge, MA, 1992).
[4]N. Orio, S. Lemouton, and D. Schwartz, "Score following: State of the art and new developments," in International Conference on New Interfaces for Musical Expressions, (2003), pp. 36–41.

[5]M. Puckette and C. Lippe, "Getting the acoustic parameters from a live performance," in Proceedings of the 3rd International Conference on Music Perception and Cognition (1994), pp. 328–333.
[6]R. W. Young, "Terminology for logarithmic frequency units," J. Acoust. Soc. Am. **11**, 134–139 (1939).
[7]R. Dick, *The Other Flute—Performance Manual of Contemporary Techniques*, 2nd ed. (Multiple Breath Music Company, New York, 1989).
[8]P.-Y. Artaud, *Harmonics Exercise Book: Simple Notes and Multisounds for Flute* (Gérard Billaudot, Paris, 1992).
[9]M. M. Wanderley and P. Depalle, "Gestural control of sound synthesis," Proc. IEEE **92**, 632–644 (2004).
[10]D. Pousset, "La flûte-MIDI, l'historique et quelques applications (The MIDI-flute: History and applications)," MS thesis, Université de Paris-Sorbonne, Paris (1992).
[11]S. Ystad and T. Voinier, "A virtually real flute," Comput. Music J. **25**, 13–24 (2001).
[12]C. Traube, P. Depalle, and M. M. Wanderley, "Indirect acquisition of instrumental gesture based on signal, physical and perceptual information," in International Conference on New Interfaces for Musical Expressions (2003), pp. 42–47.
[13]N. Orio, "The timbre space of the classical guitar and its relationship with the plucking techniques," in Proceedings of the International Computer Music Conference (ICMC'99) (1999), pp. 391–394.
[14]A. Krishnaswamy and J. O. Smith, "Inferring control inputs to an acoustic violin from audio spectra," in *International IEEE Conference on Multimedia and Expo (ICME '03)* (2003), Vol. **I**, pp. 733–736.
[15]G. Assayag, M. Castellengo, and C. Malherbe, "Nouvelles techniques instrumentales—Composition et formalisation (New instrumental techniques—Composition and formalization)," Ircam Technical Report No. 38, Ircam, Paris, France, 1985.
[16]E. B. Egozy, "Deriving musical control features from a real-time timbre analysis of the clarinet," MS thesis, Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA (1995).
[17]C. Kereliuk, B. Scherrer, V. Verfaille, P. Depalle, and M. M. Wanderley, "Indirect acquisition of fingerings of harmonic notes on the flute," in *Proceedings of the International Computer Music Conference (ICMC'07)* (2007), Vol. **I**, pp. 263–266.
[18]I. T. Jolliffe, *Principal Component Analysis*, Springer Series in Statistice **XXIX**, 2nd ed. (Springer, New York, 2002).
[19]P.-Y. Artaud, *Flûtes au présent/Present Day Flute*, 2nd ed. (Gérard Billaudot, Paris, 1995).
[20]J. W. Coltman, "Jet offset, harmonic content, and warble in the flute," J. Acoust. Soc. Am. **120**, 2312–2319 (2006).
[21]J. Wolfe, J. Smith, J. Tann, and N. H. Fletcher, "Acoustic impedance spectra of classical and modern flutes," J. Sound Vib. **243**, 127–144 (2001).
[22]J. Wolfe, J. Smith, J. Tann, and N. H. Fletcher, "Acoustics of classical and modern flutes: A compendium of impedance spectra, sound spectra, sounds and fingerings," J. Sound Vib. Electronic supplement, available at http://www.phys.unsw.edu.au/music/flute/virtual (Last viewed May 12, 2009).
[23]J. Smith, J. Wolfe, and M. Green, "Head joint, embouchure hole and filtering effects on the input impedance of flutes," in Proceedings of the Stockholm Music Acoustics Conference (SMAC 03) (2003), pp. 295–298.
[24]A. M. Noll, "Short-time spectrum and "cepstrum" techniques for vocal pitch detection," J. Acoust. Soc. Am. **36**, 296–302 (1964).
[25]G. J. McLachlan, *Discriminant Analysis and Statistical Pattern Recognition*, Wiley Series in Probability and Statistics Vol. **544** (Wiley-Interscience, New York, 2004).
[26]R. Kohavi, "A study of cross-validation and bootstrap for accuracy estimation and model selection," in International Joint Conference on Artificial Intelligence (1995), pp. 1137–1145.

A unifying fractional wave equation for compressional and shear waves

Sverre Holm
Department of Informatics, University of Oslo, P.O. Box 1080, NO-0316 Oslo, Norway

Ralph Sinkus
Institute Langevin, ESPCI ParisTech, 10 rue Vauquelin, F-75231 Paris Cedex 05, France

(Received 15 June 2009; revised 17 October 2009; accepted 19 October 2009)

This study has been motivated by the observed difference in the range of the power-law attenuation exponent for compressional and shear waves. Usually compressional attenuation increases with frequency to a power between 1 and 2, while shear wave attenuation often is described with powers less than 1. Another motivation is the apparent lack of partial differential equations with desirable properties such as causality that describe such wave propagation. Starting with a constitutive equation which is a generalized Hooke's law with a loss term containing a fractional derivative, one can derive a causal fractional wave equation previously given by Caputo [Geophys J. R. Astron. Soc. **13**, 529–539 (1967)] and Wismer [J. Acoust. Soc. Am. **120**, 3493–3502 (2006)]. In the low $\omega\tau$ (low-frequency) case, this equation has an attenuation with a power-law in the range from 1 to 2. This is consistent with, e.g., attenuation in tissue. In the often neglected high $\omega\tau$ (high-frequency) case, it describes attenuation with a power-law between 0 and 1, consistent with what is observed in, e.g., dynamic elastography. Thus a unifying wave equation derived properly from constitutive equations can describe both cases.
 [DOI: 10.1121/1.3268508]

PACS number(s): 43.80.Cs, 43.20.Hq, 43.20.Jr, 43.20.Bi [CCC] Pages: 542–548

I. INTRODUCTION

It has been well established for quite some time that wave attenuation in many materials including tissue and even on the cellular level follows a power-law over several decades of frequency variation, i.e.,

$$\alpha = \alpha_0 |\omega|^y, \tag{1}$$

with α_0 and y being constants that characterize the medium. Typically compressional waves in the megahertz regime are attenuated with an exponent y between 1 and 2,[1] and, in particular, in human tissue. Shear wave dispersion properties have recently gained increasing interest in the medical community for the additional characterization of pathologies that it may provide. This is due to the development of new imaging modes such as magnetic resonance (MR) elastography[2,3] and ultrasound elastography.[4] Shear wave attenuation in the hertz regime in tissue normally demonstrates exponents below 1,[5] and the same range is found for plastics in the megahertz range.[6]

The aim of this paper is to demonstrate that these different dispersion behaviors can be explained theoretically using one common wave equation where attenuation has been introduced via fractional derivative operators. Causality of the wave equation has been a problem in many previously proposed fractional wave equations. This is taken care of when the wave equation is derived from a proper constitutive equation; in this case, it is a generalized Hooke's law with a fractional derivative term. The high and low value ranges for the power-law exponent y can be related to two different regimes, corresponding to two different approximations. Thus this study should provide one step toward a better understanding of the relationship between observed behavior and the underlying physics. The objective is in the end to understand the role of the microstructure of the underlying material. In medical imaging such insight could lead to better and more precise diagnoses.

We start by reviewing the various previous approaches for introducing loss and dispersion into the wave equation and then introduce the Caputo wave equation. Then we show that it can be used to find two different approximations depending on the product of the frequency and the relaxation time. The low-frequency case describes compressional wave propagation and some shear wave propagation with power-law exponents between 1 and 2. Then an often neglected high-frequency approximation is derived which agrees with observations of various cases of shear wave propagation with power-law exponents down to almost 0.

II. LOSSY WAVE EQUATIONS

For both compressional and shear waves, propagation may be described by a wave equation with a loss operator L,

$$\nabla^2 u - \frac{1}{c_0^2}\frac{\partial^2 u}{\partial t^2} + Lu = 0, \tag{2}$$

where c_0 is a propagation velocity. For simplicity, in the rest of this section, we will only discuss compressional waves. As is well established, the form of the loss operator is only simple to justify from the underlying physics for a few values of y, namely, $y=2$ and $y=0$.

A. Quadratic frequency dependence, $y=2$

The first case is the classical viscoelastic wave equation which applies, e.g., to attenuation in air and water. In that case,

$$Lu = \tau \frac{\partial}{\partial t}(\nabla^2 u), \tag{3}$$

which gives a power-law with exponent $y=2$ under the condition that $\omega\tau \ll 1$, where τ is a relaxation time that characterizes the medium.[7] This case is the most usual one, but for attenuation in saturated sand in marine sediments, a high $\omega\tau$ product is also of interest, giving a power-law $y=0.5$.[8] The way the loss term has been introduced here is actually in agreement with basic requirements concerning the invariance of the wave equation with respect to rotation and translation.[9] Therefore, the viscosity tensor must be linked to second order spatial derivatives of the displacement, hence the presence of $\nabla^2 u$. The first order time derivative represents a dashpot in terms of a mechanical representation. Thus the wave equation represents a spring and dashpot in parallel, which is the classical Voigt model.

B. Frequency-independent attenuation, $y=0$

Another physics-based lossy wave equation is given by $Lu \propto -\partial u/\partial t$. It describes a medium which in the high-frequency limit has a constant loss, independent of frequency ($y=0$). However, this case applies primarily to electromagnetic waves, i.e., an electric field in a conducting medium or transverse electric waves in a homogeneous isotropic plasma.[1]

C. Fractional approaches

Szabo[1] developed a wave equation for y between 0 and 2 which matched the last case at $y=0$. For $y=2$ it was similar to an approximation to Eq. (3) where the viscoelastic loss term had been approximated by a third order time derivative, $Lu \propto \partial^3 u/\partial t$.[3] His work was based on the observation that the order of the time derivative for the loss term in the partial differential equation is one more than the exponent y.

Chen and Holm[10] interpreted the Szabo operator in terms of fractional derivatives, as it was found that it implicitly implies the Riemann–Liouville fractional time derivative. It was also found that better numerical properties can be obtained if it is replaced by a fractional operator due to Caputo.[11]

For the purpose of this paper, the fractional derivative is easiest to define in the frequency domain where it is an extension of the Fourier transform (FT) of an nth order derivative:

$$\mathrm{FT}\left(\frac{d^n u(t)}{dt}\right) = (i\omega)^n U(\omega). \tag{4}$$

The fractional derivative of arbitrary order can be understood as a generalization where the integer n is replaced with a real number. The time domain equivalent involves a convolution integral, see Sec. IV C. This implies that the fractional derivative, in contrast to integer order derivatives, has memory.[10]

Using this concept, the modified Szabo loss operator is proportional to

$$Lu \propto -\frac{\partial^{y+1} u}{\partial t^{y+1}}. \tag{5}$$

There is an implicit sign change for $y=2$.

Chen and Holm[12] reformulated the lossy wave equation in terms of a fractional derivative in the spatial domain, i.e., a fractional Laplacian. The loss operator in that case is

$$Lu \propto -\frac{\partial}{\partial t}(-\nabla^2)^{y/2} u. \tag{6}$$

The advantage of this approach is that it falls exactly back to the loss operator of Eq. (3) for $y=2$, not just its approximation with a third order time derivative.

A disadvantage of both of these formulations is that they do not guarantee a causal solution. For this reason, Szabo's original loss operator is based on a parabolic approximation to Eq. (5) which was shown to satisfy the Kramer–Kronig relations.[1] A similar approach could probably have been applied to Eq. (6) as well.

Both the loss operators of Eqs. (5) and (6) agree with the wave equation with constant loss with frequency ($y=0$). By neglecting this case which usually does not describe mechanical waves anyway, Wismer[13] proposed a new loss operator

$$Lu = \tau^{z_0} \frac{\partial^{y-1}}{\partial t^{y-1}}(\nabla^2 u), \tag{7}$$

where the order of the derivative is between 0 and 1. It is equivalent to the loss operator of the viscoelastic equation for $y=2$. Just like the loss operators proposed by Szabo[1] and Chen and Holm,[12] Eq. (7) seems to have been found by inspection, as little justification is given with a basis in physics arguments.[13]

Equation (7) will give a causal wave equation,[13] unlike Eqs. (5) and (6). One can ask oneself why this is the case when no causality constraint has been forced upon it. This seems to hint at some deeper justification in the underlying physics. Another feature which points in the same direction is the presence of the $\nabla^2 u$ operator in the loss term and the implied invariance with respect to rotation and translation. Equation (7) actually turns out to be based on a constitutive equation which is a fractional stress-strain relationship. It was first used to derive a lossy wave equation by Caputo in 1967,[11] as pointed out by Kelly and McGough.[14]

III. FRACTIONAL CAPUTO WAVE EQUATION

The derivation starts with the assumption of a fractional relationship between stress T and strain S. This is a generalized version of the lossy Hooke's law, as used by Caputo:[11]

$$T = cS + \eta \frac{\partial^{z_0} S}{\partial t^{z_0}}, \tag{8}$$

where c is the stiffness and η the viscosity. z_0 covers the range from $z_0=0$ (no viscosity) to $z_0=1$ (classical viscosity,

i.e., dashpot). The case with z_0 approaching 0 is of special interest and will be discussed in Sec. IV C. Even higher values for z_0 than 1 could also be possible as we will see later. Bagley and Torvik[15] traced the origins of this relationship back to 1921 and motivated it by reference to molecular theories. Since then, this relationship has been used extensively in the shear wave domain and found to agree well with measurements.[5] Equation (8) is combined with Newton's law:

$$\rho\frac{\partial^2 u}{\partial t^2} = \frac{\partial T}{\partial x}, \tag{9}$$

and because strain and displacement u are related via $S = \partial u/\partial x$, one obtains a wave equation with a loss operator similar to Eq. (7), i.e.,

$$\nabla^2 u - \frac{1}{c_0^2}\frac{\partial^2 u}{\partial t^2} + \tau^{z_0}\frac{\partial^{z_0}}{\partial t^{z_0}}(\nabla^2 u) = 0. \tag{10}$$

The mechanical representation of this equation is a spring and a so-called spring-pot in parallel.[16] It therefore represents the fractional generalization of the classical Voigt model where the spring-pot can continuously interpolate between a pure spring, $z_0=0$, and a pure dashpot, $z_0=1$. In this equation, the parameters are the relaxation time, $\tau^{z_0}=\eta/c$, and the propagation velocity, $c_0^2=c/\rho$.

A. Compressional and shear wave equations

In the rest of this derivation we will follow Chap. 4.2.6 of Royer and Dieulesaint[7] who generalize the derivation above to an isotropic solid. The only difference is that Royer and Dieulesaint assumed $z_0=1$ throughout their derivation.

In an isotropic solid, the only stiffness parameters are $c_{11}=K$ and $c_{44}=\mu$ for compression and shear, where K is the bulk modulus or the inverse of the compressibility, and μ is the shear modulus. The viscosity coefficients are η_{11} and η_{44} for compression and shear, respectively.

The wave equation of Royer and Dieulesaint[7] describes combined compressional and shear wave propagation. If only one wave type is present at a time in the form of a plane wave, the equation can further be separated into a compressional wave equation and a shear wave equation, with $j=1$ or $j=4$, respectively. They both obey the same fractional wave equation

$$c_{jj}\frac{\partial^2 u}{\partial x^2} - \rho\frac{\partial^2 u}{\partial t^2} + \eta_{jj}\frac{\partial^{z_0}}{\partial t^{z_0}}\left(\frac{\partial^2 u}{\partial x^2}\right) = 0 \tag{11}$$

or

$$\frac{\partial^2 u}{\partial x^2} - \frac{1}{c_{0,jj}^2}\frac{\partial^2 u}{\partial t^2} + \tau_{jj}^{z_0}\frac{\partial^{z_0}}{\partial t^{z_0}}\left(\frac{\partial^2 u}{\partial x^2}\right) = 0. \tag{12}$$

The parameters are the zero-frequency propagation speed, $c_{0,jj}=\sqrt{c_{jj}/\rho}$, and the relaxation time $\tau_{jj}^{z_0}=\eta_{jj}/c_{jj}$. In the rest of the derivation, the subscripts jj will in most cases be dropped as it will be clear from the context whether both compressional and shear waves are meant, or only a single wave type. It is also straight-forward to extend the wave equation to the three dimensional case, as in Eq. (10).

It is evident that insertion of $z_0=1$ in Eq. (12) gives the viscoelastic equation, Eq. (3). When $z_0=0$ the equation degenerates to the elastic, loss-less, wave equation as the first and the last terms merge into one.

B. Dispersion relation

The dispersion relation for the Caputo wave equation, Eq. (10), is found by assuming a harmonic, plane wave solution, $u(x,t)=\exp i(\omega t-kx)$. The one dimensional case is also assumed:

$$k^2 - \frac{\omega^2}{c_0^2} + (i\omega\tau)^{z_0}k^2 = 0. \tag{13}$$

It can be solved to yield

$$k^2 = \frac{\omega^2}{c_0^2}\frac{1}{1+(i\omega\tau)^{z_0}} = \frac{\omega^2}{c_0^2}\frac{1}{1+(\omega\tau)^{z_0}\left(\cos\frac{\pi z_0}{2} + i\sin\frac{\pi z_0}{2}\right)}. \tag{14}$$

IV. LOW-AND HIGH-FREQUENCY DISPERSION RELATIONS

In most analysis of dispersion relations, such as Eq. (14), it is assumed that $\omega\tau$ is small to make further approximations. It is here we will make it a point to analyze both cases, i.e., either the regime where $\omega\tau \ll 1$ or the regime where $\omega\tau \gg 1$.

A. Low $\omega\tau$ approximation, $y=z_0+1$

This case is determined by $(\omega\tau)^{z_0} \ll 1$, which means that elasticity dominates over viscosity. It covers most of the compressional wave cases of interest because the compressional modulus in biological tissue is very large (of the order of gigapascals). In, e.g., air $\tau_{11} \approx 10^{-10}$ s, implying that for $z_0=1$, frequencies up to the megahertz range are in this regime. In water $\tau_{11} \approx 10^{-12}$ s,[17], so frequencies up to hundreds of megahertz should be covered.

In the human body, however, τ_{44} is in the order of 0.002 s,[2] so only frequencies much less than about 10 Hz, such as in quasi-static elastography, are in this range.

In this case, the cosine and sine terms in the denominator of Eq. (14) are small compared to unity and one can approximate as follows:

$$k \approx \frac{\omega}{c_0}\left(1 - \frac{(\omega\tau)^{z_0}}{2}\cos\frac{\pi z_0}{2} - i\frac{(\omega\tau)^{z_0}}{2}\sin\frac{\pi z_0}{2}\right). \tag{15}$$

The first term is the dispersion-less and attenuation-less propagation term. It is followed by a dispersion term and by the imaginary part which describes attenuation. Note that the dispersion and attenuation terms form a Hilbert transform pair, showing that the wave equation is causal.[18] Since $k=\beta-i\alpha$, the attenuation term is

$$\alpha = \alpha_0|\omega|^{z_0+1}, \quad \alpha_0 = \frac{\tau^{z_0}}{2c_0}\sin\frac{\pi z_0}{2}. \tag{16}$$

When written in the form of Eq. (1), the power-law is

$$\alpha = \alpha_0 |\omega|^y, \quad \alpha_0 = \frac{\tau^{y-1}}{2c_0}\left|\cos\frac{\pi y}{2}\right|. \tag{17}$$

This is similar to an expression found by Kelly and McGough.[14] The relationship between z_0 and y is

$$y = z_0 + 1. \tag{18}$$

As mentioned previously, experimental data suggest that the exponent y is in the range (1,2], i.e., $z_0 \in (0,1]$ which agrees with the initial boundaries foreseen for z_0.

The phase velocity can be found from the real part of the wave number in Eq. (15):

$$c_p = \frac{\omega}{k_{\Re}} \approx c_0\left(1 + \frac{(\omega\tau)^{z_0}}{2}\cos\frac{\pi z_0}{2}\right). \tag{19}$$

The velocity dispersion can be found by subtracting c_0:

$$\Delta c_p \approx \frac{c_0}{2}\tau^{y-1}\sin\frac{\pi y}{2}\omega^{y-1} = c_0^2\alpha_0\tan\frac{\pi y}{2}\omega^{y-1}. \tag{20}$$

This result is the same as Szabo's equation (18b) which he presented for a medium which obeys causality.[18] It shows that the velocity starts from c_0 at very low frequencies and increases with frequency. It also shows that there is no velocity dispersion for $y=2$ as expected.

Numerical simulation of wave propagation in the low $\omega\tau$ case has been developed by Wismer.[13] She used a spatial finite element method and a temporal finite difference method for the fractional operator. This paper also shows several examples of interesting simulations of viscous attenuation with $1 < y \leq 2$.

B. High $\omega\tau$ approximation, $y=1-z_0/2$

In the other case, $(\omega\tau)^{z_0} \gg 1$, viscosity will dominate over elasticity. This covers shear waves in tissue above about 10 Hz, i.e., the interesting range up to 500 Hz or so used in various forms of dynamic elastography or rheometric measurements. In this case, the 1 in the numerator of Eq. (14) is small compared to the other terms and can be neglected providing the following approximation:

$$k \approx \frac{\omega}{c_0}(i\omega\tau)^{-z_0/2} = \frac{\tau^{-z_0/2}}{c_0}\left(\cos\frac{\pi z_0}{4} - i\sin\frac{\pi z_0}{4}\right)\omega^{1-z_0/2}. \tag{21}$$

Again it can be seen that the dispersion term, the first term, and the attenuation term, the second one, form a Hilbert transform pair.[18] The attenuation term in this case is

$$\alpha = \alpha_0 |\omega|^{1-z_0/2}, \quad \alpha_0 = \frac{\tau^{-z_0/2}}{c_0}\sin\frac{\pi z_0}{4}. \tag{22}$$

Expressed in the terminology of Eq. (1), it is

$$\alpha = \alpha_0 |\omega|^y, \quad \alpha_0 = \frac{\tau^{y-1}}{c_0}\cos\frac{\pi y}{2}. \tag{23}$$

Thus the relationship between the power-law exponent and the order of the fractional derivative is

$$y = 1 - z_0/2. \tag{24}$$

The range for z_0 in Ref. 13 is $0 < z_0 \leq 1$, implying that $y \in [0.5, 1)$. But there is nothing fundamental in the stress-strain relationship, Eq. (8), that should restrict the range to $z_0 \leq 1$, so we will let z_0 approach 2, implying that y can be in the range (0, 1). But, as shown by Knopoff,[19] an even z_0 implies that the stress-strain relationship becomes lossless, so $z=2$ should be avoided. Therefore, the maximum range is $0 < z_0 < 2$.

The phase velocity as a function of frequency can in this case also be found from the real part of the wave number, Eq. (21):

$$c_p = \frac{\omega}{k_{\Re}} = c_0\frac{\tau^{z_0/2}}{\cos\frac{\pi z_0}{4}}\omega^{z_0/2} = c_0\frac{\tau^{1-y}}{\sin\frac{\pi y}{2}}\omega^{1-y}. \tag{25}$$

C. Linearly increasing attenuation, $y=1$

The case for $y=1$ ($z_0=0$) is special in that both for the low and high $\omega\tau$ cases, the generalized Hooke's law, Eq. (8), degenerates to a simple linear stress-strain relation. Therefore all losses vanish, i.e., $\alpha_0 \to 0$ [Eqs. (17) and (23)]. However, losses described by $y=1$ are interesting for many materials, e.g., in seismics. One way to deal with it is to let $z_0 = \epsilon$ be very close to 0.

This was the approach taken by Buckingham[20] for modeling unconsolidated granular materials including marine sediments. He used a loss operator which was expressed as a convolution with a relaxation or memory function and which was justified by reference to earlier work on porous dissipative media:

$$Lu \propto \frac{\partial}{\partial t}(\nabla^2\{h \otimes u\}). \tag{26}$$

The relaxation function and its Fourier transform are

$$h(t) = \frac{t_0^{z_0-1}}{t^{z_0}}, \quad t > 0 \Leftrightarrow H(\omega) = \Gamma(1-z_0)(i\omega t_0)^{z_0-1}, \tag{27}$$

where Γ is the gamma-function and the Fourier transform is valid for $0 < z_0 < 1$. The transform is of the same form as Eq. (4), and this formulation is equivalent to Eq. (10) using the Riemann–Liouville fractional derivative.[10] This can be seen by Fourier transformation of the loss operator, Eq. (26), which gives $\Gamma(1-z_0)(i\omega t_0)^{z_0}k^2$. Except for a constant, this is the same as the loss term in Eq. (13). Szabo and Wu's[6] Hooke's law with a convolution operator is also related to this. Buckingham was interested in the case when z_0 approaches 0 in order to model attenuation with a power-law close to $y=1$ and used a series expansion in order to find the attenuation and the phase velocity. He also gave a numerical example with $z_0 = \epsilon = 0.05$ to illustrate this.

In a similar way, when the first term for the McLaurin series for the sine function is used in Eqs. (16) and (22), the same result is obtained for both the low-and high-frequency cases:

$$\alpha = \alpha_0 |\omega|^{1+\epsilon}, \quad \alpha_0 = \frac{\pi}{4c_0}\epsilon. \tag{28}$$

In this way attenuation with power-law behavior very close to $y=1$ can be described for the two cases.

D. Link to complex shear modulus

The frequency dependence of the shear modulus can be found from the ratio of stress and strain in Eq. (8) by inserting a harmonic plane wave, and where it is implicitly understood that the indices of c and η are $j=4$. The complex shear modulus is often denoted by $G^*=G_d+iG_l$ or $G'+iG''$ and is[5]

$$G^* = c + \eta(i\omega)^{z_0} = \mu + \eta\omega^{z_0}\cos\frac{\pi z_0}{2} + i\eta\omega^{z_0}\sin\frac{\pi z_0}{2}. \tag{29}$$

For the low $\omega\tau$ product approximation, this yields

$$G^* \approx \mu + i\eta\omega^{z_0}\sin\frac{\pi z_0}{2}, \tag{30}$$

and a ratio of imaginary and real parts which follows a power-law:

$$G_l/G_d \approx (\omega\tau)^{z_0}\sin\frac{\pi z_0}{2}. \tag{31}$$

For the high $\omega\tau$ product approximation, the complex shear modulus is

$$G^* \approx \eta\omega^{z_0}\cos\frac{\pi z_0}{2} + i\eta\omega^{z_0}\sin\frac{\pi z_0}{2}, \tag{32}$$

where both the real and the imaginary parts of the complex shear modulus have a power-law dependency of frequency with exponent z_0 and where their ratio only depends on the exponent, z_0:

$$G_l/G_d \approx \tan\frac{\pi z_0}{2}. \tag{33}$$

Thus the behavior of the real and the imaginary parts of the complex shear modulus and their ratio can help in deciding whether measured data belong in the low- or high-frequency regimes.

It should be noted that one can equally well write a similar dispersion relation as Eq. (29) for the compressional modulus. Therefore, z_0 is a universal scaling parameter governing the frequency dependence of the shear modulus as well as of the compressional modulus.

E. Potential relation to self-similar materials

It has not escaped our attention that the power-law model in the low $\omega\tau$ approximation case that we have found agrees with that recently found by Garnier and Sølna[21] for compressional waves in a random medium with long range correlation. There it was found that the power-law can be derived from a random medium where c_{11}, the bulk modulus, has a stochastic variation described by a parameter α which may be the same as our z_0. It is related to the Hurst parameter H of the underlying stochastic process as $\alpha=2-2H$ with $H \in (0.5,1)$, where the tail of the correlation function of the stochastic model decays as c_α/x^α for large lags x and where c_α is a positive constant. This kind of behavior of the correlation function implies that it is not integrable and is characteristic of self-similar or fractal processes.

If it turns out that the long range correlation property is equivalent to our low-frequency approximation, then this suggests a link to the stress-strain relationship, as the same parameter may describe the tail of the medium's correlation function, the stress-strain relationship, and the power-law of the attenuation.

V. FIT TO MEASURED DATA

Shear modulus data from cell rheology suggest values for $z_0 \approx 0.2$ valid over three orders of magnitude in the frequency range from 0.01 to 10 Hz.[22] Data for breast parenchyma obtained at about 80 Hz suggest values for $z_0 \approx 1.7$.[2] The low-frequency cut-off value c (or μ) is rather low, as can be seen from rheological brain data.[23] Thus, all those shear experiments typically fall into the high-frequency approximation of Eq. (14).

The shear wave attenuation data of Szabo and Wu for lexan, and low- and high-density polyethylene show y between 0.695 and 0.95 in the 2–8 MHz range.[6] This suggests that the high $\omega\tau$ approximation is valid, and that values for $z_0=2-2y$ are in the range from 0.1 to 0.61. On the other hand, some materials which support shear wave propagation may also be in the low $\omega\tau$ regime, e.g., yttrium indium garnet with $y \approx 2$ in the 100–1000 MHz range.[6]

We also analyzed data published by Asbach *et al.*[3] who did MR-based measurements of shear wave propagation in liver in the range from 22.5 to 67.5 Hz. We took the best data set shown in their Fig. 5, those for a healthy volunteer, and fitted power-law models. First such a model was fitted to their velocity data, and according to Eq. (25), we found that $z_0/2=0.18\pm0.07$. Second, we fitted a power-law to their attenuation data, and according to Eq. (1), the result was $y=0.73\pm0.07$. Thus within the uncertainty of the data, these exponents are in agreement with the relationship for the high-frequency approximation case, Eq. (24). The model presented here could therefore be an alternative to the standard linear solid model Asbach *et al.* had fitted, using three viscoelastic modeling parameters. Actually our model behaves better at low frequencies as it does not predict a non-zero velocity for 0 Hz.

Good data are hard to find in the megahertz range, but Gómez Álvarez-Arenas *et al.*[24] published attenuation and dispersion data for shear wave propagation in aerogels. They fitted a power-law over the range 0.44–1.2 MHz and found $y=0.5\pm0.15$. They then used the dispersion data to find the shear modulus as $\mu=\rho c_{0,44}^2$. The measured values for the real part of the shear modulus are plotted as stars in Fig. 1. Gómez Álvarez-Arenas *et al.* commented that the data do not fit simple and basic viscoelastic single-relaxation models (Voigt and Maxwell). In our framework, a value of y less than 1 suggests that the high-frequency regime applies and that Eq. (24) can be used to find a value of $z_0=1\mp0.3$. Indeed when these curves are plotted on top of the data in Fig.

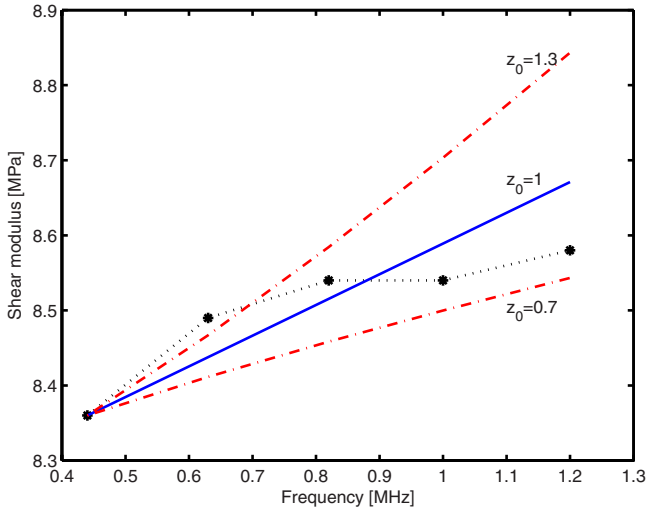


FIG. 1. (Color online) Power-law functions plotted over real part of measured shear modulus of aerogel from Ref. 24.

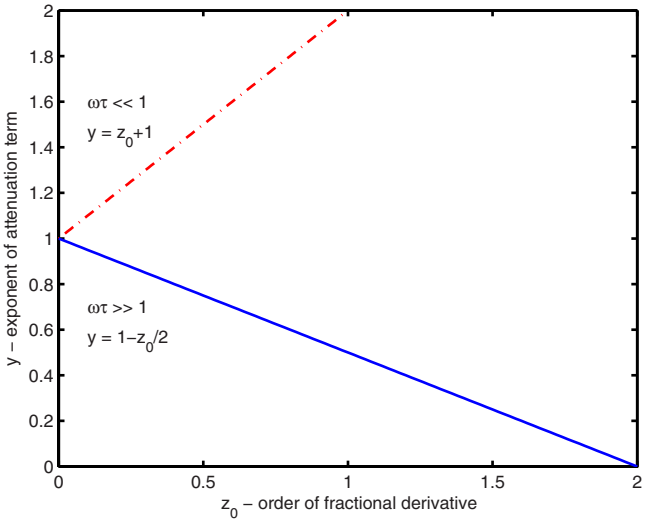


FIG. 2. (Color online) Power-law of attenuation as a function of power-law of viscoelastic term.

1, it is evident that within this rather restricted range of frequencies and within the uncertainty in y, the data seem to follow the power-law quite well. However, the actual variation in the data is very low compared to the absolute value, only 0.18 MPa for a shear modulus in the 8.5 MPa range. Therefore the fit will be very sensitive to measurement uncertainties, as indeed may be the case for the imaginary part which according to Eq. (32) should follow a similar power-law, but which it does not for this data set.

As noted, acoustic data from marine sediments have been modeled using special cases of the proposed model. The $z_0=1$ case for low and high $\omega\tau$ products can be used for saturated sand[8] and z_0 close to 0 and power-law $y\approx 1$ can be used for unconsolidated granular materials.[20]

VI. DISCUSSION

This analysis has been motivated by the observation that power-law behavior for the frequency dependence of viscoelastic parameters can be introduced by fractional derivatives. In order to remain invariant with respect to rotation and translation of the coordinate system, the fractional time derivative must be introduced according to Eq. (8). This leads to power-law dispersion relations for the complex mechanical moduli where the power-law exponent is equal to the order of the fractional derivative $z_0>0$. The obtained dispersion equations satisfy causality. Most experimental ultrasound related literature shows power-law exponents for the compressional wave attenuation $\alpha=\alpha_0|\omega|^y$ between $y\in[1,2]$ and establishes the following link between y and z_0: $y=1+z_0$. However, most experimental low-frequency shear wave data have attenuation with power-law exponents $y<1$, which are different from the previous findings for compressional waves if we exclude the possibility of fractional derivatives of negative order.

This dilemma is made even stronger by the observation that Eq. (29) holds equally for the compressional as well as for the shear modulus. Thus, how come that different intervals for y are observed when studying the attenuation properties of compressional waves in the megahertz regime and of shear waves in the hertz regime?

The solution to unify both worlds is by starting from the Caputo equation (10). Since it is derived from physics, it automatically gives a causal solution, and its dispersion equation satisfies the Kramers–Kronig relations. Thus, there is no need to force the Kramers–Kronig relations afterward or to modify the descriptions to fit the Kramers–Kronig relations. In studying the exact dispersion relations, Eq. (14), particular care must be used when doing approximations concerning the $\omega\tau$-term. We found that depending on the magnitude of the $\omega\tau$-term, different relationships between z_0 and y hold (see Fig. 2).

- When considering compressional modulus properties in tissue, the $\omega\tau$-term is small compared to 1 due to the fact that tissue is quasi-incompressible leading to a compressional parameter value of the order of gigapascals; thus, here we must use the low-frequency approximation leading to the relationship $y=1+z_0$.
- When considering shear modulus properties in tissue, the $\omega\tau$-term is of the order or large compared to 1 due to the fact that the low-frequency cut-off value for the shear modulus is rather low (tenth of kilopascals); thus, here we must use the high-frequency approximation leading to the relationship $y=1-z_0/2$.

VII. CONCLUSION

We have derived a fractional wave equation which is based on a fractional stress-strain relationship. The order of the fractional operator in both the wave equation and the stress-strain relationship is z_0, where $0<z_0<2$. The wave equation gives a causal solution as it is founded in a constitutive equation. It also satisfies the requirement for invariance to rotation and translation as second order spatial derivatives are found in the loss operator.

The equation results in attenuation which increases with $|\omega|^y$ for both pressure and shear waves. In the low-frequency

case, the exponent is $y=z_0+1$ and falls between 1 and 2. This is typical for most cases of compressional waves as found in, e.g., tissue and also some cases of shear waves and agrees with observations.

In the high-frequency case, which often has been overlooked, the exponent is $y=1-z_0/2$ and falls in the range between 0 and 1. This agrees with what has been observed for shear waves in, e.g., dynamic elastography.

The fractional wave equation thus unifies two regimes that up to now have been described by partly different theories due to the different ranges of the power-law exponents.

ACKNOWLEDGMENTS

This work was done while the first author (S.H.) was on a sabbatical at Institute Langevin, ESPCI ParisTech, France. He wants to thank Professor Mathias Fink for the hospitality, Dr. Mickaël Tanter for interesting discussions, and Professor Daniel Royer for his lectures on elastic waves in solids. The authors also want to thank Professor Josselyn Garnier (Université Paris VII) and Professor Knut Sølna (University of California Irvine) for comments on the derivations of this paper.

[1]T. Szabo, "Time domain wave equations for lossy media obeying a frequency power law," J. Acoust. Soc. Am. **96**, 491–500 (1994).
[2]R. Sinkus, K. Siegmann, T. Xydeas, M. Tanter, C. Claussen, and M. Fink, "MR elastography of breast lesions: Understanding the solid/liquid duality can improve the specificity of contrast-enhanced MR mammography," Magn. Reson. Med. **58**, 1135–1144 (2007).
[3]P. Asbach, D. Klatt, U. Hamhaber, J. Braun, R. Somasundaram, B. Hamm, and I. Sack, "Assessment of liver viscoelasticity using multifrequency MR elastography," Magn. Reson. Med. **61**, 373–379 (2008).
[4]T. Deffieux, G. Montaldo, M. Tanter, and M. Fink, "Shear wave spectroscopy for in vivo quantification of human soft tissues visco-elasticity," IEEE Trans. Med. Imaging **28**, 313–322 (2009).
[5]M. Kiss, T. Varghese, and T. Hall, "Viscoelastic characterization of in vitro canine tissue," Phys. Med. Biol. **49**, 4207–4218 (2004).
[6]T. Szabo and J. Wu, "A model for longitudinal and shear wave propagation in viscoelastic media," J. Acoust. Soc. Am. **107**, 2437–2446 (2000).
[7]D. Royer and E. Dieulesaint, *Elastic Waves in Solids* (Springer, Berlin, 2000), Vol. **I**.
[8]J. M. Hovem and G. D. Ingram, "Viscous attenuation of sound in saturated sand," J. Acoust. Soc. Am. **66**, 1807–1812 (1979).
[9]L. D. Landau, L. P. Pitaevskii, E. M. Lifshitz, and A. M. Kosevich, *Theory of Elasticity*, Course of Theoretical Physics Vol. **7** (Butterworth-Heinemann, Burlington, MA, 1986).
[10]W. Chen and S. Holm, "Modified Szabo's wave equation models for lossy media obeying frequency power law," J. Acoust. Soc. Am. **114**, 2570–2574 (2003).
[11]M. Caputo, "Linear models of dissipation whose Q is almost frequency independent-II," Geophys J. R. Astron. Soc. **13**, 529–539 (1967).
[12]W. Chen and S. Holm, "Fractional Laplacian time-space models for linear and nonlinear lossy media exhibiting arbitrary frequency power-law dependency," J. Acoust. Soc. Am. **115**, 1424–1430 (2004).
[13]M. Wismer, "Finite element analysis of broadband acoustic pulses through inhomogenous media with power law attenuation," J. Acoust. Soc. Am. **120**, 3493–3502 (2006).
[14]J. F. Kelly and R. J. McGough, "Fractal ladder models and power law wave equations," J. Acoust. Soc. Am. **126**, 2072–2081 (2009).
[15]R. Bagley and P. Torvik, "A theoretical basis for the application of fractional calculus to viscoelasticity," J. Rheol. **27**, 201–210 (1983).
[16]H. Schiessel and A. Blumen, "Mesoscopic pictures of sol-gel transition: Ladder models and fractal networks," Macromolecules **28**, 4013–4019 (1995).
[17]L. Kinsler, A. Frey, A. Coppens, and J. Sanders, *Fundamentals of Acoustics* (Wiley, New York, 1982).
[18]T. Szabo, "Causal theories and data for acoustic attenuation obeying a frequency power law," J. Acoust. Soc. Am. **97**, 14–24 (1995).
[19]L. Knopoff, "On the dissipative viscoelastic constants of higher order," J. Acoust. Soc. Am. **26**, 183–186 (1954).
[20]M. Buckingham, "Theory of acoustic attenuation, dispersion, and pulse propagation in unconsolidated granular materials including marine sediments," J. Acoust. Soc. Am. **102**, 2579–2596 (1997).
[21]J. Garnier and K. Sølna, "Pulse propagation in random media with long-range correlation," Multiscale Model. Simul. **7**, 1302–1324 (2009).
[22]B. Fabry, G. Maksym, J. Butler, M. Glogauer, D. Navajas, N. Taback, E. Millet, and J. Fredberg, "Time scale and other invariants of integrative mechanical behavior in living cells," Phys. Rev. E **68**, 041914 (2003).
[23]S. Nicolle, M. Lounis, R. Willinger, and J. Palierne, "Shear linear behavior of brain tissue over a large frequency range," Biorheology **42**, 209–223 (2005).
[24]T. E. Gómez Álvarez-Arenas, F. R. Montero de Espinosa, M. Moner-Girona, E. Rodríguez, A. Roig, and E. Molins, "Viscoelasticity of silica aerogels at ultrasonic frequencies," Appl. Phys. Lett. **81**, 1198–1200 (2002).

Thermal effects generated by high-intensity focused ultrasound beams at normal incidence to a bone surface

Diane M. Nell and Matthew R. Myers
Center for Devices and Radiological Health, U.S. Food and Drug Administration, Building 62, 10902 New Hampshire Avenue, Silver Spring, Maryland 20993

(Received 26 March 2009; revised 1 October 2009; accepted 2 October 2009)

Experiments and computations were performed to study factors affecting thermal safety when high-intensity focused ultrasound (HIFU) beams are normally incident (i.e., beam axis normal to the interface) upon a bone/soft-tissue interface. In particular, the temperature rise and thermal dose were determined as a function of separation between the beam focus and the interface. Under conditions representative of clinical HIFU procedures, it was found that the thermal dose at the bone surface can exceed the threshold for necrosis even when the beam focus is more than 4 cm from the bone. Experiments showed that reflection of the HIFU beam from the bone back into the transducer introduced temperature fluctuations of as much as $\pm 15\%$ and may be an important consideration for safety analyses at sufficiently high acoustic power. The applicability of linear propagation models in predicting thermal dose near the interface was also addressed. Linear models, while underpredicting thermal dose at the focus, provided a conservative (slight overprediction) estimate of thermal dose at the bone surface. Finally, temperature rise due to absorption of shear waves generated by the HIFU beam in the bone was computed. Modeling shear-wave propagation in the thermal analysis showed that the predicted temperature rise off axis was as much as 30% higher when absorption of shear waves is included, indicating that enhanced heating due to shear-wave absorption is potentially important, even for normally incident HIFU beams. [DOI: 10.1121/1.3257547]

PACS number(s): 43.80.Gx, 43.35.Wa, 43.80.Sh [CCC] Pages: 549–559

I. INTRODUCTION

In clinical applications of high-intensity focused ultrasound (HIFU), the absorption of acoustic energy by bone is an important safety factor. The high acoustic absorption of bone can lead to large temperature rises near the surface of the bone and potential damage to nearby sensitive tissues. In HIFU therapy for uterine leimyoma, sciatic nerve damage has been reported due to elevated temperatures in the pelvic bone (Kim *et al.*, 2008; Ren *et al.*, 2009). Civale *et al.* (2006) noted that when the targeted organ in the HIFU procedure is the liver or kidneys, ultrasound absorption in the ribs can produce damage to the skin.

Mathematical models have proven useful for estimating the temperature increase arising from ultrasound absorption in bone, particularly in diagnostic imaging and hyperthermia applications. Many of the early models (Shaw, 1994; Carstensen *et al.*, 1990; Jago *et al.*, 1999) were based upon a point-source solution to the bioheat equation, as proposed by Nyborg (1988). The point-source approach is often utilized in an absorbing-disk model. A limitation of the point-source approach is that only one set of thermal properties can be used in the model. Doody *et al.* (2000) discussed the accuracy of heated-disk models in their comparison with finite-element simulations for diagnostic ultrasound intensities and durations.

Chan *et al.* (1973) considered a general heat source (applicable to acoustic or electromagnetic cases) that decayed exponentially with distance into the bone. They also considered multiple tissue layers and perfusion in their finite-difference computations of steady-state temperature. Wu and Du (1990) studied a focused Gaussian beam normally incident upon a bone/soft-tissue interface. They determined the dependence of the steady-state temperature upon operational parameters such as the amount of focusing and level of perfusion. Fujii *et al.* (1999) considered plane waves incident upon a bone/soft-tissue interface at various angles. They computed the steady-state temperature increase arising from absorption of both longitudinal and shear waves.

To analyze the safety issues associated with relatively short-duration procedures such as radiation force imaging, Herman and Harris (2002) and Herman and Myers (2003) derived transient models that quantify safety in terms of thermal dose (Sapareto and Dewey, 1984). Myers (2004, 2005) analyzed the effect of pulse characteristics on thermal effects for plane-wave sonication at a bone/soft-tissue interface.

Experimentally, O'Neill *et al.* (1994) examined ultrasound heating in a reusable bone/tissue phantom. They validated the output display standard used to estimate steady temperature rise at a bone/soft-tissue interface. Duckett *et al.* (2004) used phantoms containing mouse skulls to examine thermal effects of diagnostic beams operating at very high (40 MHz) frequencies. Nahirnyak *et al.* (2007) measured temperature rise in a human skull fragment under low exposure levels. In these three experimental studies just cited, the primary interest was the long-term temperature rise, occurring over hundreds of seconds. To the authors' knowledge, phantom studies involving HIFU procedures targeted near a bone surface have not been performed. Such procedures involve large temperature rises (tens of degrees) in a short time interval (tens of seconds).

This paper describes a combined experimental/computational study that investigated HIFU beam sonication near a bone/soft-tissue interface. An important motivation for the study was the need to quantify safety issues as a function

of separation ("standoff distance") between the HIFU focus and the bone. To pursue this objective, the experiments were performed to validate the computational model, which was then used to compute thermal effects as a function of standoff distance.

The experiments incorporated bovine cortical bone pieces embedded in a tissue-mimicking material (TMM) used in previous phantom studies (King *et al.*, 2007; Hariharan *et al.*, 2007). A HIFU transducer was used to expose the bone sections to beams directed at normal incidence to the bone surface. The surface of the bone was instrumented with thermocouples to measure the ensuing temperature rise. The position of the beam focus was adjusted to lie inside each material, as well as at the interface. In this way, temperature rise could be measured as the standoff was varied.

Finite-element simulations of the HIFU beam propagation and tissue temperature rise were performed to compare with thermocouple measurements. Following the validation of the computational model, numerical calculations were used to investigate a variety of safety questions relevant to HIFU procedures. As noted above, the most important is the focal-zone standoff at which adverse thermal effects occur; another is when linear models can be safely used to estimate potential damage near the bone-tissue interface. An additional question addressed in the computations is whether shear-wave propagation needs to be accounted for in safety predictions, for while the beam axis may be normal to the bone-tissue interface, the finite width of the beam and the focusing effect lead to shear-wave generation in the bone. The effect of these shear waves are accounted for in the experiments and computations.

II. METHODS

A. HIFU apparatus and procedure

Bovine bone samples (3) obtained from a pet store (Petsmart Inc., Silver Spring/Burtonsville, MD) were used in the experiments; the bones were obtained pre-cleaned and sterilized (not fresh). The thermal conductivity of the bone was measured using a thermal property analyzer (KD2 Thermal Properties Analyzer, Decagon Devices, Inc., Pullman, WA) employing a thin probe that is heated during the measurement procedure. To facilitate measurement, a hole just large enough to insert the probe was bored out of the cortical section (along the long-axis) of the bone. To simulate the HIFU testing conditions, the bone samples were submerged in water for several days prior to testing, and all air bubbles were removed from the test holes. [Davidson and James (2000) reported substantial differences in conductivities between wet and dry bones.] During testing, the bone samples remained submerged, and the probe was inserted into the hole. The resulting value for the thermal conductivity was 0.49 W/m °C; this is in reasonable agreement with Davidson and James (2000), who reported a conductivity of 0.56 W/m °C for bovine cortical bone. The longitudinal wave speed of the bone was also measured in-house using a time-delay spectroscopy method and two bone samples of two thicknesses. Other properties of the bone (used in the computations below) were taken to be published values (Duck, 1990; Goss *et al.*, 1978; ICRU, 1998). Properties of the bone are summarized in Table I.

TABLE I. Properties of bone (cortical; water-saturated).

Density (kg/m^3)	2025
Longitudinal wave speed (m/s)	3736
Shear-wave speed (m/s)	1995 [average of minimum and maximum values reported in Goss *et al.* (1978)]
Longitudinal wavelength at 1.105 MHz (mm)	3.4
Shear wavelength at 1.105 MHz (mm)	1.8
Longitudinal attenuation coefficient (dB/cm MHz)	6.9 (Duck, 1990)
Shear attenuation coefficient (dB/cm MHz)	10.0 [roughly 1.5 times longitudinal attenuation, per Chan *et al.* (1974), Goss *et al.* (1980), and Haken *et al.* (1992)]
B/A	N/a (modeled as linear)
Thermal conductivity (W/m °C)	0.487+−0.005
Specific heat (J/kg °C)	1300 (ICRU, 1998)
Thermal diffusivity (mm^2/s)	0.186

The soft-tissue material was developed previously by King *et al.* (2007). The physical properties, closely matching those of human muscles, are shown in Table II. Most of the properties of the TMM were measured in our laboratory using the procedures described by King *et al.* The TMM is formed in a liquid state for subsequent pouring into a fixture containing the cortical bone pieces.

Bone pieces in the form of flat bricks were cut from additional cortical bovine bone samples (purchased as previously described) and machined to a roughly 2×2 cm^2 cross section and 0.7 cm thickness. Two bone/soft-tissue phantoms were prepared by first gluing the bone pieces to the back face of the cylindrical fixture (acrylonitrile butadiene styrene). On the unglued surface, a grid was marked off in 2-mm increments, and eight 125-μm diameter thermocouples [T-type, copper-constantan thermocouples from Omega Engineering, Stamford, CT; Fig. 1(a)] were taped to the piece of cortical bone. A phantom with the thermocouples attached is shown prior to the pouring of the TMM in Fig. 1(b). From the standpoint of thermocouple artifact (discussed further below), smaller thermocouples are preferable. However, multiple attempts with smaller thermocouples resulted in breakage of one or more thermocouples, and a more robust approach was sought.

TABLE II. Properties of TMM.

Density (kg/m^3)	1000
Longitudinal wave speed (m/s)	1561 ± 22
Wavelength at 1.105 MHz (mm)	1.4
Attenuation coefficient	$0.53f^{1.04}\pm0.024f^{0.009}$ dB/cm MHz
B/A	8
Thermal conductivity (W/m °C)	0.58 ± 0.03
Specific heat (J/kg °C)	3660 ± 10
Thermal diffusivity (mm^2/s)	0.15 ± 0.015

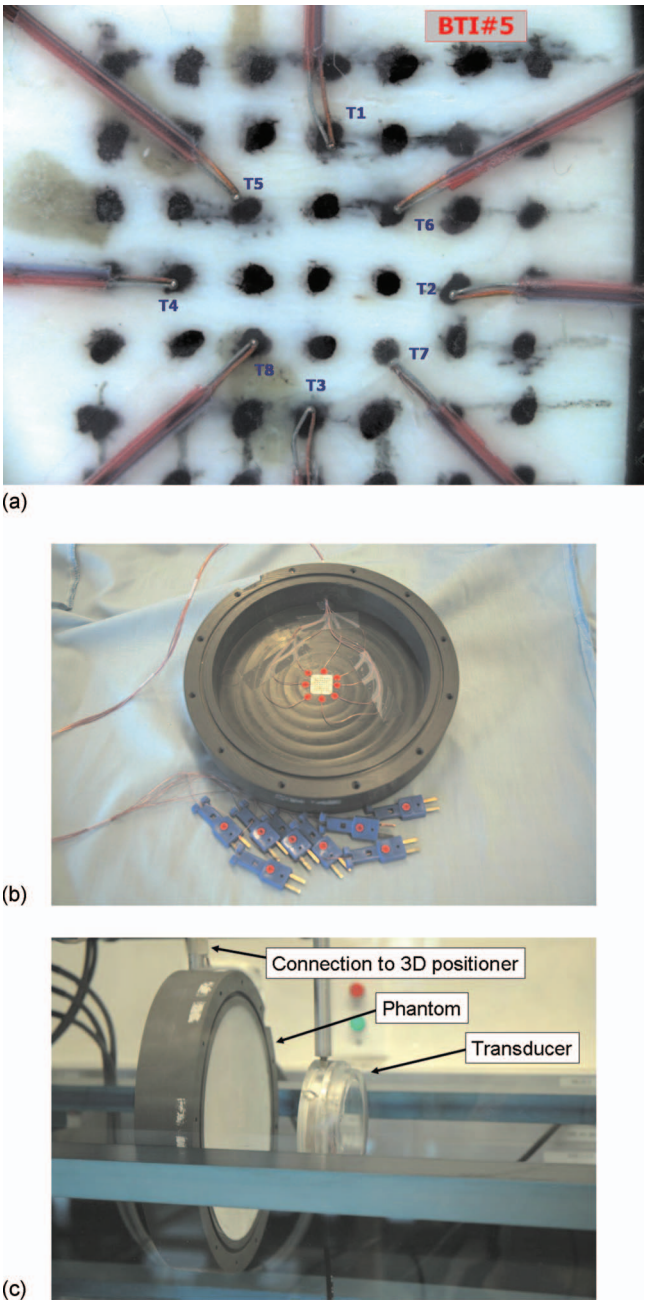


FIG. 1. (a) Cortical bone specimen with grid points marked and thermocouples attached at selected locations on the grid. (b) Cylindrical fixture containing the bone specimen and thermocouples in (a). (c) Bone/soft-tissue phantom mounted in tank coaxially with HIFU transducer. Vertical rods on the phantom and transducer are connected to an *x*-*y*-*z* positioning system.

The HIFU transducer used for sonicating the phantom is a 1.1 MHz, 8 cm-diameter transducer, with a 10 cm focal length. The lateral beamwidth (−6 dB) at the focus is 2.7 mm, and the axial beam length (−6 dB) is approximately 30 mm. The transducer was driven by an Agilent 33250A wave form generator and amplified with an Amplifier Research 150A-100B amplifier. Power output was computed from voltages across the transducer (using calibration data from a radiation force balance; Maruvada *et al.*, 2007) monitored on a 100 MHz oscilloscope (Agilent 54622A). The transducer and phantom were located in a tank of water, and the position of the transducer varied using an *x*-*y*-*z* positioner (Acoustic Intensity Measurement System, NTR Systems Inc., Seattle, WA). The HIFU transducer and bone/soft-tissue phantom are shown in Fig. 1(c).

Sonication periods between 10 and 15 s were used in the experiments. These exposure times are representative of HIFU tumor-ablation procedures. Cooling was observed for about 30 s, at which time temperatures had reduced to a few degrees above ambient.

Some of the eight thermocouples were positioned to measure temperature for oblique-incidence sonication and were not used in this study. Also, using all thermocouples reduced the sampling rate for each thermocouple. Sampling was performed with the use of a data acquisition system (OMB-DAQ-56 by Omega Engineering, Inc.) connected to a desktop computer. Depending on the experiment, either one or three thermocouples were used, with a sampling rate of 0.134 or 0.348 s, respectively.

An acoustic power of 1 W was used in most of the experiments. This level of power produced a temperature rise well above the noise level—around 10°, but low enough that cooling times prior to subsequent sonications were not excessive. In cases where measurable temperature rise was desired on thermocouples other than the target thermocouple, a higher power of 5 W was used.

In order to establish the position of the target thermocouple, a scanning procedure was employed. The transducer was moved in 0.1 mm increments using the three-dimensional positioning system, and at each location temperature rise during a 10-s sonication was recorded. The location showing the largest temperature increase over the 10-s time interval was taken to be the thermocouple position.

B. Numerical simulations

Simulations of the bone/soft-tissue sonication experiments were performed using the PZFLEX (Weidlinger Associates, Mountain View, CA) finite-element commercial software. Prior to use in the bone/soft-tissue study, PZFLEX was compared against a finite-difference solution of the KZK equation, as implemented in the software HIFU_Simulator (Soneson, 2009). The medium for the comparison was soft tissue having the properties of the TMM used in the experiments (Table II). Focal pressure values agreed to within 4%, and temperature rise after 30 s of sonication agreed within 8%.

For the bone/soft-tissue model, the rotational symmetry of the transducer and the large cross-sectional area of the bone (relative to the distance heat diffused in the time of interest) allowed for an axisymmetric model. A uniform mesh consisting of 40 elements per wavelength was used in the simulations. A mesh-dependency study showed that sparser grids were unable to accurately resolve the interference of the incident field and the reflected field near the bone/soft-tissue interface. A denser grid (e.g., 50 elements/wavelength) produced little change in the solution but increased computation time. With the 40 element/wavelength mesh density, the model required roughly 1–3 h run-times (depending on standoff) on an 8 Gbyte random access memory, 4 cpu desktop computer.

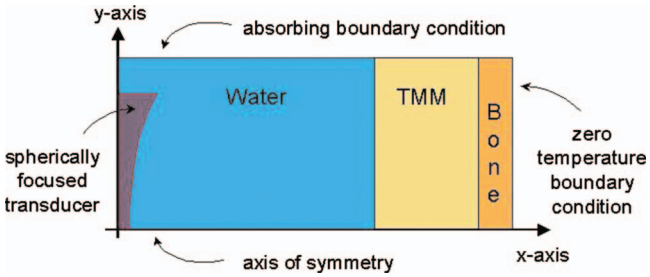


FIG. 2. Computational geometry and boundary conditions.

At the edges of the computational model, absorbing boundary conditions for the acoustic field and zero temperature-rise conditions for the thermal field were employed. To motivate the zero temperature-rise condition, we consider the back face of the model, which was taken to be 1.5 cm behind the interface (bone thickness of 1.5 cm) in the computations. If the length associated with thermal diffusion is quantified by $(\kappa t)^{1/2}$, where κ is the thermal diffusivity and t is the total time (heating and cooling) of interest, then for a diffusivity of 0.186 mm^2/s (Table I) and a time of 60 s, the thermal length is about 3 mm. The end of the computational boundary is thus well beyond the point where heat generated at the interface can diffuse during the time of interest. Also, for an attenuation in bone of 6.9 dB/cm, the beam power has decayed to about 9% of the interface value at the end of the computational domain, and little local heat is generated. The computational geometry and boundary conditions are provided in Fig. 2.

Material properties used in the computational model simulated those of the experimental model. As noted above, some of the properties of the TMM and bone were measured. Otherwise, published values were used. The material properties employed in the simulations are contained in Tables I and II.

The transient temperature field emerging from the solution of the heat-conduction equation was used in thermal-dose calculations. Thermal dose was computed using the method developed by Sapareto and Dewey (1984). This method was based upon cell experiments but has been validated in animal models (O'Brien *et al.*, 2008). In the thermal-dose formulation of Sapareto and Dewey, an exponential relation exists between the tissue temperature and the exposure time required to necrose cells. Thermal dose is quantified in terms of the number of equivalent minutes of heating at 43 °C,

$$t_{43}(x,y,z) = \int_{t=0}^{t=t_{\text{final}}} R^{43-T(t)} dt,$$

where t_{final} is the treatment duration, $T(t)$ is the temperature (in °C) as a function of time, and

$$R = \begin{cases} 0.5 & \text{if } T(t) \geq 43 \text{ °C} \\ 0.25 & \text{otherwise} \end{cases}.$$

According to this relation, thermal dose resulting from heating the tissue to 43 °C for 240 min is equivalent to that achieved by heating to 56 °C for 1.75 s. The threshold for cell necrosis was taken to be 240 equivalent minutes at 43 °C (Sapareto and Dewey, 1984), though this threshold is actually different for different organs.

III. RESULTS

The standoff distance mentioned in the following sections is the distance from the bone/soft-tissue interface to the focal plane. Negative standoffs imply distances into the bone medium; positive distances extend into the soft tissue. Hence, increasing the standoff implies backing the focus away from the bone/soft-tissue interface (i.e., the focus becomes increasingly proximal to the bone/soft-tissue interface).

The acoustic power used in the experiments was 1 W unless otherwise stated. Similarly, the experimental sonication time was 10 s if not specified. A wider variety of powers and sonication times was taken up in the computations to reflect a wider variety of clinical conditions.

A. Thermocouple artifact

Prior to presenting experimental results, we estimate the extent to which our temperature measurements were affected by thermocouple artifact. Two types of thermocouple artifact were considered potentially significant in our temperature measurements. The first is viscous heating, arising from the relative motion of the thermocouple and the TMM. The second is conduction of heat away from the volume of interest by the highly conducting copper wire. We first address viscous heating.

To estimate the contribution of viscous heating in the experimentally measured temperature rises, the temperature rises of a directly sonicated thermocouple (T5) and two neighboring thermocouples (T1 and T4) were experimentally measured and compared with computational predictions (thermocouples T5, T1, and T4 are shown in Fig. 1). Experiments were performed at a power of 5 W in order for temperatures at T1 and T4 to be elevated sufficiently above the noise level. The assumption was made that agreement between the experimental and computational traces on the remote thermocouples meant that the numerical model was simulating the propagation and heat transfer in the experiment with acceptable accuracy. Differences between the experimental and theoretical temperatures on the directly sonicated thermocouple could then be taken to be the artifact. The validity of this approach of employing remote thermocouples to estimate the contribution of viscous heating requires that the thermocouples be sufficiently distant to minimize the viscous heating effect relative to the absorptive heating. Morris *et al.* (2008) investigated the presence of viscous heating in thermocouples at various radial locations relative to the beam axis. For a thermocouple located on the beam axis, the ratio of viscous temperature rise to absorptive temperature rise (measured by a thin film thermocouple) was approximately 1.5 at the end of a 10-s sonication [Fig. 8 of Morris *et al.* (2008)]. At a radial location where the intensity was approximately 30% of the peak (on-axis) value, they found that the viscous temperature rise was approximately the same as the absorptive temperature rise (ratio ≈ 1). At a radial location where the intensity was approximately 15%

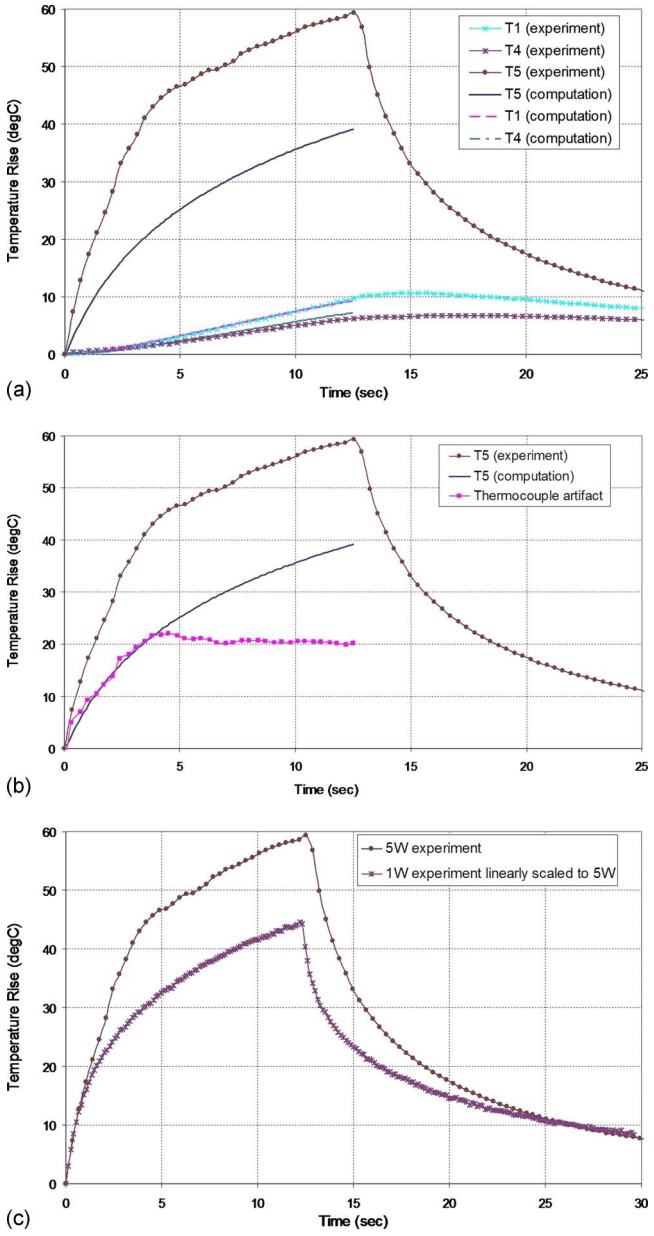


FIG. 3. (Color online) (a) Temperature traces on thermocouples T1, T4, and T5 for a sonication at 5 W power. Also shown are the finite-element predictions of temperature at the same locations. (b) Experimental temperature rise at thermocouple T5, computational temperature rise at T5, and the difference between the two, which is taken to be the thermocouple artifact. (c) Experimental temperature traces measured at T5 for a power of 5 W and for a power of 1 W scaled by a factor of 5.

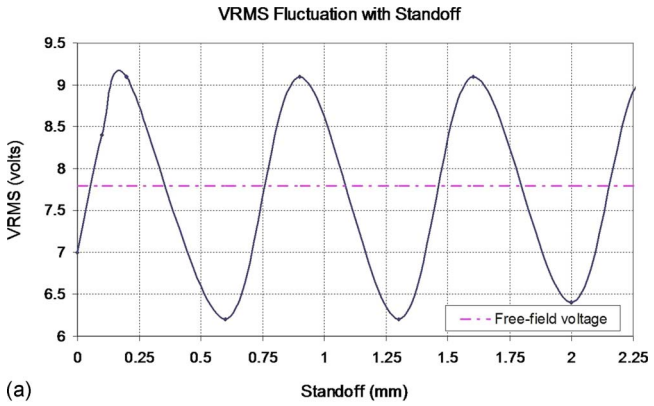
of the on-axis value, the ratio of viscous temperature rise to absorptive temperature rise was approximately 50%. The 15% relative intensity was the smallest considered by Morris *et al.* (2008). Our thermocouples T1 and T4 were located at radial locations corresponding to 0.6% and 1.5% of the on-axis intensity. Based upon an extrapolation of the results of Morris *et al.* to our locations, along with the agreement between experimental and computational temperatures at the remote locations shown subsequently, we conclude that viscous heating artifact at the remote locations can be neglected.

The experimentally measured temperature rises at the remote locations T1 and T4 are shown in Fig. 3(a), along with the temperature at the focus (T5). The temperature rises predicted by the finite-element computations are also shown. The bottom pairs of curves applied to T1 and T4 show reasonable overlap of the experimental and computational values. On the contrary, the top two curves in Fig. 3(a) (i.e., the experimentally measured and computationally predicted temperature rises at the focus; thermocouple T5) differ significantly due to the presence of viscous heating at T5. The difference between the experimental and theoretical temperatures, i.e., the viscous heating artifact, is shown in Fig. 3(b). The thermocouple artifact in this study was found to be roughly 30% of the measured temperature rise [Fig. 3(b)] at the end of the 13-s sonication at 5 W.

As a check on our approach to quantifying thermal artifact, a comparison was made to the theoretical study of viscous heating presented in Morris *et al.* (2008). Specifically, Morris *et al.* reported the viscous heating for a 120-μm diameter thermocouple in a beam having $1/e$ focal width of 1.2 mm (thermocouple diameter $\approx$6.2% of beam width). They found the ratio of viscous temperature rise to absorptive temperature rise to be about 0.76 at the end of 5 s of sonication. In comparison, we utilized a 125-μm wire in a beam of $1/e$ width equal to 1.4 mm (thermocouple diameter 4.5% of beam width). The ratio of viscous temperature rise to absorptive temperature rise for our situation was roughly 0.84 at the end of 5 s [Fig. 3(b)]. The thermal artifact quantified in this study seems to be comparable with that determined in Morris *et al.* (2008).

The second artifact considered was conduction artifact. To quantify the rate at which heat was conducted away from the tissue volume surrounding the thermocouples, we developed a model similar to that of Fry and Fry (1954). Only the copper wire, which has conductivity 17 times that of constantan (Dickinson, 1985), was considered. The copper wire was modeled as a cylinder having uniform temperature in the radial direction and a temperature gradient in the axial direction. The details of the model are provided in the Appendix. Results of the model show that the conduction artifact was approximately 18% at T1 and T4; i.e., the measured temperature was lowered by about 18% at the remote locations. As explained in the Appendix, the 18% value is likely an overestimate. At the focus, the conduction artifact was about 6%.

To summarize our findings on thermocouple artifact, it was determined that viscous heating tends to increase temperature readings at the target thermocouple T5 by approximately 30%, but viscous heating had little effect on the remote thermocouples T1 and T4. Conduction artifact tended to lower the readings on the remote thermocouples by about 18% and about 6% on the target thermocouple. Because quantitative predictions regarding HIFU safety featured later in the paper were made on the basis of computations, no effort was made to adjust temperature measurements to account for thermocouple artifact. Additionally, it was found that artifacts derived at 5 W power levels did not scale proportionately to other powers, e.g., the 1 W powers used in subsequent experiments. Figure 3(c) contains the temperature trace measured at 5 W, along with the trace measured at 1 W scaled by a factor of 5. A difference was found between these two traces, with the 5-W experiment having an apparent larger artifact than the 1 Watt experiment, indicating a



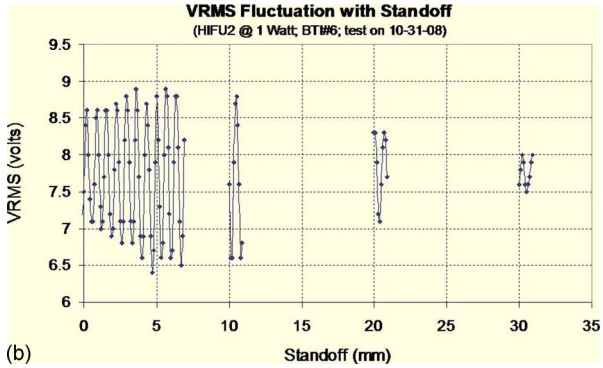


FIG. 4. (Color online) (a) Variation in transducer voltage as a function of standoff distance between the transducer and the bone surface. (b) Variation in transducer voltage as a function of standoff distance for expanded range of standoffs.

nonlinear or power-dependent artifact. Hence, rather than trying to determine an appropriate amount for quantitatively adjusting data at 1 W of power, we give estimates of the thermocouple artifact to help provide a measure of uncertainty in our temperature measurements.

B. Influence of back reflected field

A consideration in examining HIFU beams incident upon an interface (e.g., bone-tissue) is the effect of waves reflected back into the transducer. To observe the effect and significance of waves reflected off of the bone surface back into the transducer in these experimental studies, the voltage across the transducer was measured as the transducer was moved relative to the interface. The measured transducer voltage (rms) as a function of standoff distance is shown in Figs. 4(a) and 4(b) (two data sets recorded on two different days; Fig. 4(a) presents the data on a smaller scale to facilitate visualization of the fluctuations). For targeting near the interface (standoff near zero), the amplitude of the sinusoidal variation is approximately ±15%–20% of the voltage measured in the absence of the bone. The spatial scale characterizing the undulation is roughly 0.7 mm. The influence of the reflected field on the transducer output voltage is still near ±10% at a standoff of 2 cm and drops to approximately ±3% at a standoff of 3 cm [Fig. 4(b)].

The variation in temperature associated with the voltage fluctuations of Figs. 4(a) and 4(b) is given in Fig. 5. Here, the temperature at the bone surface is plotted as a function of

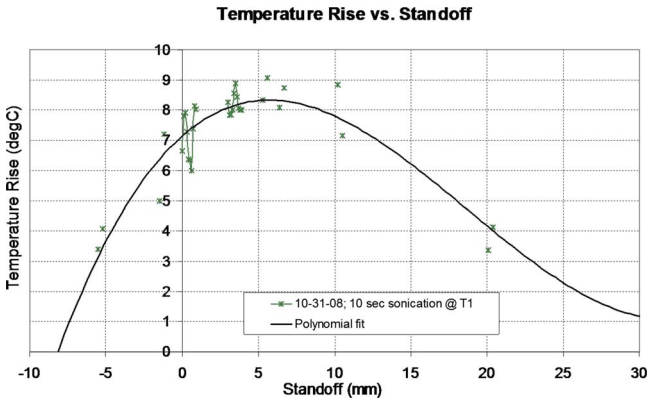


FIG. 5. (Color online) Temperature rise measured at the bone surface as a function of distance between bone surface and transducer. Acoustic power is 1 W. Solid line is a polynomial fit through experimental data.

standoff. The nonuniform range of standoff distances chosen for measurement was selected to explore the overall trend in temperature rise (as the distance from the transducer to interface was varied over multiple centimeters), as well as to resolve the rapid (sub-millimeter) variations occurring due to reflection from the interface back into the transducer. A fitting polynomial extending from a standoff distance of −8 to 30 mm represents an average of the small-scale temperature fluctuations arising from the reflected field. Data from only one trial (of 3) is featured in Fig. 5 in order to more clearly display the small-scale fluctuations associated with the reflected field; however, data from the other trials (presented in Sec. III C) are also well represented by the fitting curve, which provides an accurate mean over the small-scale fluctuations. The small-scale fluctuations are connected in two places by straight line segments, which facilitate visualization of the spatially fluctuating temperature relative to the spatial mean temperature. Temperature fluctuations were as much as ±15%.

C. Thermal effects as a function of standoff

Temperature rise was measured as a function of standoff on three different days. The reproducibility error for the three trials can be estimated by fitting curves through the three sets of experimental data, as was done in Fig. 5 for the 10-31-08 data set. Upon fitting the data to the curves (not shown) and computing the average distance between the curves at 5 mm standoff increments, we find the reproducibility error for the temperature measurement to be approximately 5%.

A comparison between the thermocouple measurements and finite-element predictions is provided in Fig. 6. Experimental data for all the three trials are shown. The computations generally underpredict the experimental temperatures, though we reiterate that the experimental values are not corrected for thermocouple artifact.

Extending the temperature predictions, the finite-element model was used to predict thermal dose as a function of axial distance. To simulate conditions of a uterine-fibroid ablation, we considered target locations well away from the interface (beyond 3 cm), powers of 100 and 200 W, and a sonication time of 30 s. In Fig. 7, thermal dose is plotted as a function of distance along the beam axis. The

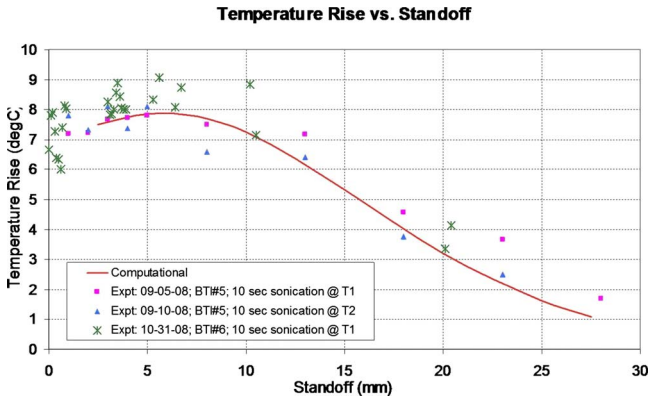


FIG. 6. (Color online) Temperature rise measured at the bone surface as a function of distance between bone surface and transducer for three different trials. Also shown is the computational prediction of temperature for the same standoffs.

transducer is located at the origin, the focal plane resides at a coordinate of 10 cm, and the bone surface is at an axial coordinate of 14 cm. The thermal doses are truncated at the 240 equivalent minute mark; any doses higher than this threshold were assumed to achieve cell necrosis. For the 200 W power, the threshold for necrosis is surpassed at the bone surface over an axial distance of about a quarter of 1 cm.

D. Linear vs nonlinear models

Continuing with thermal-dose calculations at clinically relevant powers, the thermal dose was computed as a function of standoff for 100 and 200 W powers, with a 30-s sonication time. In addition, for this set of computations, both linear and nonlinear propagation models were used to compute the acoustic field in the soft tissue. (Linear model was used in the bone in all cases.)

The thermal dose at the bone surface is shown in Fig. 8(a) for standoffs between 3 and 6 cm. The 240-min threshold for cell necrosis is represented by the horizontal line. For the 200-W exposure, the threshold is exceeded for standoffs less than about 5 cm. The thermal dose at the bone predicted by the linear model exceeds that of the nonlinear model by roughly a factor of 2. At the 100-W power, the thermal dose

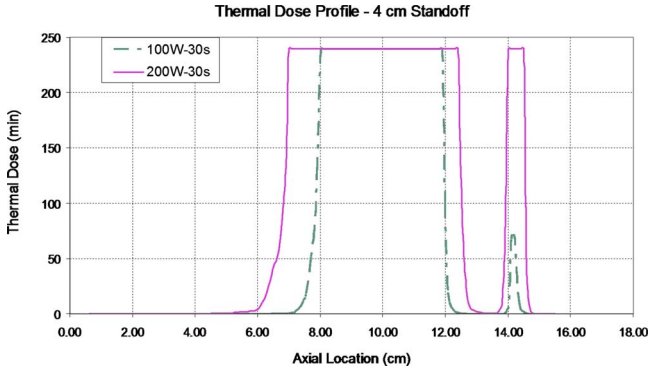


FIG. 7. (Color online) Computations of thermal dose as a function of axial position for 100 and 200 W powers. Transducer focus is located 4 cm from bone surface. Thermal-dose values above the threshold of 240 equivalent min are truncated at the 240-min value.

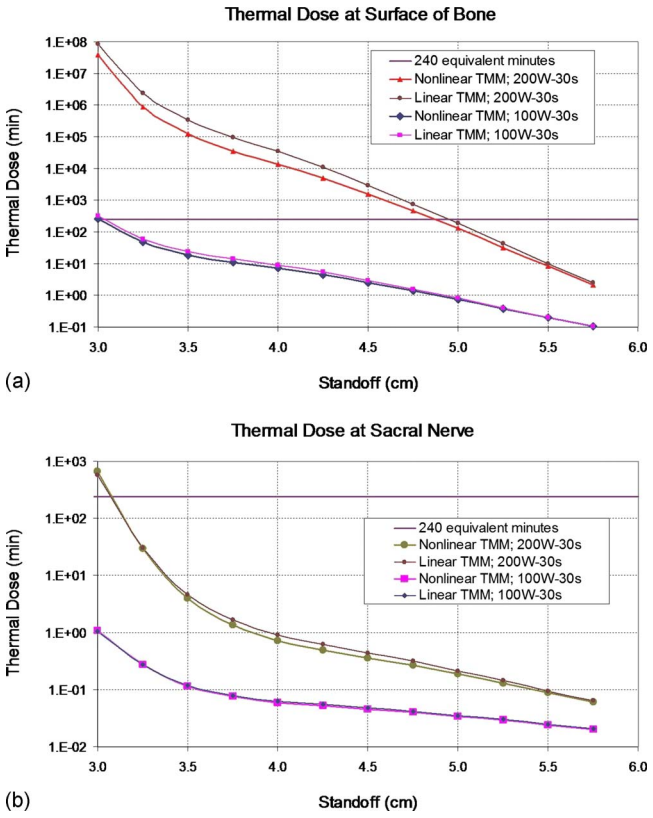


FIG. 8. (Color online) (a) Computations of thermal dose at the surface of the bone as a function of distance between the bone surface and the transducer for 100 and 200 W power levels. Sonication time is 30 s. Values predicted by both linear and nonlinear models are presented. (b) Computations of thermal dose at the sacral nerve as a function of distance between the bone surface and the transducer for 100 and 200 W power levels. Sonication time is 30 s. Values predicted by both linear and nonlinear models are presented.

is below the threshold for necrosis at all standoff values considered. The linear model provides very slightly higher values of thermal dose.

Thermal dose was also computed at a location 5 mm from the bone surface in the soft-tissue medium. The calculations at this location were intended to predict cell damage at a location representative of the sacral nerve due to ultrasound absorption within the sacral bone. In Fig. 8(b), it can be seen that the threshold for cell necrosis (240 equivalent min) is exceeded for the 200 W sonication when the standoff is less than about 31 mm. The linear model predicts a slightly higher thermal dose for standoffs around 35–50 mm, but the dose is below threshold. For the 100-W sonication, the thermal dose is below threshold for all standoffs considered (30 mm and greater). The linear and nonlinear predictions are virtually identical for the 100-W power.

E. Effect of shear waves

To examine the influence of shear-wave generation within the elastic bone, the temperature rise was computed by considering the bone as both a liquid medium (no-shear waves) and an elastic solid (shear waves supported). The properties of the elastic medium are contained in Table I. The density and longitudinal wave speed were kept the same (values in Table I) in the two models. The acoustic power was 5 W, and the beam focus was located at the bone surface

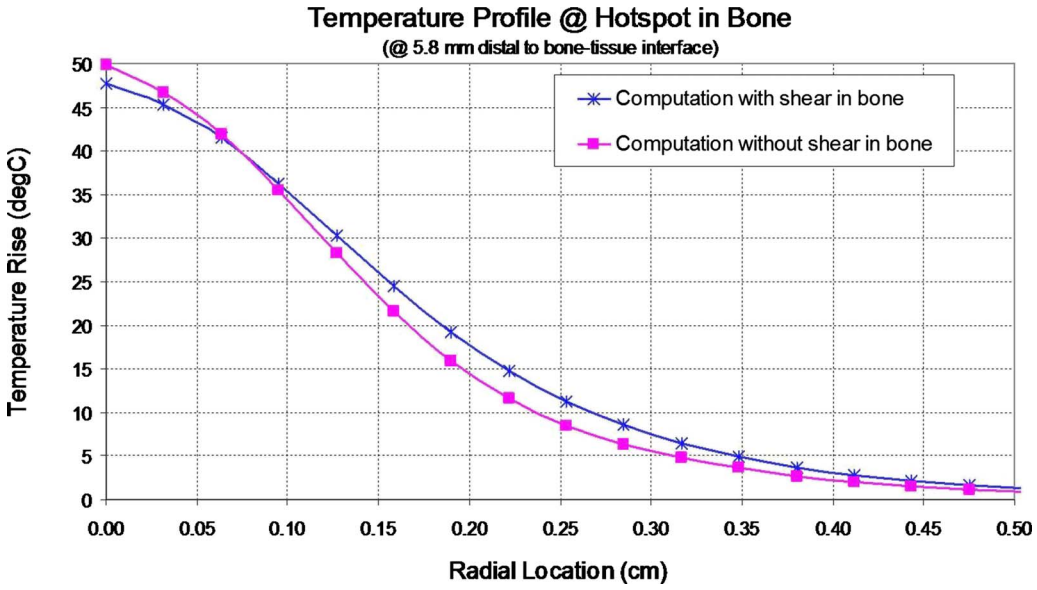


FIG. 9. (Color online) Computed temperature vs radial position, predicted by models accounting for shear waves in bone and ignoring shear waves. Axial location is point of maximum temperature rise in bone.

(zero standoff). Temperature rise as a function of radial position, at the axial coordinate corresponding to the highest temperature increase within the bone (roughly 6 mm within the bone surface), is shown in Fig. 9. The temperatures apply to the end of the sonication period (10 s). The no-shear bone model manifests a few-degree higher temperature rise on axis, but at increasing radial distances from the axis, the temperature rise is greater in the shear-supporting bone model. The largest discrepancy between the temperature rises accounting for shear and those neglecting it is about 30% (off axis).

IV. DISCUSSION

Viscous heating artifact measured in the 5 W experiment constituted a significant fraction of the measured temperature rise—about 30% after a 10 s sonication. Huang (2002) measured a similar level of artifact in experiments performed in soft-tissue phantoms. If we presume a power-independent artifact and reduce the 1 W experimental data (shown in Fig. 6) by 30%, the computational results will then exceed the artifact-adjusted experimental data by about 25% at target locations near the interface (standoffs less than 10 mm). (Without artifact adjustment, the 1 W experimental results slightly exceed the computational ones, as presented in Fig. 6.) At larger standoffs, the computational curve would fit the artifact-adjusted data more closely, less than 10% average discrepancy for standoffs greater than 15 mm. As noted in Sec. III, it is our sense that the 30% artifact determined at a power of 5 W is an overestimate at the 1 W power level. We note that Morris *et al.* (2008) found that the ratio of viscous heating to absorptive heating was roughly independent of power for the intensities studied. The propagation medium considered by Morris *et al.* (2008) is not the same as that of this study, owing to the presence of the bone. It is difficult for us to assess the influence of the bone on thermocouple artifact in any simple manner; further study is required.

Assuming adjustments (30% or less) in the experimental data for artifacts, the computational results then provide a usefully conservative measure of thermal effects from a safety standpoint. That is, the temperatures predicted by the computations predict the trends (e.g., with standoff) correctly and overestimate risk of thermal damage without grossly exaggerating it. With this in mind, we felt sufficiently confident in the computational model to employ it in further safety-related analyses.

The length scale of the undulations in the transducer voltage as a function of standoff [Figs. 4(a) and 4(b)] is consistent with a wave that has traversed the roundtrip distance between the transducer and the bone surface. The phase change of such a reflected wave is $2*k*L$, where k is the acoustic wave number and $2L$ is the roundtrip distance. The wavelength associated with the spatial variations as the standoff L is varied [Fig. 4(a)] is then half of the acoustic wavelength, or about 0.7 mm.

The substantial transducer voltage fluctuations due to back reflection [Figs. 4(a) and 4(b)] are evidence that not only estimates of the forward field, but the back reflected field as well, may be required for safety determinations. For a standoff distance of 1 cm, for example, the voltage fluctuations with distance are still on the order of $\pm 15\%$ [Fig. 4(b)]. This corresponds to a temperature variation of about $\pm 10\%$ at the location cited (standoff of 1 cm) based upon the results of the previous section (Fig. 5). If the temperature rise is 60° [e.g., Fig. 3(c) for 5 W power], this implies that transducer locations differing by only 0.375 mm [distance between maxima and minima of the fluctuations in Fig. 4(a)] will produce temperatures on the bone surface differing by 6 °C. In terms of thermal dose, this represents a difference of 2^6, or 64. Thus, for sufficiently high acoustic powers, back reflection may be important from a safety perspective.

The substantial temperature rises near the bone surface for very low power levels (e.g., 1–5 W; Figs. 3 and 6) emphasize the potential safety risk associated with HIFU pro-

cedures targeted even moderately close to bone. Figure 7, which relates to clinically relevant powers, illustrates that even when the HIFU target is 4 cm from the bone surface, a significant zone (multiple millimeters) exists where the threshold for necrosis is exceeded. The axial extent of the lesion (length along beam axis where thermal dose exceeds 240 equivalent min) within the bone is about 10% of the axial length of the lesion at the focus.

Figures 8(a) and 8(b) provide some insights for establishing safe standoff distances for HIFU procedures. For the 100 W power (30-s duration), a 3 cm standoff proved ample for avoiding cellular necrosis at a location representative of the sacral nerve, though barely sufficient for avoiding necrosis at the bone surface. For the 200 W power, cellular necrosis occurs at the sacral nerve when the standoff is less than about 3.1 cm. However, at the bone surface, safety is only achieved with a significantly greater standoff of 5 cm. Further computation is required to establish safety limits for other powers and sonication times, as well as other transducer designs.

For the conditions considered in this study, linear models provided a slight overestimate of the thermal dose delivered to the bone surface in the simulated HIFU procedures. This slightly conservative property is advantageous from a safety standpoint for it predicts safety risk without unnecessarily restricting the effectiveness of the procedure due to inflated risk. The greater heating at the bone surface predicted by the linear models (compared to nonlinear models) is accompanied by reduced heating at the focus; hence, linear predictions may not be desirable for estimating unwanted necrosis near the focus.

Accounting for shear-wave propagation in the bone yielded a slightly lower bone temperature on axis (Fig. 9), but a higher temperature at other radial locations, as well as a higher amount of absorbed energy overall. Hence, it appears that even at normal beam incidence, it may be important to account for shear-wave propagation in the determination of thermal effects near the bone/soft-tissue interface.

The ultrasound absorption in bone—6.9 dB/cm—used throughout this study is appropriate for the human spine (Duck, 1990), as well as for bovine cortical bone (Duck, 1990). However, this absorption is probably low for human cortical bone; published absorption values range from 14–22 dB/cm at 1 MHz (ICRU, 1998). Thermal effects (temperature rise, thermal dose, lesion size) at the bone surface are likely to be higher for absorption values characteristic of human cortical bone. This difference in absorption values highlights the need for calculations using values of tissue properties and transducer characteristics that are specific to the procedure of interest whenever analyses are performed to determine safety thresholds.

This study has concentrated on the safety issues associated with bone exposure during HIFU procedures. In addition to the safety perspective, HIFU absorption at a bone/soft-tissue interface is important from a therapeutic standpoint. HIFU therapy is currently being investigated for palliation of pain in patients with bone metastases (Catane *et al.*, 2007; Gianfelice *et al.*, 2008). In predicting the efficacy of such palliation procedures, it is important to quantify the

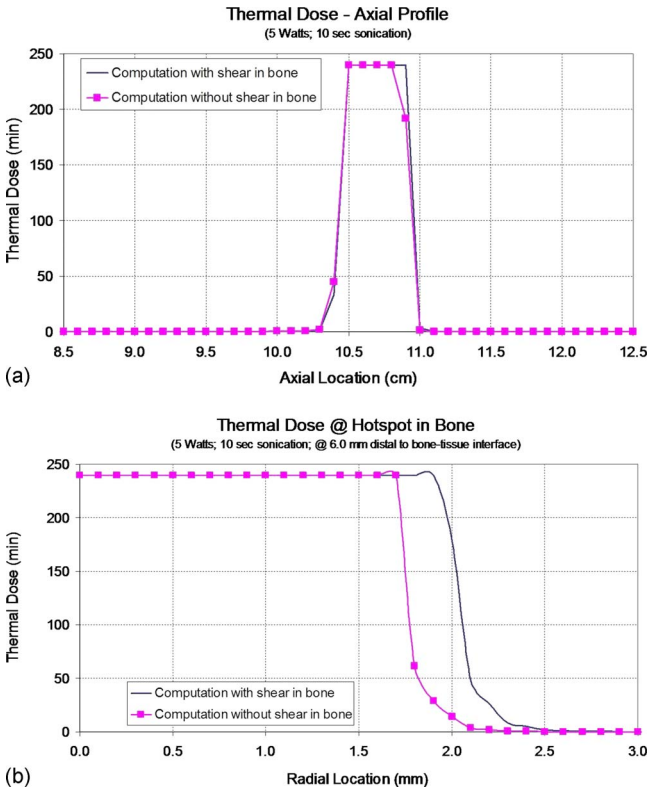


FIG. 10. (Color online) (a) Thermal dose as a function of axial position with and without the influence of shear waves. Exposure duration is 10 s, and power level is 5 W. (b) Thermal dose as a function of radial position with and without the influence of shear waves. Depth into bone is 6 mm. Exposure duration is 10 s, and power level is 5 W.

level of nerve-killing capacity arising from the absorbed ultrasound energy at a particular location. In Fig. 9, we see that a 50 °C temperature rise on axis was achievable with only 5 W of power and 10 s of sonication. The size of the lesion will be affected by shear-wave generation and absorption within the bone. Figure 10(a) shows the thermal dose as a function of axial position for the 10-s exposure at 5 W. The predicted bone-lesion (thermal dose in excess of 240 equivalent min) thickness is approximately 4 mm when shear waves are accounted for and 3 mm when shear is neglected. In the radial direction [Fig. 10(b)], the width of the lesion is about 1.9 mm when shear waves are included and 1.7 mm when they are neglected. This radial dependence of the thermal dose was computed at the axial location of highest temperature rise, roughly 6 mm into the bone.

A limitation of this study is the lack of a precise measure of thermocouple artifact. The large artifact is due in part to the use of relatively large (125-μm) thermocouples. To draw more exact conclusions on the basis of the experimental measurements, more precise estimates of thermocouple artifacts would be required. However, our primary objective was to provide sufficient empirical validation of the numerical model used to examine the various factors affecting HIFU safety. Given an agreement between empirical and numerical temperature values within 25% (depending upon the amount of artifact selected) and an accurate prediction by the computations of the trend in temperature rise as a function of standoff, it was felt that the numerical model was sufficiently validated.

Another limitation of the study is the abbreviated study of shear waves due to the restriction to normal beam incidence and the restriction to zero standoff. As the angle of incidence of the beam to the bone surface changes from 90° (arising from oblique incidence of the beam axis with the bone surface and also arising from nonzero standoffs at normal incidence), shear waves generated and absorbed in the bone play an increasingly important role. The interaction of bone with HIFU beams at oblique incidence and nonzero standoffs will be taken up in a subsequent study.

V. CONCLUSIONS

Experiments involving a HIFU beam targeted at various standoff distances from a bone surface showed that back reflection of the beam into the transducer, which alters the effective intensity relative to no reflection, may be an important consideration for safety analyses at sufficiently high acoustic power. It was also shown computationally that at clinically relevant power levels and exposure times, the thermal-dose threshold necessary for cell damage near the bone can be exceeded, even when the focus is 4 cm from the bone. Linear models provided conservative estimates of the thermal dose at the bone surface, even though they underpredicted the level of cell damage at the focus. Finally, the enhanced heating due to shear-wave absorption is potentially important, even for the normally incident HIFU beams.

ACKNOWLEDGMENTS

The authors are grateful to Yunbo Liu for measuring the properties of the bone and TMM. They also wish to thank Subha Maruvada, Gerald Harris, and Bruce Herman for helpful discussions on back reflection and thermocouple artifacts.

APPENDIX

To estimate the conduction artifact in the thermocouple measurements, we consider only the copper wire (Fig. 1), which has much higher conductivity than the constantan wire (Dickinson, 1985). We model the copper wire as a cylinder of radius r_0 and length L. Thermocouples T1, T4, and T5 will experience a temperature gradient along this length and conduct heat away from the measurement volume. For convenience in the following derivation, we consider the artifact at T1 during sonication at T5 (Fig. 1). Modifications for other situations will be described.

The heat per unit time conducted away from the measurement volume by the copper wire can be written as

$$k_c \frac{\partial T}{\partial x} \pi r_0^2, \tag{A1}$$

where k_c is the thermal conductivity of the copper wire and $\partial T/\partial x$ is the axial temperature gradient. We examine the temperature rise in the measurement volume at small times when the temperature rise is proportional to time, i.e.,

$$T \sim \frac{\alpha I t}{\rho_0 c_p}, \tag{A2}$$

where α, ρ_0, and c_p are the absorption coefficient, density, and specific heat of tissue.

We take the beam to be centered upon thermocouple T5 (Fig. 1). If r is the radial coordinate measured from T5, the distance x along the copper wire is $r\cos(\theta)$, θ being the angle between the copper wire and the radius connecting T5 and T1. After incorporation of Eq. (A2), the heat per unit time lost due to conduction along the copper wire becomes

$$k_c \pi r_0^2 \cos(\theta) \frac{\alpha t}{\rho_0 c_p} \frac{\partial I}{\partial r}. \tag{A3}$$

We wish to compare Eq. (A3) with the deposited energy affecting the temperature rise measured by T1. Following Fry and Fry (1954), if we surround the copper wire with a tissue-filled cylinder of radius r_1, where r_1 marks the boundary of the volume of energy affecting T4, then the deposited energy/time is

$$\alpha I \pi r_1^2 L. \tag{A4}$$

The ratio of the conducted heat to the total deposited heat is Eq. (A3) divided by Eq. (A4),

$$\frac{\partial I}{\partial r} \frac{k_c r_0^2 \cos(\theta) t}{\rho_0 c_p I r_1^2 L}. \tag{A5}$$

We model the beam in the focal plane with a Gaussian profile,

$$I = I_0 \exp(-r^2/b^2), \tag{A6}$$

where b is the $1/e$ beam width. Then,

$$\frac{\partial I}{\partial r} = \frac{-2r}{b^2} I, \tag{A7}$$

and Eq. (A5) becomes

$$\frac{2 r k_c r_0^2 \cos(\theta) t}{\rho_0 c_p r_1^2 L b^2}. \tag{A8}$$

An estimate for r_1 can be obtained by considering the model of a cylinder of radius r_1 with surface temperature V and zero initial temperature. The value of r_1 is chosen so that the axis of the cylinder is just inside the volume affected by diffusion of heat from the surface in time t. From Carslaw and Jaeger (1959), this occurs for

$$\frac{\kappa t}{r_1^2} = \frac{k_t t}{\rho_0 c_p r_1^2} \approx 0.04 \tag{A9}$$

or

$$r_1^2 = \frac{25 k_t t}{\rho_0 c_p}, \tag{A10}$$

where k_t is the conductivity of the tissue medium surrounding the thermocouple. Upon inserting Eq. (A10) into Eq. (A8), we find that the ratio of heat conducted away (per unit time) to the energy deposited in the volume affecting thermocouple T1 is

$$\frac{2 r k_c r_0^2 \cos(\theta)}{25 k_t L b^2}. \tag{A11}$$

The length L characterizes the axial scale over which the intensity field changes. We take L to be the same as the beam radius b of the transducer. From hydrophone measurements, the beam radius b of the transducer was found to be about 1.4 mm. For thermocouple T1, we take r to be the distance between the junctions of T5 and T1, roughly 2.8 mm in Fig. 1. Using $r_0=0.0625$ mm, $\cos(\theta)=0.7$, $k_c=385$ W/(m K) (Dickinson, 1985), and $k_t=0.487$ W/(m K) (Table I), Eq. (A11) yields a value of about 18%. That is, temperatures measured by thermocouple T1 are underestimates of the true temperature by about 18% due to heat conducted away by the copper wire. The 18% value is probably excessive since hydrophone measurements show that T1 and T4 reside near the first minimum in the radial intensity profile, where the derivative in Eq. (A5) is zero. The Gaussian model [Eq. (A6)] does not capture the minima and maxima in the intensity profile as a function of radial position.

Given the symmetry between T1 and T4 (Fig. 1), an 18% artifact also applies to T4 during the sonication at T5. At the target location T5, Eq. (A11) can again be used with slight modification. The value of $\cos(\theta)$ is 1 since the copper wire lies along the radial direction emanating from T5. An appropriate value of r must be selected; we take it to be half the length of the cylindrical section under consideration, i.e., $L/2=0.7$ mm. With the parameter values appropriate for T5, Eq. (A11) yields an artifact of 6%. That is, conduction artifact causes temperature values measured by the target thermocouple T5 to be reduced by about 6% relative to the true value.

Carslaw, H. S., and Jaeger, J. C. (**1959**). *Heat Conduction in Solids* (Oxford University Press, New York).

Carstensen, E. L., Child, S. Z., Norton, S., and Nyborg, W. (**1990**). "Ultrasonic heating of the skull," J. Acoust. Soc. Am. **87**, 1310–1317.

Catane, R., Beck, A., Inbar, Y., Rabin, T., Shabshin, N., Hengst, S., Pfeffer, R. M., Hanannel, A., Dogadkin, O., Liberman, B., and Kopelman, D. (**2007**). "MR-guided focused ultrasound surgery (MRgFUS) for the palliation of pain in patients with bone metastases—Preliminary clinical experience," Ann. Oncol. **18**, 163–167.

Chan, A. K., Sigelmann, R. A., and Guy, A. W. (**1974**). "Calculations of therapeutic heat generated by ultrasound in fat-muscle-bone layers," IEEE Trans. Biomed. Eng. **BME-21**, 280–284.

Chan, A. K., Sigelmann, R. A., Guy, A. W., and Lehmann, J. F. (**1973**). "Calculation by the method of finite differences of the temperature distribution in layered tissues," IEEE Trans. Biomed. Eng. **BME-20**, 86–90.

Civale, J., Clarke, R., Rivens, I., and ter Haar, G. (**2006**). "The use of a segmented transducer for rib sparing in HIFU treatments," Ultrasound Med. Biol. **32**, 1753–1761.

Davidson, S. R. H., and James, D. F. (**2000**). "Measurement of thermal conductivity of bovine cortical bone," Med. Eng. Phys. **22**, 741–747.

Dickinson, R. J. (**1985**). "Thermal conduction errors of manganin-constantan thermocouple arrays," Phys. Med. Biol. **30**, 445–453.

Doody, C., Duck, F. A., and Humphrey, V. F. (**2000**). "Comparison of finite-element and heated disk models of tissue heating by ultrasound," Ultrasound Med. Biol. **26**, 1347–1355.

Duck, F. A. (**1990**). *Physical Properties of Tissue* (Academic, New York).

Duckett, A. S., Reid, A. D., Leamen, L., Cucevic, V., and Foster, F. S. (**2004**). "Thermal assessment of 40-MHz ultrasound at soft tissue-bone interfaces," Ultrasound Med. Biol. **30**, 665–673.

Fry, W. J., and Fry, R. B. (**1954**). "Determination of absolute sound levels and acoustic absorption coefficients by thermocouple probes—Theory," J. Acoust. Soc. Am. **26**, 294–310.

Fujii, M., Sakamoto, K., Toda, Y., Negishi, A., and Kanai, H. (**1999**). "Study of the cause of the temperature rise at the muscle-bone interface during ultrasound hyperthermia," IEEE Trans. Biomed. Eng. **46**, 494–504.

Gianfelice, D., Gupta, C., Kucharczyk, W., Bret, P., Havill, D., and Clemons, M. (**2008**). "Palliative treatment of painful bone metastases with RM imaging-guided focused ultrasound," Radiology **249**, 355–363.

Goss, S. A., Johnston, R. L., and Dunn, F. (**1978**). "Comprehensive compilation of empirical ultrasonic properties of mammalian tissues," J. Acoust. Soc. Am. **64**, 423–457.

Goss, S. A., Johnston, R. L., and Dunn, F. (**1980**). "Compilation of empirical ultrasonic properties of mammalian tissues. II," J. Acoust. Soc. Am. **68**, 93–108.

Haken, B. A., Frizzell, A., and Carstensen, E. L. (**1992**). "Effect of mode conversion on ultrasonic heating at tissue interfaces," J. Ultrasound Med. **11**, 393–405.

Hariharan, P., Myers, M. R., and Banerjee, R. K. (**2007**). "HIFU procedures at moderate intensities—Effect of large blood vessels," Phys. Med. Biol. **52**, 3493–3513.

Herman, B. A., and Harris, G. R. (**2002**). "Models and regulatory considerations for transient temperature rise during diagnostic ultrasound pulses," Ultrasound Med. Biol. **28**, 1217–1224.

Herman, B. A., and Myers, M. R. (**2003**). "An analytic derivation for the transient temperature rise during an ultrasound pulse focused on bone," Ultrasound Med. Biol. **29**, 771–773.

Huang, J. (**2002**). "Heating in vascular tissue and flow-through tissue phantoms induced by focused ultrasound," Ph.D. dissertation, Boston University, Boston, MA.

ICRU (**1998**). "Tissue substitutes, phantoms and computational modelling in medical ultrasound," ICRU Report No. 61, International Commission on Radiation Units and Measurements, Bethesda, MD, December 31.

Jago, J. R., Henderson, J., Whittingham, T. A., and Mitchell, G. (**1999**). "A comparison of the AIUM/NEMA thermal indices with calculated temperature rises for a simple third-trimester pregnancy tissue model," Ultrasound Med. Biol. **25**, 623–628.

Kim, Y., Rhim, H., Choi, M. J., Lim, H. K., and Choi, D. (**2008**). "High-intensity focused ultrasound therapy: An overview for radiologists," Korean J. Radiol. **9**, 291–302.

King, R. L., Herman, B. A., Maruvada, S., Wear, K. A., and Harris, G. R. (**2007**). "Development of a HIFU phantom," *Proceedings of the Sixth International Symposium on Therapeutic Ultrasound* (American Institute of Physics, Melville, NY).

Maruvada, S., Harris, G. R., Herman, B. A., and King, R. (**2007**). "Acoustic power calibration of high-intensity focused ultrasound transducers using a radiation force technique," J. Acoust. Soc. Am. **121**, 1434–1439.

Morris, H., Rivens, I., Shaw, A., and ter Haar, G. (**2008**). "Investigation of the viscous heating artifact arising from the use of thermocouples in a focused ultrasound field," Phys. Med. Biol. **53**, 4759–4776.

Myers, M. R. (**2004**). "Transient temperature rise due to ultrasound absorption at a bone/soft-tissue interface," J. Acoust. Soc. Am. **115**, 2887–2891.

Myers, M. R. (**2005**). "Effect of pulse characteristics on temperature rise due to ultrasound absorption at a bone/soft-tissue interface," J. Acoust. Soc. Am. **117**, 3281–3287.

Nahirnyak, V., Mast, T. D., and Holland, C. K. (**2007**). "Ultrasound-induced thermal elevation in clotted blood and cranial bone," Ultrasound Med. Biol. **33**, 1285–1295.

Nyborg, W. L. (**1988**). "Solutions of the bio-heat transfer equation," Phys. Med. Biol. **33**, 785–792.

O'Brien, W. D., Deng, C. X., Harris, G. R., Herman, B. A., Merritt, C. R., Sanghvi, N., and Zachary, J. F. (**2008**). "The risk of exposure to diagnostic ultrasound in postnatal subjects—Thermal effects," J. Ultrasound Med. **27**, 517–535.

O'Neill, T. P., Winkler, A. J., and Wu, J. (**1994**). "Ultrasound heating in a tissue-bone phantom," Ultrasound Med. Biol. **20**, 579–588.

Ren, X.-L., Zhou, X.-D., Yan, R.-L., Liu, D., Zhang, J., He, G.-B., Han, Z.-H., Zheng, M.-J., and Yu, M. (**2009**). "Sonographically guided extracorporeal ablation of uterine fibroids with high-intensity focused ultrasound: Midterm results," J. Ultrasound Med. **28**, 95–103.

Sapareto, S. A., and Dewey, W. C. (**1984**). "Thermal dose determination in cancer therapy," Int. J. Radiat. Oncol. **10**, 787–800.

Shaw, A. (**1994**). "Prediction of temperature rise in layered media from measured ultrasonic intensity data," Phys. Med. Biol. **39**, 1203–1218.

Soneson, J. (**2009**). "A user-friendly software package for HIFU simulation," *Proceedings of the Eighth International Symposium on Therapeutic Ultrasound*, Minneapolis, MN, September 2008 (American Institute of Physics, Melville, NY).

Wu, J., and Du, G. (**1990**). "Temperature elevation generated by a focused Gaussian ultrasonic beam at a tissue-bone interface," J. Acoust. Soc. Am. **87**, 2748–2755.

Propagation of narrow-band-high-frequency clicks: Measured and modeled transmission loss of porpoise-like clicks in porpoise habitats

Stacy L. DeRuiter
IFREMER, Service Acoustique et Sismique, BP 70, 29280 Plouzané, France and Department of Biology, Woods Hole Oceanographic Institution (WHOI), MS 50, Woods Hole, Massachusetts 02543

Michael Hansen
Department of Biological Sciences, Zoophysiology, Aarhus University, C.F. Møllers Allé, Building 1131, DK-8000 Aarhus C, Denmark

Heather N. Koopman and Andrew J. Westgate
Department of Biology and Marine Biology, University of North Carolina Wilmington, Wilmington, North Carolina 28403-3297 and Grand Manan Whale and Seabird Research Station, 24 Route 776, Grand Manan, New Brunswick E5G 1A1, Canada

Peter L. Tyack
Department of Biology, Woods Hole Oceanographic Institution (WHOI), MS 50, Woods Hole, Massachusetts 02543

Peter T. Madsen
Department of Biology, Woods Hole Oceanographic Institution (WHOI), MS 50, Woods Hole, Massachusetts 02543 and Department of Biological Sciences, Zoophysiology, Aarhus University, C.F. Møllers Allé, Building 1131, DK-8000 Aarhus C, Denmark

(Received 16 July 2009; revised 11 September 2009; accepted 25 September 2009)

Estimating the range at which harbor porpoises can detect prey items and environmental objects is integral to understanding their biosonar. Understanding the ranges at which they can use echolocation to detect and avoid obstacles is particularly important for strategies to reduce bycatch. Transmission loss (TL) during acoustic propagation is an important determinant of those detection ranges, and it also influences animal detection functions used in passive acoustic monitoring. However, common assumptions regarding TL have rarely been tested. Here, TL of synthetic porpoise clicks was measured in porpoise habitats in Canada and Denmark, and field data were compared with spherical spreading law and ray-trace (Bellhop) model predictions. Both models matched mean observations quite well in most cases, indicating that a spherical spreading law can usually provide an accurate first-order estimate of TL for porpoise sounds in porpoise habitat. However, TL varied significantly (± 10 dB) between sites and over time in response to variability in seafloor characteristics, sound-speed profiles, and other short-timescale environmental fluctuations. Such variability should be taken into account in estimates of the ranges at which porpoises can communicate acoustically, detect echolocation targets, and be detected via passive acoustic monitoring. © *2010 Acoustical Society of America.* [DOI: 10.1121/1.3257203]

PACS number(s): 43.80.Ka, 43.30.Es, 43.80.Lb [WWA] Pages: 560–567

I. INTRODUCTION

Harbor porpoises (*Phocoena phocoena*) are small toothed whales that inhabit temperate and subarctic waters of the northern hemisphere, and like all toothed whale species investigated, they use echolocation for foraging and navigation. However, little is known about the echolocation strategies of porpoises in the wild (Au, 1993; Akamatsu *et al.*, 2005, 2007). Estimating the range at which porpoises can detect prey items and other environmental landmarks is an integral part of studies of harbor porpoise biosonar. It is also a key to understanding obstacle detection and avoidance behavior relevant to fishery bycatch reduction strategies (Au and Jones, 1991; Kastelein *et al.*, 2000; Mooney *et al.*, 2004, 2007). To predict the maximum range at which an object can be detected using echolocation, one must measure or estimate click source level, minimum detectable received echo level (in noise), target strength of the prey item or other target, and propagation loss between the echolocating animal and the target. Such estimates have been published for species including harbor porpoises (Au *et al.*, 2007; Mooney *et al.*, 2007; Villadsgaard *et al.*, 2007), bottlenose dolphins (Au *et al.*, 2007), orcas (Au *et al.*, 2004), false killer whales and Risso's dolphins (Madsen *et al.*, 2004), and Hector's and hourglass dolphins (Kyhn *et al.*, 2009). To estimate transmission loss (TL), all of those studies used a spherical spreading law with a frequency dependent attenuation factor. Because toothed whale echolocation clicks are generally short in duration and emitted in a narrow beam (Au, 1993), temporal overlap of surface- and bottom-reflected arrivals will be rare

except at longer ranges in shallower habitats, and the spherical spreading assumption should hold unless significant water-column refraction occurs. However, this common assumption is not normally validated empirically (Au, 1993; Madsen and Wahlberg, 2007; but see Miksis-Olds and Miller, 2006; Villadsgaard *et al.*, 2007). Testing this assumption and quantifying porpoise-habitat-specific and temporal variability in transmission loss could thus provide valuable data on the accuracy of detection-range estimates. It would also provide insight into the biophysical basis of porpoise acoustic ecology in the context of foraging, predator avoidance, and social communication.

Data on transmission loss are not only important for studies of porpoise biosonar and communication. Passive acoustic monitoring (PAM) with automatic click detection devices such as T-POD porpoise detectors (Thomsen *et al.*, 2005) has become an increasingly common method for monitoring the presence and abundance of toothed whales, especially harbor porpoises; T-PODs have been used both to study habitat utilization patterns (Carlström, 2005; Philpott *et al.*, 2007; Verfuss *et al.*, 2007; Todd *et al.*, 2009) and to quantify changes in detection rates in response to anthropogenic noise (Cox *et al.*, 2001; Culik *et al.*, 2001; Koschinski *et al.*, 2003; Carstensen *et al.*, 2006; Leeney *et al.*, 2007; Carlström *et al.*, 2009). PAM provides data on the time and intensity of detected sounds, not the spatial abundance of animals. To determine the probability of call detection as a function of range and to convert detection rates into estimates of density and abundance, current methods rely on either (1) model predictions of TL coupled with a conversion factor relating click trains detected to the number of individuals present (Stafford *et al.*, 2007; Zimmer *et al.*, 2008) or (2) distance sampling techniques requiring calibration data, for example, PAM detections of sounds produced at known times by animals at known positions (Marques *et al.*, 2009). In the absence of such experimental data collected under equivalent transmission loss conditions, testing assumptions about transmission loss is critical for correct interpretation of passive acoustic monitoring data.

Potential variability in TL in response to environmental fluctuations is likely a significant source of error in PAM density estimates just as it is for biosonar detection range estimates, and since that variability is not generally taken into account in model-based TL estimates, field measurements of habitat-specific and temporal variations in TL are critical.

Here, we experimentally test the hypothesis that a spherical spreading law with attenuation can accurately predict the transmission loss of harbor porpoise-like clicks in porpoise habitats, comparing field measurements of the transmission loss of porpoise-like clicks in porpoise habitats in Canada and Denmark with values predicted by a spherical spreading law with attenuation and by a ray-trace acoustic propagation model (Bellhop). In addition, we outline temporal and spatial variability in TL at our experimental sites and discuss implications for passive acoustic monitoring and the acoustic ecology of porpoises.

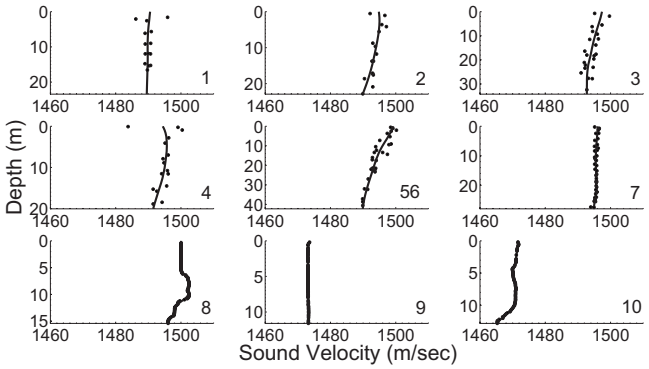


FIG. 1. Sound-speed profiles at the experimental sites. The black dots indicate raw CTD data, and the solid black lines indicate the smoothed sound-speed profiles used for acoustic modeling. Experiment numbers are indicated in the lower right corner of each plot.

II. METHODS

A. Description of field sites

Field measurements of transmission loss of porpoise-like clicks were made in two areas: near Grand Manan, New Brunswick, Canada (44.74°–44.79°N, 66.72°–66.77°W) and in Aarhus Bay near Aarhus, Denmark (56.15°N, 10.3°E). Both locations are characterized by high densities of harbor porpoises and represent important habitat. The Grand Manan experiments (1–7) were carried out at six sites along the northeast coast of the island between 8–13 August 2006. In Aarhus Bay, measurements were made at the same site in the middle of the bay on 5 September 2006, 29 November 2006, and 16 April 2007 (experiments 8–10).

CTD measurements were taken in conjunction with each experiment to determine a sound-speed profile for each site, and echosounder measurements were used to characterize the bathymetry at each site. Figure 1 shows the sound-speed profiles calculated from the CTD (Conductivity, Temperature, Depth) data for experiments 1–10. The sound-speed profiles for experiments 1, 7, and 9 show nearly isovelocity water columns with minimal variation in sound speed with depth. In contrast, experiment sites 2–6 and 10 have downward-refracting sound-speed profiles, and site 8 has a lower-velocity sound channel of about 6–12-m depth. All Grand Manan sites had relatively flat bathymetry, with a maximum downward slope of about 3 deg at sites 5 and 6. The sites in Aarhus Bay had flat bottoms with water depths of 15, 12, and 13 m for experiments 8–10, respectively.

While we did not measure bottom properties at our sites, relevant data are available from several sources. Paskevich *et al.* (2001) and Poppe *et al.* (2005) provided sediment grain-size data for sites about 15 km from the experimental sites, but further from the coast of Grand Manan; they found mainly silty clay, occasionally with a small amount of sand or shells. The sediments in Aarhus Bay are also mainly silty clay or sandy clay (Lund-Hansen *et al.*, 2002; Røy *et al.*, 2005).

B. Acoustic data collection and processing

The experimental setup for transmission loss measurements required two platforms: one to transmit the signals

from a fixed location, and one to receive the signals at a variety of ranges. One boat was anchored or tied up to a fixed object and was attached to the other by a line. The transducer was deployed over the side of the transmit boat, continuously transmitting synthetic porpoise clicks. The length of the line between boats was adjusted to position the receiving boat at stations 5, 10, 25, 50, 100, 150 (experiments 9 and 10 only), and 200 m (experiments 1–7 only) from the transmitter. Source-receiver ranges were verified by radar and range finder when possible. At each station, the receiving boat made 1–5-min recordings of the transmitted signal using two hydrophones deployed at 3- and 5-m depths.

The transmitted signal in all experiments was a series of synthetic porpoise clicks with acoustic properties very similar to those of real clicks. We chose to use synthetic clicks rather than real ones for consistency and to avoid the difficulty of selecting a single click from a single animal to designate as typical. Each synthetic click consisted of 11 (experiments 1–8) or 15 (experiments 9 and 10) equal-amplitude cycles of a 135-kHz tone, with a 10-ms click interval. The duration of each synthetic click was about 82 μs (11 cycles) or 111 μs (15 cycles). Thus, the duration, frequency, and measured bandwidth (about 7 kHz) of the synthetic clicks were similar to the duration (about 75–250 μs), peak frequency (120–142 kHz), and bandwidth (6–26 kHz) of typical harbor porpoise echolocation clicks (Au, 1993; Au *et al.*, 1999; Villadsgaard *et al.*, 2007). We used an Agilent 33220A signal generator (Agilent Technologies, Santa Clara, CA) to produce the clicks at a peak-peak amplitude of 1 V, amplified the signal by 46 dB using a custom-built amplifier (courtesy of Niels U. Kristiansen) for a total amplitude of about 200 V peak-peak, and transmitted them into the water with a Brüel & Kjær 8105 spherical hydrophone (Brüel & Kjær Sound & Vibration Measurement A/S, Nærum, Denmark; transmit sensitivity at 135 kHz of 138 dB relative to 1 μPa/V at 1 m) deployed at either 5-m depth (experiments 1–7 and 10) or 3-m depth (experiments 8 and 9). Although porpoises produce directional echolocation clicks with a −3-dB beam width of about 16 deg (Au *et al.*, 1999), we chose to use an omnidirectional transducer to avoid the problem of ensuring that the receiving hydrophones were on the acoustic axis of a directional transducer. The measured source level of the transmissions was 184 dB re 1-μPa peak-peak at 1 m, which is within the expected range for wild harbor porpoise echolocation clicks [178–205 dB re 1-μPa peak-peak at 1 m (Villadsgaard *et al.*, 2007)]. We used two types of hydrophones as receivers: a Reson TC4034 hydrophone (Reson, Slangerup, Denmark) with a receiving sensitivity of −220 dB relative to 1 V/μPa at 135 kHz and a Reson 4014 hydrophone with a receiving sensitivity of −186 dB relative to 1 V/μPa at 135 kHz.

For the Grand Manan experiments (Nos. 1–7), at each station, 1–5-min sound recordings of the signal were collected with hydrophones deployed at 3- and 5-m depths. The signal at 3-m depth was recorded on a Reson TC4034 hydrophone, amplified either 40 or 60 dB with a custom-built amplifier, and band-pass filtered with an analog filter between 1.7 kHz (one pole) and 160 kHz (four poles). The signal at 5-m depth was recorded on a Reson TC4014 hydrophone, amplified 32 dB with an etec amplifier (etec, Frederiksvaerk, Denmark), and high-pass filtered with an analog filter (one pole) at 1 kHz. All signals from the hydrophones were digitized using a 333-kHz sampling rate (16-bit resolution) on a Wavebook 516E analog to digital converter (IOtech, Cleveland, Ohio), and the resulting files were saved on a laptop computer. In order to maximize resolution in the recordings, amplification on the 3-m hydrophone was varied between 40 and 60 dB, and the clip level of the digital recordings was varied between 0.2- and 10-V peak-peak.

The recording setup was slightly different for the Danish experiments (experiments 8–10). At each station, 1–5-min sound recordings of the signal were collected using hydrophones deployed at 3- and 5-m depths. Both channels were recorded on Reson TC4034 hydrophones amplified either 40 or 60 dB. Signals from the hydrophones were filtered with an analog band-pass filter between 1 kHz (one pole) and 200 kHz (four poles) and digitized at 500-Hz sampling rate (12-bit resolution) on an ADLINK analog-digital converter (ADLINK Technology Inc., Taipei, Taiwan); the resulting data files were saved to a laptop computer. The peak-peak clip level of the digital recordings was 10 V (experiment 8) or 4 V (experiments 9 and 10).

The recordings from each station were band-pass filtered between 100 and 160 kHz with an eighth order Butterworth filter in ADOBE AUDITION (Adobe, San Jose, CA). Using custom-written scripts in MATLAB (The MathWorks, Natick, MA), we applied an envelope-based click detector to extract 100 clicks from each file and calculate the peak-peak received level (RL) of each click. The click detection routine outputs the peak RL of the highest amplitude acoustic arrival only (and not the combined level of several multipath arrivals) as long as the delay in arrival time between the arrivals is greater than about 100 μs and thus the arrivals do not overlap in time. We subtracted the RLs from the measured transmitter source level, 184 dB re 1-μPa peak-peak, to obtain the transmission loss of each click, recording the observed transmission loss of 100 clicks at each station for comparison with model predictions. For experiment 10, we also carried out a pulse-compression analysis (cross-correlation of the data waveform with the theoretical transmitted signal) to examine arrival structure as a function of time.

C. Transmission loss predictions

1. Spreading law calculations

We first applied a spherical spreading law [TL $=20\log_{10}$(range)] to predict transmission loss. At high frequencies such as the 135 kHz considered in these experiments, absorption also contributes significantly to the transmission loss, so we included an additional absorption loss of 0.04 dB/m [representative of values calculated according to Francois and Garrison (1982a, 1982b) for actual environmental conditions at the study sites] in the transmission loss calculation. We expected this spreading law/attenuation TL approximation to be accurate only in areas where the sound speed was relatively homogeneous and the sound did not

interact with the bottom or the sea surface before arriving at the receiver. Because the transmitted signal was very short, one would predict those conditions to hold unless receiver depth was very shallow or source-receiver range was large (Madsen and Wahlberg, 2007).

2. Bellhop acoustic propagation model

We also applied an acoustic propagation model, which can take into account the sound-speed profile, bathymetry, and bottom properties as well as multipath acoustic propagation, to predict transmission loss at each experiment site. Because of the high frequency of the sound source used in these experiments, we chose Bellhop, a ray-tracing model, for these predictions (Porter and Bucker, 1987). We used the ACTUP MATLAB front-end (Amos Maggi and Alec Duncan, available at http://www.cmst.curtin.edu.au/products/actoolbox/, last viewed 6/30/09) to interface with Bellhop. We did not gather data on the bottom properties at each site, but as noted earlier, published data indicate that the experimental sites are dominated by silty clay sediments (Paskevich *et al.*, 2001; Lund-Hansen *et al.*, 2002; Poppe *et al.*, 2005; Røy *et al.*, 2005). In this sediment type, the ratio of sound speed in the surface sediments to sound speed in the water overlying the sediments is generally about 0.984 (Jackson and Richardson, 2007). We combined that ratio with the sound speed at the base of the sound-speed profile to calculate a sediment sound speed for each experimental site; we then used Hamilton's (1978) equations to estimate sediment density. Finally, following Jackson and Richardson (2007) for silty clay sediments, we estimated bottom attenuation in the sediments to be about 0.45 dB per wavelength. The aforementioned sound speed, density, and attenuation values defined an isovelocity bottom layer for each of the experiment sites. We carried out Bellhop model runs to determine the incoherent transmission loss and arrival-time delay of each arrival at the appropriate receiver depth/range locations at each site. Using those results, we calculated a transmission loss estimate for each combination of experiment sites and receiver stations. To match our transmission loss measurements, this estimate included the transmission loss of the highest-amplitude arrival at the receiver (generally the first, direct arrival), summed with the amplitudes of any other arrivals that overlapped temporally with the highest-amplitude arrival. In practice, no temporal overlap of modeled arrivals occurred, so our Bellhop TL estimate was based on the amplitude of a single arrival.

D. Comparison of TL data and predictions

We used two measures to compare our observed TL with the spreading law and Bellhop model predictions. First, we simply calculated the error of each model (for each station in each experiment) by subtracting the average observed TL from the predicted TL. Second, we calculated a root-mean-squared-error (RMSE) value for each experiment and for the set of ten experiments according to RMSE $= \sqrt{(\text{mean}(\text{TL}_{\text{predicted}} - \text{TL}_{\text{data}})^2)}$, using the mean observed TL at each site and range. For RMSE calculations, the error values in decibels ($\text{TL}_{\text{predicted}} - \text{TL}_{\text{data}}$) were converted to linear units (amplitude), and then the resulting RMSE values were converted back into decibels.

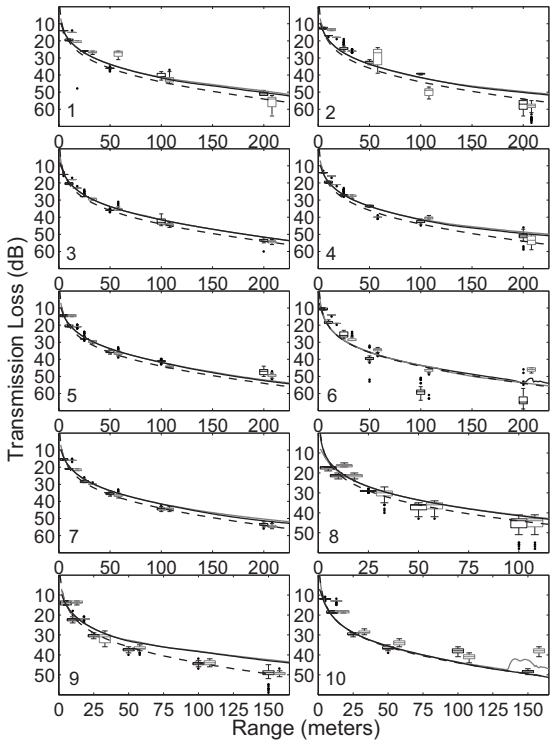


FIG. 2. Measured and modeled transmission losses as a function of range (experiment numbers are indicated in the lower left corner of each panel). Measured received levels are presented as box-and-whisker plots (black boxes show data collected at 3-m receiver depth, and gray boxes show data collected at 5-m receiver depth). The boxes extend to the upper and lower quartiles of observed field TL values, with a horizontal line inside each box indicating the median. The whiskers extend to the largest and smallest observed values in the time bin, up to 1.5 times the interquartile range; values outside this range are plotted as black dots. For readability, the 5-m receiver depth data are offset slightly along the x axis (to higher range values), but were collected at the same time and range as the 3-m data. The dashed black lines show the transmission loss predicted by a spherical spreading model with attenuation. The black and gray traces show transmission loss predicted by the Bellhop acoustic propagation model (black for 3-m receiver depth and gray for 5-m receiver depth).

III. RESULTS

Figure 2 compares measured and modeled transmission losses for experiments 1–10. For most of our experiments (1–4 and 7–9), both Bellhop model output and data inspection confirmed that transmission loss increased with range, the first arrival at each receiver had the highest peak-peak amplitude, and it did not overlap in time with other arrivals. For experiments 5, 6, and 10, however, we found that the highest-amplitude recorded click was sometimes not the first arrival, but a later arrival that appeared to be composed of several overlapping arrivals. We observed this phenomenon at source-receiver ranges as short as 10 m in experiment 5, 25 m in experiment 6, and 50 m in experiment 10. As an example, Fig. 3 shows data from experiment 10 for a receiver depth of 3 m and a source-receiver range of 50 m. The figure includes the waveforms of received arrivals from 100 clicks as well as the results of the pulse-compression analy-

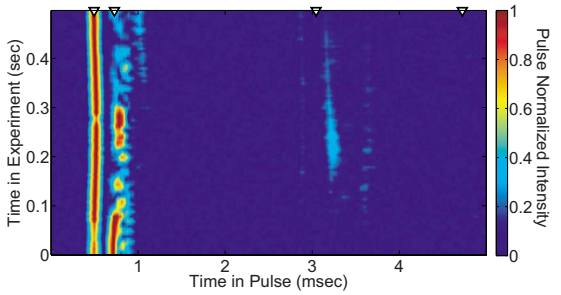


FIG. 3. (Color online) This figure contains data from experiment 10 for a receiver depth of 3 m and a source-receiver range of 50 m. Time is on the *x* axis; 0 points are arbitrary with respect to pulse transmission and reception, but are 10 ms apart to match the inter-click interval of transmissions. Every horizontal line shows data for one pulse (100 consecutive pulses, received over 0.5 s, are plotted). Color indicates pulse intensity resulting from pulse-compression analysis. Arrows on the upper *x* axis indicate Bellhop-predicted arrival times for the first (direct), second (surface-reflected), third (bottom-reflected), and fourth (bottom then surface-reflected) acoustic arrivals.

sis, indicating that the largest peak in the data waveform is actually composed of several overlapping arrivals. We also noted that, in experiments 6 and 10, transmission loss did not increase as smoothly with range as in the other experiments.

Figure 4 summarizes our results regarding the relative accuracy of the spreading law and Bellhop TL predictions. Figure 4 (top panel), which plots the error of both types of predicted TL as a function of source-receiver range, shows that the prediction error did tend to increase with range. However, the error points remain relatively evenly scattered around zero at all ranges, indicating that neither model has a tendency to consistently over- or under-estimate TL as range increases. Error is plotted separately for each experiment in the middle panel of Fig. 4, and the bottom panel shows the RMSE for each experiment. The two plots show that both models predicted TL quite accurately (errors not exceeding 6 dB, RMSE less than 3 dB) for experiments 3 and 7–9 and somewhat accurately for experiments 1, 2, and 4 (errors not exceeding 10 dB, RMSE less than 5 dB). They both performed poorly for experiment 6, though the spreading law

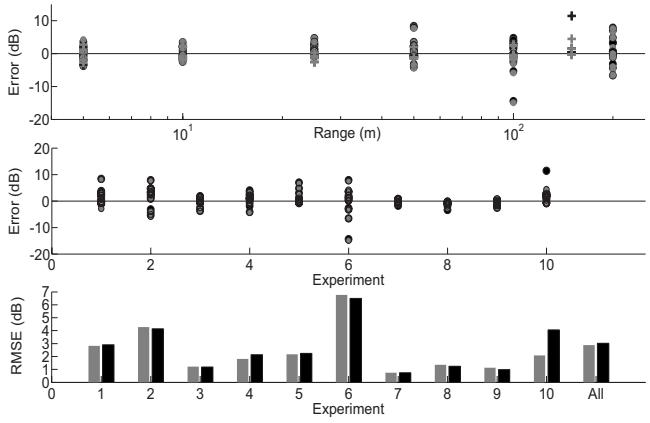


FIG. 4. Error in model predictions of mean transmission loss as a function of range and experiment. Gray symbols indicate Bellhop model error, and black indicates error by the spherical spreading law with attenuation. In the top panel, round symbols represent data from Grand Manan, while crosses are data from Denmark. In the top two panels, each data point gives the error for one source-receiver range in one experiment; in the bottom panel, bars indicate the RMSE for each experiment.

model performed somewhat better than the Bellhop model. The spreading law also performed well for experiment 5, where the Bellhop model performed less well. However, for experiment 10, the Bellhop model performed accurately while the spreading law did not. According to Fig. 4(C), the RMSE for the whole set of ten experiments was between 3 and 4 dB for both the spreading law prediction and the Bellhop model, with the spreading law predictions performing slightly better than the Bellhop model predictions.

IV. DISCUSSION

In this paper, we have compared ray-trace (Bellhop) model predictions of transmission loss with a simple spherical spreading law with attenuation, which is commonly applied in predictions of transmission loss related to marine mammal echolocation and communication (Au, 1993; Madsen and Wahlberg, 2007). Both models performed quite well on average, with the possible exceptions of experiments 5 and 6, where the environment was probably not adequately characterized (see Sec. IV E). The spreading law was able to predict TL with an average error of just over 3 dB at source-receiver ranges up to 200 m, with less error within 50-m range (Fig. 4). We quantified the level of only the highest-amplitude arrival at each station, which rarely interacted with the surface or bottom and usually underwent minimal refraction. Thus, most of the assumptions underlying the spreading law were upheld and its good performance was not surprising.

One might argue that surface and bottom interacting propagation paths are unlikely ever to influence the transmission loss of clicks made by porpoises because, unlike the transducer used in our experiments, porpoises produce highly directional clicks with a −3-dB beam width of about 16 deg in the vertical plane (Au *et al.*, 1999). Given this narrow beam and assuming that porpoises always echolocate directly toward targets in the water column, direct-path propagation likely dominates. However, this argument does not hold for PAM detections, for other sound recordings, or for cases in which the porpoise directs its echolocation beam at an angle toward the surface or bottom rather than directly at a potential target.

A. Effects of water-column refraction and boundary interactions

In some particular cases, the spreading law may not perform well, while Bellhop does better; experiment 10 is an example. There, the spreading law prediction failed due to the presence of a sound channel at about 5-m depth (Fig. 1); multiple ray paths passed through a focus near the receiver at 5-m depth and 150-m range (Fig. 5), significantly reducing the transmission loss at that location. In addition, the most powerful arrivals were not always the first, direct arrivals (Fig. 3). The spreading law model cannot be relied upon in such cases, where surface/bottom interactions and/or refraction in the water column significantly affect transmission loss.

The Bellhop model output also underestimated TL at 150-m range somewhat for experiment 10; given the obser-

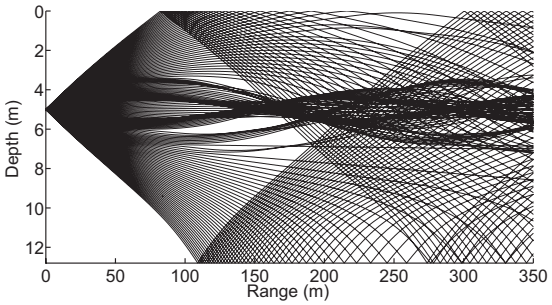


FIG. 5. Bellhop ray-trace output for the experiment 10 site. For clarity, only rays with launch angles between ±20 deg from the horizontal are shown. Some rays pass through a focus at about 5-m depth and 150-m range.

vation that the amplitude and timing of the surface-reflected arrival were rather variable and sometimes seemed to comprise several arrivals (Fig. 3), it seems possible that interactions with surface waves may explain these results (Tindle *et al.*, 2009). We did not measure surface waves or related environmental parameters in the field and did not include them in our model.

B. Implications for passive acoustic monitoring

Previous work has shown that click source levels of free-ranging porpoises average 191 dB re 1-μPa peak-to-peak at 1 m (Villadsgaard *et al.*, 2007), and that T-POD porpoise detectors can detect porpoise clicks at levels as low as 123–132 dB re 1-μPa peak-to-peak (Kyhn *et al.*, 2008). Thus, T-PODs may detect on-axis porpoise clicks as long as TL is less than about 68 dB [28 dB for off-axis clicks since their source levels are 40 dB lower (Hansen *et al.*, 2008)]. TL of this order is within or slightly exceeds the values considered in this study (measured TL first exceeded 28 dB at 25- or 50-m range and did not exceed 68 dB at maximum ranges of 150–200 m). Our results can thus be used to approximate the errors that would result from use of a spreading law-based model of transmission loss in T-POD detection range estimates for areas similar to our study sites. Such errors should be relatively low on average for off-axis clicks detected at ranges of 50 m or less, but would probably be greater for on-axis clicks that are likely to be detected at greater ranges (Fig. 4).

Specifically, the variability in TL as a function of depth discussed in Sec. IV A has implications for passive acoustic monitoring since it results in detection probabilities and distances that vary with animal depth and detector depth. Assuming that spherical spreading with attenuation accurately estimates transmission loss, one would expect a T-POD to detect porpoises at ranges of up to about 700 m. However, in conditions like those at the experiment 10 site, a T-POD in the sound channel would actually be able to detect porpoises in the sound channel at ranges of up to about 1200 m (calculated using Bellhop output for the experiment 10 site). If not accounted for, variability of this magnitude could easily result in very large errors in estimates of animal density. This effect may be particularly important to consider when comparing seasonal trends in animal detections at a given location since the presence of a sound duct during part of the year could lead to the erroneous conclusion that more animals are present during that period compared to periods without a sound duct. Therefore, variability in TL as a function of depth and time should be considered when selecting detector deployment depths and reviewing passive acoustic monitoring data.

C. Implications for porpoise biosonar and communication ranges

The results of this study should be relevant to estimation of porpoise communication and echolocation detection ranges as long as our source-receiver ranges match the distances over which porpoises actually communicate and echolocate. Direct observations of such distances have not been made, but they can be estimated. Consider a porpoise producing echolocation clicks with source levels of 191 dB re 1-μPa peak-to-peak at 1 m (Villadsgaard *et al.*, 2007), equivalent to an energy flux density of 140 dB re 1 μPa^2 s (Kastelein *et al.*, 1999). Assume that the porpoise echolocates on a herring with a target strength of −37 dB (Ona, 2003), with a returning echo detection threshold of about 44 dB re 1 μPa^2 s (Au *et al.*, 2007). On average, echoes with one-way transmission loss of 29.5 dB or less would be detectable to the porpoise [TL=(140−44+(−37))/2 =29.5 dB]. Our observed TL first exceeded 29.5 dB at ranges of 25–50 m (Fig. 2). For a similar calculation of communication call detection ranges, assume that the threshold is as above, and that communication click source levels average about 180 dB re 1-μPa peak-to-peak at 1 m (Clausen *et al.*, 2008), equivalent to 129 dB re 1 μPa^2 s (Kastelein *et al.*, 1999). Communication sounds with TL of no more than 85 dB would then be detectable to conspecifics assuming that both transmitter and receiver are oriented toward one another (Hansen *et al.*, 2008). 85 dB is much higher than the largest TL value we observed (61 dB at ranges up to 200 m). Even that value could lead to an underestimate of communication range if the passive detection threshold is lower than the echo detection threshold, or if wild porpoise communication click source levels are higher than those measured in captivity.

While the spherical spreading law will in many cases provide a good estimate of the active space of porpoise clicks, experiment 10 illustrates that refraction in the water column can strongly affect transmission loss in some cases. Porpoises may be able to exploit such sound channels to increase the range at which they can detect prey or other targets with their echolocation or to increase the active space of their communication calls. For example, assuming that spherical spreading with attenuation applies and that communication calls remain detectable until TL exceeds 85 dB (see above calculations), the predicted maximum active space of an on-axis porpoise communication call would be about 700 m (100 m for an off-axis call with a source level 40 dB lower). By taking advantage of a sound channel like the one at the experiment 10 site, porpoises could significantly increase their maximum active space to 1.1 km for an on-axis call (187 m for an off-axis call, both calculated using Bellhop output for the experiment 10 site). Future studies could test the hypothesis that porpoises exploit sound channels by

looking for a relationship between sound-speed profiles and the depth distribution of porpoise when they are calling and receiving acoustic signals, preferably using data from acoustic and depth-recording tags.

D. Temporal variability

While the modeling approaches we used predicted measured TL with relatively low error when results were averaged over time and space, variability in measured TL was high (± 10 dB) over timescales of about 1 s (Figs. 2 and 3), indicating that short-timescale environmental variability can strongly influence the TL of individual porpoise clicks. Variability of this magnitude on similar timescales has been previously observed in several studies and attributed to reflection of underwater sound from surface waves (Tindle *et al.*, 2009) or the presence of a steep thermocline (where the speed of sound changed by about 4 m/s per meter of depth) (Wilson *et al.*, 1994). However, because our data set contains mostly direct (rather than surface-reflected) arrivals, and the sound-speed profiles at our study sites did not seem to vary rapidly with depth, neither of those sources of variability seem a likely explanation for the short-timescale variability in our results. This variability is, however, similar to that observed in TL measurements in manatee habitats (Miksis-Olds and Miller, 2006).

Temporal variability in TL, like the effects discussed in Sec. IV A-B, has potential to strongly affect density estimates calculated from passive acoustic monitoring data. If possible, this variability should be quantified at PAM sites, and sites with lower variability should be selected for detector placement.

E. Sources of error in the data set

We would not necessarily expect a perfect match between our data and models due to several potential sources of error. First, we estimate that error in the source-receiver range measurements could have been as much as 10% (especially at longer ranges) due to the effects of wind, currents, and tide on the boats. Second, tidal action (especially at the Grand Manan sites) may have caused the hydrophones to hang slightly obliquely, causing errors in the nominal source and receiver depths.

Predictions from both the spherical spreading law and the Bellhop model were less accurate for experiments 5 and 6, especially at longer ranges. We suspect that the environment model we used as Bellhop input was incomplete especially for the experiment 5 and 6 sites, as the area around the site contains boulders and a rocky ledge or outcropping, features that were not adequately characterized for inclusion in the environmental model.

V. CONCLUSIONS

On average, for most of the cases we considered, spherical spreading with attenuation provided relatively accurate estimates of transmission loss of synthetic porpoise clicks in natural porpoise habitats. Both of our modeling approaches were able to predict transmission loss with overall RMSE of less than about 4 dB, and in half of the study sites with errors of less than 2 dB. However, there was considerable variability in TL around the average value. Characterization of this variability is important before using animal sounds to estimate density and may also provide insight into porpoise acoustic ecology. The cases where spherical spreading failed to predict TL accurately involved situations where the dominant acoustic arrival was strongly affected by surface reflection or refraction in the water column. The existence of those conditions depends on site-specific source-receiver geometry, bathymetry, sediment properties, and sound-speed profile, so spherical spreading may not accurately describe transmission loss in porpoise habitats that differ significantly from the sites considered here. When adequate data on the acoustic environment are available, a more detailed model such as Bellhop can be a useful tool to assess the effects of environmental conditions on transmission loss at a particular site.

ACKNOWLEDGMENTS

We are very grateful to everyone at the Grand Manan Whale and Seabird Research Station for their advice and assistance during data collection in Canada, especially Sarah Wong, Marla Bojarski, and Hillary Lane. We also thank the crew of RV Genetica, Aarhus University, Denmark, for their help in collecting the Danish field data. Ying-Tsong Lin, James Lynch, and Arthur Newhall provided advice related to data interpretation and acoustic propagation modeling, and Ying-Tsong Lin also provided a MATLAB script used as the basis of the pulse-compression analysis. Regina Campbell-Malone, Line Kyhn, Xavier Lurton, and Jakob Tougaard provided helpful comments. Field data collection was partially supported by a Student Research Award from the WHOI Ocean Life Institute (Grant No. 25051351). P.T.M. and M.H. were funded by Steno and frame grants from the Danish Natural Science Foundation.

Akamatsu, T., Teilmann, J., Miller, L. A., Tougaard, J., Dietz, R., Wang, D., Wang, K. X., Siebert, U., and Naito, Y. (**2007**). "Comparison of echolocation behaviour between coastal and riverine porpoises," Deep-Sea Res., Part II **54**, 290–297.

Akamatsu, T., Wang, D., Wang, K. X., and Naito, Y. (**2005**). "Biosonar behaviour of free-ranging porpoises," Proc. R. Soc. London, Ser. B **272**, 797–801.

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

Au, W. W. L., Benoit-Bird, K. J., and Kastelein, R. A. (**2007**). "Modeling the detection range of fish by echolocating bottlenose dolphins and harbor porpoises," J. Acoust. Soc. Am. **121**, 3954–3962.

Au, W. W. L., Ford, J. K. B., Horne, J. K., and Allman, K. A. N. (**2004**). "Echolocation signals of free-ranging killer whales (*Orcinus orca*) and modeling of foraging for chinook salmon (*Oncorhynchus tshawytscha*)," J. Acoust. Soc. Am. **115**, 901–909.

Au, W. W. L., and Jones, L. (**1991**). "Acoustic reflectivity of nets—Implications concerning incidental take of dolphins," Marine Mammal Sci. **7**, 258–273.

Au, W. W. L., Kastelein, R. A., Rippe, T., and Schooneman, N. M. (**1999**). "Transmission beam pattern and echolocation signals of a harbor porpoise (*Phocoena phocoena*)," J. Acoust. Soc. Am. **106**, 3699–3705.

Carlström, J. (**2005**). "Diel variation in echolocation behavior of wild harbor porpoises," Marine Mammal Sci. **21**, 1–12.

Carlström, J., Berggren, P., and Tregenza, N. J. C. (**2009**). "Spatial and temporal impact of pingers on porpoises," Can. J. Fish. Aquat. Sci. **66**, 72–82.

Carstensen, J., Henriksen, O. D., and Teilmann, J. (**2006**). "Impacts of offshore wind farm construction on harbour porpoises: Acoustic monitoring of echolocation activity using porpoise detectors (T-PODs)," Mar. Ecol.: Prog. Ser. **321**, 295–308.

Clausen, K. T., Madsen, P. T., and Wahlberg, M. (**2010**). "Click communication in harbour porpoises, (*Phocoena phocoena*)," Bioacoustics (in press).
Cox, T. M., Read, A. J., Solow, A., and Tregenza, N. (**2001**). "Will harbour porpoises (*Phocoena phocoena*) habituate to pingers?," J. Cetacean Res. Manage. **3**, 81–86.
Culik, B. M., Koschinski, S., Tregenza, N., and Ellis, G. M. (**2001**). "Reactions of harbor porpoises *Phocoena phocoena* and herring *Clupea harengus* to acoustic alarms," Mar. Ecol.: Prog. Ser. **211**, 255–260.
Francois, R. E., and Garrison, G. R. (**1982a**). "Sound absorption based on ocean measurements: Part I: Pure water and magnesium sulfate contributions," J. Acoust. Soc. Am. **72**, 896–907.
Francois, R. E., and Garrison, G. R. (**1982b**). "Sound absorption based on ocean measurements: Part II: Boric acid contribution and equation for total absorption," J. Acoust. Soc. Am. **72**, 1879–1890.
Hamilton, E. L. (**1978**). "Sound velocity-density relations in sea-floor sediments and rocks," J. Acoust. Soc. Am. **63**, 366–377.
Hansen, M., Wahlberg, M., and Madsen, P. T. (**2008**). "Low-frequency components in harbor porpoise (*Phocoena phocoena*) clicks: Communication signal, by-products, or artifacts?," J. Acoust. Soc. Am. **124**, 4059–4068.
Jackson, D. R., and Richardson, M. D. (**2007**). *High-Frequency Seafloor Acoustics* (Springer Science+Business Media L.L.C., New York).
Kastelein, R. A., Au, W. W. L., and De Haan, D. (**2000**). "Detection distances of bottom-set gillnets by harbour porpoises (*Phocoena phocoena*) and bottlenose dolphins (*Tursiops truncatus*)," Mar. Environ. Res. **49**, 359–375.
Kastelein, R. A., Au, W. W. L., Rippe, H. T., and Schooneman, N. M. (**1999**). "Target detection by an echolocating harbor porpoise (*Phocoena phocoena*)," J. Acoust. Soc. Am. **105**, 2493–2498.
Koschinski, S., Culik, B. M., Henriksen, O. D., Tregenza, N., Ellis, G., Jansen, C., and Kathe, G. (**2003**). "Behavioural reactions of free-ranging porpoises and seals to the noise of a simulated 2 MW windpower generator," Mar. Ecol.: Prog. Ser. **265**, 263–273.
Kyhn, L. A., Tougaard, J., Jensen, F., Wahlberg, M., Stone, G., Yoshinaga, A., Beedholm, K., and Madsen, P. T. (**2009**). "Feeding at a high pitch: Source parameters of narrow band, high-frequency clicks from echolocating off-shore hourglass dolphins and coastal Hector's dolphins," J. Acoust. Soc. Am. **125**, 1783–1791.
Kyhn, L. A., Tougaard, J., Teilmann, J., Wahlberg, M., Jorgensen, P. B., and Bech, N. I. (**2008**). "Harbour porpoise (*Phocoena phocoena*) static acoustic monitoring: Laboratory detection thresholds of T-PODs are reflected in field sensitivity," J. Mar. Biol. Assoc. U.K. **88**, 1085–1091.
Leeney, R. H., Berrow, S., McGrath, D., O'Brien, J., Cosgrove, R., and Godley, B. J. (**2007**). "Effects of pingers on the behaviour of bottlenose dolphins," J. Mar. Biol. Assoc. U.K. **87**, 129–133.
Lund-Hansen, L. C., Laima, L. C., Mouritsen, K. N., Lam, Y., and Hai, A. (**2002**). "Effects of benthic diatoms, fluff layer, and sediment conditions on critical shear stress in a non-tidal coastal environment," J. Mar. Biol. Assoc. U.K. **82**, 929–936.
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. Exp. Biol. **207**, 1811–1823.
Madsen, P. T., and Wahlberg, M. (**2007**). "Recording and quantification of ultrasonic echolocation clicks from free-ranging toothed whales," Deep-Sea Res., Part I **54**, 1421–1444.
Marques, T. A., Thomas, L., Ward, J., DiMarzio, N., and Tyack, P. L. (**2009**). "Estimating cetacean population density using fixed passive acoustic sensors: An example with Blainville's beaked whales," J. Acoust. Soc. Am. **125**, 1982–1994.
Miksis-Olds, J. L., and Miller, J. H. (**2006**). "Transmission loss in manatee habitats," J. Acoust. Soc. Am. **120**, 2320–2327.
Mooney, T. A., Au, W. W. L., Nachtigall, P. E., and Trippel, E. A. (**2007**). "Acoustic and stiffness properties of gillnets as they relate to small cetacean bycatch," ICES J. Mar. Sci. **64**, 1324–1332.
Mooney, T. A., Natchigall, P. E., and Au, W. W. L. (**2004**). "Target strength of a nylon monofilament and an acoustically enhanced gillnet: Predictions of biosonar detection ranges," Aquat. Mamm. **30**, 220–226.
Ona, E. (**2003**). "An expanded target-strength relationship for herring," ICES J. Mar. Sci. **60**, 493–499.
Paskevich, V. F., Poppe, L. J., Hastings, M. E., and Hathaway, J. C. (**2001**). "Sea floor photography from the continental margin program: A pictorial survey of benthic character and habitats along the U.S. east coast," Open File Report 01-154, United States Geological Survey, Woods Hole, MA.
Philpott, E., Englund, A., Ingram, S., and Rogan, E. (**2007**). "Using T-PODs to investigate the echolocation of coastal bottlenose dolphins," J. Mar. Biol. Assoc. U.K. **87**, 11–17.
Poppe, L. J., Williams, S. J., and Paskevich, V. F. (**2005**). "U.S.G.S. East-coast sediment analysis: Procedures, database and G.I.S. data," Open File Report No. 05-1001, United States Geological Survey, Woods Hole, MA.
Porter, M. B., and Bucker, H. P. (**1987**). "Gaussian beam tracing for computing ocean acoustic fields," J. Acoust. Soc. Am. **82**, 1349–1359.
Røy, H., Huettel, M., and Jørgensen, B. B. (**2005**). "The influence of topography on the functional exchange surface of marine soft sediments, assessed from sediment topography measured *in situ*," Limnol. Oceanogr. **50**, 106–112.
Stafford, K. M., Mellinger, D. K., Moore, S. E., and Fox, C. G. (**2007**). "Seasonal variability and detection range modeling of baleen whale calls in the Gulf of Alaska, 1999–2002," J. Acoust. Soc. Am. **122**, 3378–3390.
Thomsen, F., van Elk, N., Brock, V., and Piper, W. (**2005**). "On the performance of automated porpoise-click-detectors in experiments with captive harbor porpoises (*Phocoena phocoena*) (L)," J. Acoust. Soc. Am. **118**, 37–40.
Tindle, C. T., Deane, G. B., and Preisig, J. C. (**2009**). "Reflection of underwater sound from surface waves," J. Acoust. Soc. Am. **125**, 66–72.
Todd, V. L. G., Pearse, W. D., Tregenza, N. C., Lepper, P. A., and Todd, I. B. (**2009**). "Diel echolocation activity of harbour porpoises (*Phocoena phocoena*) around North Sea offshore gas installations," ICES J. Mar. Sci. **66**, 734–745.
Verfuss, U. K., Honnef, C. G., Meding, A., Dahne, M., Mundry, R., and Benke, H. (**2007**). "Geographical and seasonal variation of harbour porpoise (*Phocoena phocoena*) presence in the German Baltic Sea revealed by passive acoustic monitoring," J. Mar. Biol. Assoc. U.K. **87**, 165–176.
Villadsgaard, A., Wahlberg, M., and Tougaard, J. (**2007**). "Echolocation signals of wild harbour porpoises, *Phocoena phocoena*," J. Exp. Biol. **210**, 56–64.
Wilson, M. A., Farwell, R. W., and Stanic, S. (**1994**). "High-frequency propagation statistics in a shallow water environment," in *OCEANS '94 Proceedings* (IEEE, Brest, France), pp. 261–265.
Zimmer, W. M. X., Harwood, J., Tyack, P. L., Johnson, M. P., and Madsen, P. T. (**2008**). "Passive acoustic detection of deep-diving beaked whales," J. Acoust. Soc. Am. **124**, 2823–2832.

Depth dependent variation of the echolocation pulse rate of bottlenose dolphins (*Tursiops truncatus*)

Peter Simard[a)]
College of Marine Science, University of South Florida, 140 7th Avenue South, St. Petersburg, Florida 33701

Ashley L. Hibbard and Kimberly A. McCallister
Eckerd College, 4200 54th Avenue South, St. Petersburg, Florida 33711

Adam S. Frankel, David G. Zeddies, and Geoffrey M. Sisson
Marine Acoustics, Inc., 4100 North Fairfax Drive, Suite 730, Arlington, Virginia 22203

Shannon Gowans and Elizabeth A. Forys
Eckerd College, 4200 54th Avenue South, St. Petersburg, Florida 33711

David A. Mann
College of Marine Science, University of South Florida, 140 7th Avenue South, St. Petersburg, Florida 33701

(Received 4 May 2009; revised 6 September 2009; accepted 25 September 2009)

Trained odontocetes appear to have good control over the timing (pulse rate) of their echolocation clicks; however, there is comparatively little information about how free-ranging odontocetes modify their echolocation in relation to their environment. This study investigates echolocation pulse rate in 14 groups of free-ranging bottlenose dolphins (*Tursiops truncatus*) at a variety of depths (2.4–30.1 m) in the Gulf of Mexico. Linear regression models indicated a significant decrease in mean pulse rate with mean water depth. Pulse rates for most groups were multi-modal. Distance to target estimates were as high as 91.8 m, assuming that echolocation was produced at a maximal rate for the target distance. A 5.29-ms processing lag time was necessary to explain the pulse rate modes observed. Although echolocation is likely reverberation limited, these results support the hypotheses that free-ranging bottlenose dolphins in this area are adapting their echolocation signals for a variety of target detection and ranging purposes, and that the target distance is a function of water depth.
 [DOI: 10.1121/1.3257202]

PACS number(s): 43.80.Ka, 43.80.Jz [WWA] Pages: 568–578

I. INTRODUCTION

Echolocation is the active use of sound by animals to detect, discriminate, and localize objects based on the qualities of returning echoes (Au, 1993). Sophisticated echolocation abilities are known only in microchiropteran bats and the odontocete cetaceans (Bradbury and Vehrencamp, 1998). The details of odontocete echolocation signals, the function of echolocation, and the abilities of echolocating odontocetes have been determined in a number of studies on trained animals (mostly on common bottlenose dolphins, *Tursiops truncatus*; for a review, see Au, 1993). Bottlenose dolphin echolocation clicks are broadband pulses with high peak frequencies (up to 130 kHz), high source levels (180–227 dB re 1 μPa peak-to-peak), short duration (individual pulses 40–70 μs), and moderate to high directionality (3-dB beamwidth 10°–40°, Au *et al.*, 1974, 1978, 1986; Moore *et al.*, 2008).

Free-ranging odontocetes routinely produce pulsed sounds, which many studies have shown to be used for echolocation. Compelling evidence for echolocation being used for foraging purposes has been found for several free-ranging species using acoustic tags with accelerometers and depth recorders (e.g., Cuvier's beaked whale, *Ziphius cavirostris*, Blainville's beaked whale, *Mesoplodon densirostris*; Johnson *et al.*, 2004; Madsen *et al.*, 2005, sperm whale, *Physeter macrocephalus*; Miller *et al.*, 2004, finless porpoise, *Neophocaena phocaenoides*; Akamatsu *et al.*, 2005). Echolocation has also been observed to be used by free-ranging bottlenose dolphins during foraging behaviors (Herzing, 1996; Nowacek, 2005). Free-ranging studies also suggest the use of echolocation for navigation purposes. For example, the initial click patterns of sperm whales have been correlated in several cases with water depth, suggesting that echolocation serves a navigation purpose (Jaquet *et al.*, 2001; Thode *et al.*, 2002).

Odontocetes appear to have good control over the timing (pulse rate) of their echolocation clicks. Most results indicate a strong inverse relationship between pulse rate and target distance (e.g., Au *et al.*, 1974), a result also observed in bat echolocation (Tian and Schnitzler, 1997). Dolphins typically produce a click that travels to a target, is reflected back to the dolphin, and is processed by the dolphin before another click

[a)]Author to whom correspondence should be addressed. Electronic mail: psimard@mail.usf.edu

is produced. Various studies have investigated this lag time and have produced estimates ranging between 19 and 45 ms for targets 0.4–120 m away in bottlenose dolphins (Au *et al.*, 1974), although lag times as low as 2.5 ms have been reported for very close target distances (Evans and Powell, 1967; Au, 1993). Several studies have investigated the relationship between pulse rate and distance to target in free-ranging odontocetes. Akamatsu *et al.* (1998) demonstrated that free-ranging bottlenose dolphins, baiji (*Lipotes vexillifer*), and finless porpoises have slower pulse rates than their counterparts housed in tanks. This was interpreted as a function of potential target distances, as captive odontocetes have a limited range in which they can echolocate (Akamatsu *et al.*, 1998). The pulse rate of free-ranging bottlenose dolphins was found to correspond to the two-way travel time plus a lag time in a manner similar to the results obtained from trained animals (Jensen *et al.*, 2009). However, odontocete echolocation does not appear to be limited to this model; Turl and Penner (1989) found that a captive beluga whale (*Delphinapterus leucas*) was able to effectively echolocate using pulses produced at a higher rate than the two-way travel time. The ability to process a series of echolocation clicks produced at a higher rate than the two-way travel time has also been demonstrated by Ivanov (2004) using a trained bottlenose dolphin in a natural environment. Timing of pulses has also been found to be influenced by the purpose of their production. For example, Penner (1988) found that when presented with a target at unknown range, captive bottlenose dolphins would begin acoustic searching behavior by echolocating at pulse rates appropriate for the maximum distance they expected to encounter. During deep dives, sperm whales and Cuvier's and Blainville's beaked whales alter their echolocation pulse rates from slow, relatively constant foraging clicks to a rapid pulse rate (known as buzzing) when in final approach to prey (Madsen *et al.*, 2002; Johnson *et al.*, 2004; Miller *et al.*, 2004; Johnson *et al.*, 2008). A similar pattern of relatively constant pulse rates for searching and increasing pulse rates for decreasing target distances during the final approach to a target has also been observed in harbor porpoises (*Phocoena phocoena*; Verfuß *et al.*, 2009).

The coastal waters of west central Florida are characterized by a series of shallow lagoons (depth typically <10 m), protected by barrier islands from the open Gulf of Mexico, itself characterized by a broad continental shelf and little bathymetric variability. Bottlenose dolphins are commonly found in this area and have been the subject of intensive studies in several locations. Consequently, a great deal is known about the ecology, community structure, acoustic behavior, and movements of these dolphins. Bottlenose dolphins in this and other areas have been classified as inshore, intermediate, and offshore types, each characterized by unique genetic, histological, morphological, and ecological differences (Barros and Odell, 1990; Hersh and Duffield, 1990; Wells *et al.*, 1999; Sellas *et al.*, 2005). Dolphins in Sarasota Bay have high rates of residency and natal philopatry (Wells and Scott, 1990), a pattern which appears to be true for at least some other lagoons nearby (e.g., Boca Ciega Bay; S. Gowans, unpublished data). However, there is some mixing both between adjacent lagoon communities and between inshore and offshore groups, resulting in a series of overlapping communities (Wells and Scott, 1990; Fazioli and Wells, 1999). As in other areas, prey species for west central Florida bottlenose dolphins are composed of several fish and cephalopod species and reflect local prey abundances (Barros and Odell, 1990; Barros and Wells, 1998; Allen *et al.*, 2001). There is some evidence that bottlenose dolphins in Sarasota Bay do not rely on echolocation for navigation while traveling, but do tend to echolocate while foraging (Nowacek, 2005). There is also evidence that Sarasota Bay bottlenose dolphins also may forage using passive listening for soniforous prey (Gannon *et al.*, 2005).

The ability to modify the qualities of echolocation clicks allows odontocetes to adapt the signals to different tasks and different environments. Despite the great deal of data on dolphin echolocation in trained dolphins, there are no studies which address the relationship between echolocation click rates and environmental parameters in multiple, independent groups of free-ranging dolphins. The goal of this study was to determine whether there are differences in echolocation pulse rates between groups of free-ranging bottlenose dolphins as a function of depth in the coastal waters of west central Florida. The water depth is likely to be an important factor in the three-dimensional navigation which odontocetes must routinely contend with, and because many prey species have depth dependent distribution, depth is also an indirect predictor of foraging behavior. Therefore, we hypothesize that the pulse rate of bottlenose dolphin echolocation clicks will be a function of depth.

II. METHODS

Field work was conducted from April to September 2008 in the Gulf of Mexico in west central Florida, including open Gulf areas extending to 50-km offshore, Tampa Bay, and Boca Ciega Bay (a shallow lagoon protected by barrier islands situated on the northern mouth of Tampa Bay). The boundaries of the study area were approximately 28°00′ N to 27°00′ N and 82°35′ W to 83°10′ W (Fig. 1). Survey depths ranged from approximately 1 to 30 m. Cetaceans were located visually by experienced marine mammal observers (naked-eye supplemented with 7×50-mm^2 binoculars when necessary) from research vessels and aided during April cruises with an observer in a Cessna 172 aircraft. Four vessels were used in the study: R/V *Eugenie Clark* (14-m Newton, twin 320-hp diesel engines), R/V *Fish Hawk* (11-m Sea Hawk, twin 450-hp diesel engines), R/V *MS*3 (5.8-m Cape Horn, 95-hp four-stroke outboard), and R/V *LRZ* (4.2-m rigid inflatable Zodiac, 25-hp four-stroke outboard). Cruises on the *Eugenie Clark* and *Fish Hawk* were conducted for the Dolphin Ecology Vocalizations and Oceanography project, while those on the *MS*3 and *LRZ* were conducted for the Eckerd College Dolphin Project. For each group of dolphins we determined species, group size, and for some groups, photo-identification data were collected. GPS (Global Positioning System) locations were recorded automatically every 60 s using a Garmin GPSMAP 76c throughout the cruises.

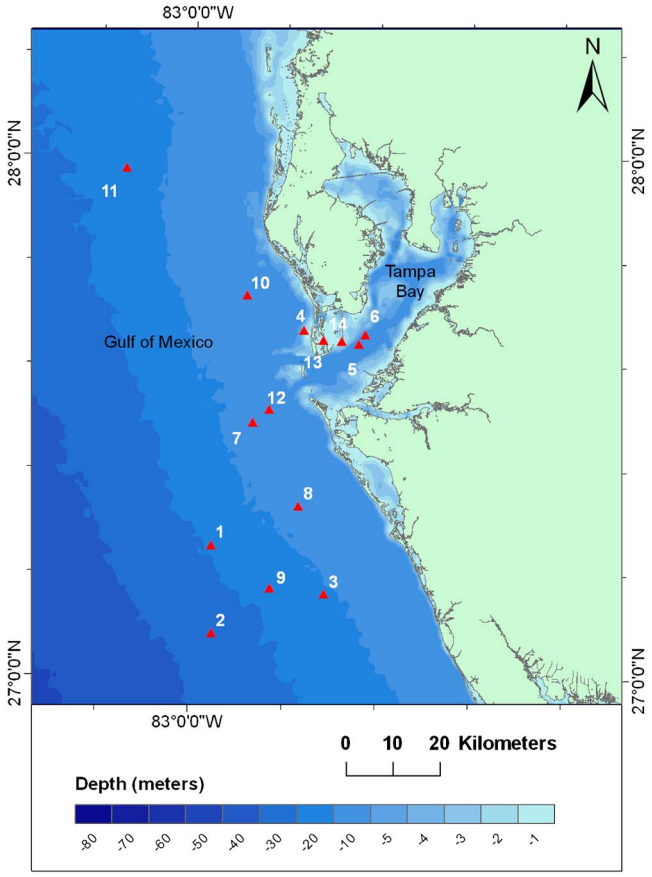


FIG. 1. (Color online) Map of study area, median locations of 14 groups of bottlenose dolphins shown, and numbers refer to groups in chronological order and are given with group identification in Table I. Depths are in meters.

Acoustic recordings of dolphin groups were made throughout each encounter whenever possible. Two recording systems were used. Aboard the R/V *Eugenie Clark*, recordings were made using a custom recording system using ISHMAEL software (D. Mellinger), a single pole high-pass filter (corner frequency of 1.5 kHz), Alligator Technologies amplifier (60-dB gain), and 32-kHz elliptical anti-aliasing filter. Recordings were sampled at 64 kHz with 16-bit resolution. A 16-element towed array was used (Innovative Transducers, Inc., Fort Worth, TX, sensitivity: −212 dBV/μPa, ±3 dB up to 4 kHz, −1.2 dBV/μPa per 1 kHz above 4 kHz), with depth varying with vessel speed. All other recordings were made with an M-Audio 24/96 digital recorder (16-bit resolution, 96-kHz sample rate, +3.9-dB gain between 20 Hz and 50 kHz, and anti-aliasing filter with corner frequency of 47.8 kHz), with a single HTI-96-MIN omnidirectional hydrophone (sensitivity: −180 dBV/μPa, ±3 dB from 10 Hz to 50 kHz). This hydrophone was kept at approximately 5-m depth for vessel speeds of 8–10 km/h when in open Gulf waters by using a 2.5-kg lead weight and at a constant depth of 1 m when in inshore waters by suspending the hydrophone, several meters of cable, and two 2.5-kg lead weights from a surface float.

Multiple recordings of sounds from a group of dolphins cannot be considered statistically independent, as there is a high probability that the same animal has produced multiple sounds. When analyzing the characteristics of multiple clicks within click trains, this lack of independence is a certainty (Soldevilla *et al.*, 2008). Therefore, the analysis of clicks was grouped by sighting: all recorded clicks within each sighting were treated as a single sample point for statistical purposes. In order to increase the probability that the recordings from each group were statistically independent from one another, we attempted to reduce multiple recordings of the same group and of the same individuals. This was accomplished in one of two ways. For inshore groups (near Tampa Bay and Boca Ciega Bay), independence was determined using photo-identification results; groups containing individuals who were already identified in previously recorded groups used in the analysis were omitted from the analysis. For groups found further offshore (≥10 km from shore), photo-identification results were unavailable, and independence was estimated by determining the distance between the first position of a group and the last positions of all previous groups seen on the same day. Mean bottlenose dolphin movement speeds have been estimated to be approximately 2–5 km/h (Irvine *et al.*, 1981; Mate *et al.*, 1995; Wells *et al.*, 1999). Therefore, as a conservative approach, groups were considered independent only if the distances between the first position of the group and the last positions of all previous groups divided by the corresponding time differences were greater than 10 km/h. This procedure was repeated each day that field work was conducted. Bottlenose dolphins more than a few kilometers from shore appear to be more transient in nature and travel longer distances than their inshore counterparts (Fazioli and Wells, 1999; Wells *et al.*, 1999). The area just offshore of Sarasota (the southern half of our study area) appears to be the confluence of two separate offshore stocks (Fazioli and Wells, 1999). Consequently our study area is potentially visited by a large number of individual dolphins, and these visits are likely ephemeral. Therefore, this method is likely sufficient to ensure statistical independence for dolphins 10 km or more from shore. Only recordings from groups containing exclusively bottlenose dolphins were used in this analysis.

Echolocation clicks were identified in acoustic files in ADOBE AUDITION (version 2, Adobe Systems, Inc., San Jose, CA). All echolocation trains were isolated into individual wav files for analysis. Only echolocation trains from single dolphins were used in analysis. Files containing overlapping echolocation from more than one dolphin were omitted from analysis (unless the amplitudes of additional echolocation clicks were comparatively very low). This was important in order to determine the pulse rate of individual dolphins; overlapping pulse trains would artificially increase the pulse rate in the analysis. Echolocation click trains were defined as sets of pulses in which each pair of pulses had a pulse period (time between amplitude maxima, also known as inter-pulse interval or inter-click interval) greater than half the preceding pulse period or less than twice the preceding pulse period; otherwise the sets of pulses were considered separate click trains. Burst-pulses were identified in ADOBE AUDITION by the presence of horizontal banding when viewed in 1024-point FFT (fast Fourier transform) spectrograms (see Watkins, 1967). As burst-pulses are believed to be used for com-

TABLE I. Group identification number, number corresponding to Fig. 1, length of encounter, group size, and water depth data for 14 groups of Gulf of Mexico bottlenose dolphins. Group identification number indicates date (mmddyy_number), depth data rounded to nearest meter for spatial interpolation of National Geophysical Data Center Coastal Relief Model (hence minimum depth of 0 m for group 091408_2, see Sec. II).

Group id.	No. in Fig. 1	Encounter length (min)	Group size	Mean water depth (m)	SD depth (m)	Min depth (m)	Max depth (m)
040708_2	1	50	15	23.6	0.56	23	25
040908_2	2	38	9	30.1	0.50	29	31
041008_5	3	45	6	17.2	0.40	17	18
062008_2	4	33	6	2.8	0.39	2	3
071108_2	5	63	15	6.9	1.55	4	9
071808_6	6	63	4	6.6	0.73	5	7
072108_2	7	21	2	13.0	0.00	13	13
072108_3	8	9	2	12.0	0.00	12	12
072108_4	9	31	4	21.8	1.00	21	23
072308_5	10	12	2	10.0	0.00	10	10
072308_7	11	29	8	19.6	0.50	19	20
072808_2	12	7	2	10.0	0.00	10	10
080108_2	13	23	4	5.0	0.00	5	5
091408_2	14	36	5	2.4	1.17	0	4
Total		460			8.84	0	31
Mean		32.9	6.0	12.9			
Range		7–63	2–15	2.4–30.1			
SD		18.03	4.40				
SE				2.25			

munication, not target detection (e.g., Lammers *et al.*, 2003), they were not used in this analysis.

Individual wav files of echolocation click trains were imported into MATLAB (version 2007b, Mathworks, Inc., Natick, MA) for further analysis in a custom signal processing program. Files were high-pass filtered (20 kHz) to eliminate boat noise (which was considerable in some files). The resulting signal was rectified and then enveloped using a low pass filter (500 Hz) in order to facilitate the automatic identification of peaks and reduce false detections. Enveloped peaks equal to or greater than 1.6 times the root mean square (rms) amplitude of the file were labeled as potential echolocation peaks. The accuracy of click detection was verified manually by inspecting plots of the click train that were overlaid with the detected clicks. Only files (or segments of files) with accurately identified echolocation peaks from single dolphin click trains were used in analysis. The pulse periods (and the corresponding pulse rates) were determined for each pair of identified echolocation clicks and measured as the distances between amplitude maxima of the enveloped signals.

The GPS positions collected during each encounter were imported into ARCGIS (version 9.2, ESRI, Redlands, CA) in order to calculate the depths at each location. For each GPS position, the interpolated depth (to the nearest meter) was determined, and the mean water depth was calculated for each group (bathymetry data from the National Geophysical Data Center Coastal Relief Model; Divins and Metzger, 2009).

Statistical analysis was carried out using MYSTAT (version 12, Systat Inc., Chicago, IL) and MATLAB (2007b, Mathworks Inc.). The mean pulse rate for each group was determined both by using all pulse rate values and by binning pulse rate data by echolocation click train. The latter method was used to reduce the bias from pulse rate interdependence within click trains and from over-representation by exceptionally long click trains. High-occurrence mode pulse rates were determined for each group from all pulse rate values. These pulse rate modes were identified as peaks in the pulse rate histograms separated from adjacent peaks by decreases in occurrence greater than half their magnitude. The maximum value rounded to the nearest integer was used as the mode value. These mode values, representing the high-occurrence values in pulse rates, were converted to theoretical distance to target values using the two-way acoustic travel time and the dolphin's lag time using the following formula:

$$d = (((1/\mathrm{pr}) - t)/2)^{*}c,$$

where d is estimated distance (m), pr is pulse rate (Hz), t is lag time (s), and c is speed of sound (m/s). A speed of sound estimate of 1534 m/s was based on the formula of MacKenzie (1981) using typical temperature and salinity values for the study area [25 °C, 34.5 PSU (Practical Salinity Units)] and 10-m depth. The lag time used for estimated distance calculations was determined by an iterative process and was defined as the maximum time value necessary to calculate distances of 0.25 m or greater (similar to methods of Akamatsu *et al.*, 2005). The minimum distance of 0.25 m was chosen as a reasonable estimate of short-range bottlenose dolphin echolocation based on results of echolocation studies on trained animals (e.g., Evans and Powell, 1967). To determine the relationship between pulse rate and depth, linear regression models were used for the mean pulse rates (calculated from all data and from data binned by click train) versus mean depth for each group, and the principal (greatest

TABLE II. Group identification number, number of pulses analyzed, and pulse rate data for 14 groups of Gulf of Mexico bottlenose dolphins. Group identification number indicates date (mmddyy_number). For mean, SD, and minimum and maximum pulse rates, first value indicates mean calculated using all pulse rate values and second value indicates mean calculated using pulse rates binned by echolocation click train. High-occurrence modes in pulse rate data rounded to nearest integer and ordered within group by decreasing level of occurrence. Corresponding theoretical target distances for each mode based on two-way travel time at 1534 m/s+5.29 ms lag time. Principal (highest magnitude) mode used in regression analysis in bold.

Group id.	Mean water depth (m)	Pulses	Click trains	Mean pulse rate (Hz)	SD. pulse rate (Hz)	Min pulse rate (Hz)	Max pulse rate (Hz)	Mode pulse rate (Hz)	Theoretical distance (m)
040708_2	23.6	259	37	29.48/19.62	14.359/10.903	5.51/6.77	64.84/57.06	**20**	34.3
040908_2	30.1	123	34	19.63/19.09	6.653/7.204	7.25/7.63	36.97/36.97	**26**	25.4
041008_5	17.2	350	31	42.29/47.42	46.721/57.431	7.10/9.87	183.38/182.86	**22**	30.8
								51	11.0
								178	0.25
062008_2	2.8	1428	108	63.28/52.59	31.784/32.176	5.73/6.08	188.98/142.42	**68**	7.2
								10	72.6
071108_2	6.9	1797	134	54.33/38.95	31.320/25.121	5.22/6.56	188.61/139.50	**57**	9.4
071808_6	6.6	1065	71	66.82/50.85	36.332/36.511	5.73/6.78	170.82/156.61	**78**	5.8
								10	72.6
								156	0.9
072108_2	13.0	735	27	32.73/22.78	21.171/14.862	7.46/11.66	138.33/61.71	**12**	59.9
072108_3	12.0	441	29	42.85/37.13	15.665/15.994	4.63/4.92	98.06/73.17	**46**	12.6
072108_4	21.8	463	66	42.24/38.25	27.079/24.828	6.09/6.33	157.12/110.31	**28**	23.3
072308_5	10.0	535	71	50.67/44.97	20.880/23.895	5.89/8.50	111.89/105.51	**68**	7.2
072308_7	19.6	1123	66	37.36/25.83	23.351/19.669	3.84/4.52	160.54/84.50	**28**	23.3
								86	4.9
072808_2	10.0	371	40	48.05/33.85	24.905/24.949	4.18/6.01	108.9/87.827	**60**	8.7
								8	91.8
080108_2	5.0	1660	108	52.39/42.41	40.400/34.780	7.27/8.08	185.69/163.33	**12**	59.9
								68	7.2
091408_2	2.4	1624	78	48.13/43.26	31.119/32.467	4.29/8.56	179.10/144.52	**38**	16.1
								77	5.9
Total		11974	900	49.99/39.75	32.765/30.382	3.84/4.52	188.98/182.86		

magnitude) mode versus mean depth of each group, with Shapiro–Wilk tests for normality.

III. RESULTS

Fourteen groups with acoustic recordings were determined to be independent samples and were used in analysis (Fig. 1). In 5 out of 14 groups, independence was determined using photo-identification results; all of these groups were found within Tampa Bay, Boca Ciega Bay, or within approximately 2 km from shore in the Gulf of Mexico. The distance traveled method was used for the remaining nine groups. These groups were all recorded 10 km or farther from shore in the Gulf of Mexico. Sighting duration ranged from 7 to 63 min (mean=32.9 min, SD=18.0 min), and group size ranged from 2 to 15 dolphins (mean=6.0, SD=4.40, Table I). Mean depths ranged from 2.4 to 30.1 m (mean=12.9 m, SE=2.26 m, Table I). For these 14 groups, a total of 11 974 pulses in 900 click trains was accurately detected in 1238 acoustic files (123–1797 pulses per group, 27–134 click trains per group, Table II), resulting in 10 736 calculated pulse rates. Mean pulse rate calculated by using the combined pulse rate values for each group ranged from 19.6 to 66.8 Hz (overall mean=49.9 Hz, SD=32.77 Hz, Table II), while the mean pulse rate calculated by using the mean values for each click train ranged from 19.1 to 52.6 Hz (overall mean=39.8 Hz, SD=30.38 Hz, Table II).

Twenty-three modes in pulse rate were identified (1–3 per group, Fig. 2). The values of the pulse rate modes for each group, along with the corresponding mean depth for each group and theoretical acoustic two-way travel times for each mode, are listed in Table II. Pulse rate modes ranged between 8 and 156 Hz. In order to calculate distance to target estimates for the 23 pulse rate modes, a lag time of 5.29 ms was required. Estimated distance to target values for these modes ranged between the predetermined minimum of 0.25 m to a maximum of 91.8 m.

The linear regression models indicated a significant trend for mean pulse rate of groups to decrease with mean depth of groups, both for means calculated using combined data ($r^2=0.719$, $p<0.001$, Shapiro–Wilk test for normality: $p=0.814$, Fig. 3) and for means calculated using data pooled by click train ($r^2=0.528$, $p=0.003$, Shapiro–Wilk test for normality: $p=0.291$, Fig. 4). Although there was a general trend for the value of the principal (largest) pulse rate mode to decrease with mean depth, the slope was not significantly different than zero ($r^2=0.260$, $p=0.063$, Shapiro–Wilk test for normality: $p=0.929$, Fig. 5).

IV. DISCUSSION

The relationship between target distance and pulse rate has been thoroughly investigated using trained dolphins under controlled conditions and in most cases as a dolphin

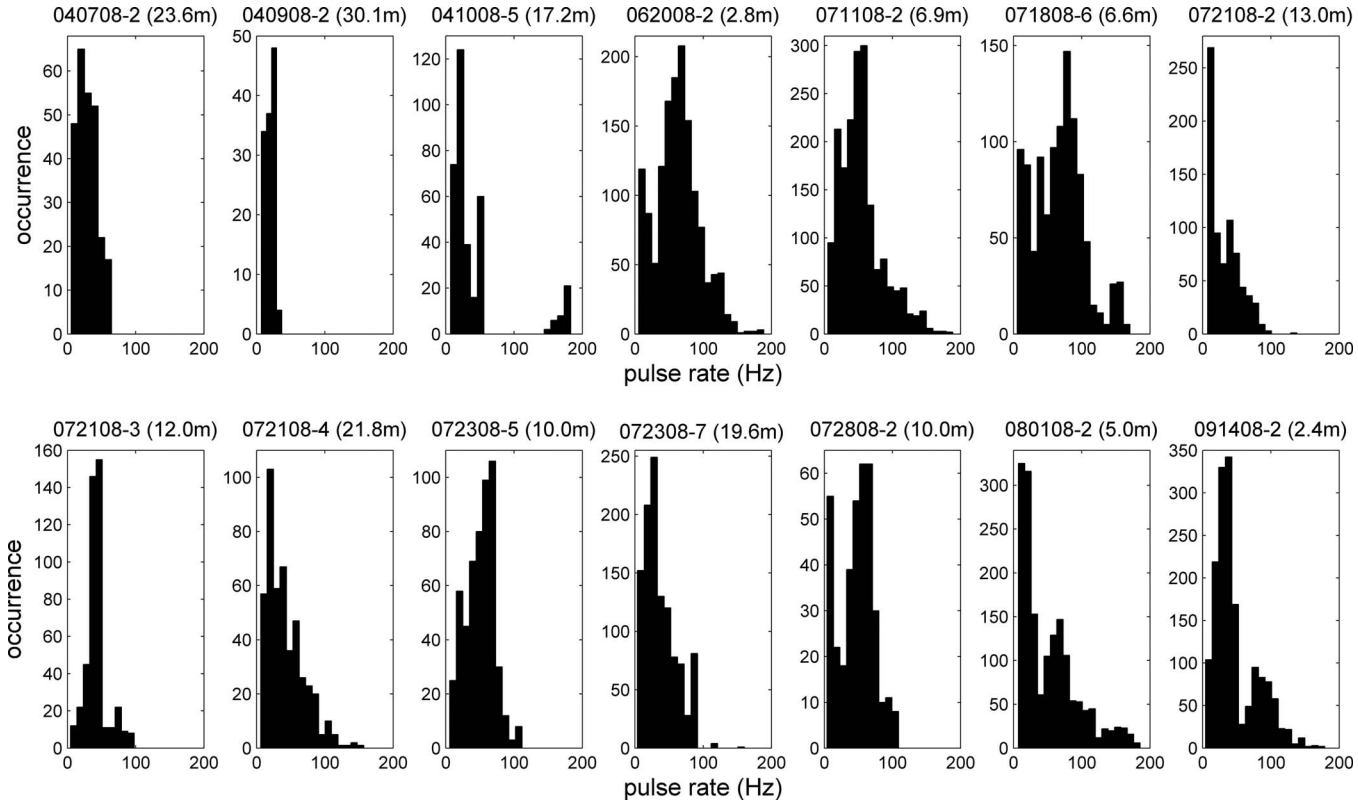


FIG. 2. Histograms (occurrence) of pulse rate data from 14 groups of bottlenose dolphins with mean water depth. Principal (largest) mode used in regression analysis. Histogram bar resolution of 10 Hz.

decreases its distance to the target, the echolocation pulse rate increases. This is thought to be a function of the two-way travel time of the echolocation click, plus a lag time, both of which must occur before the next click is produced (Au *et al.*, 1974). Several recent studies have determined that this relationship is also true for some free-ranging odontocetes (Akamatsu *et al.*, 2005; Jensen *et al.*, 2009). Other studies have investigated free-ranging odontocete echoloca-

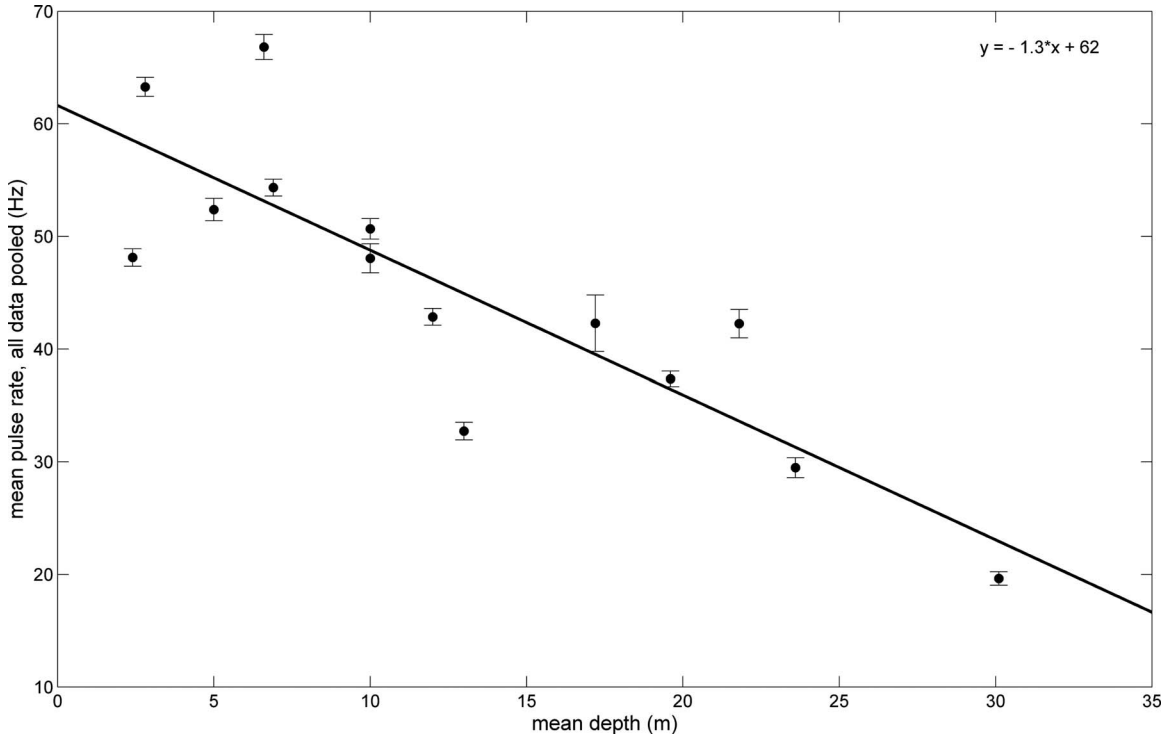


FIG. 3. Plot of mean pulse rate values (calculated using all pulse rate values) for 14 groups of bottlenose dolphins used in this study against mean depth of group with least-squares regression line. Error bars= ± 1 SE, r^2=0.719, and $p<0.001$.

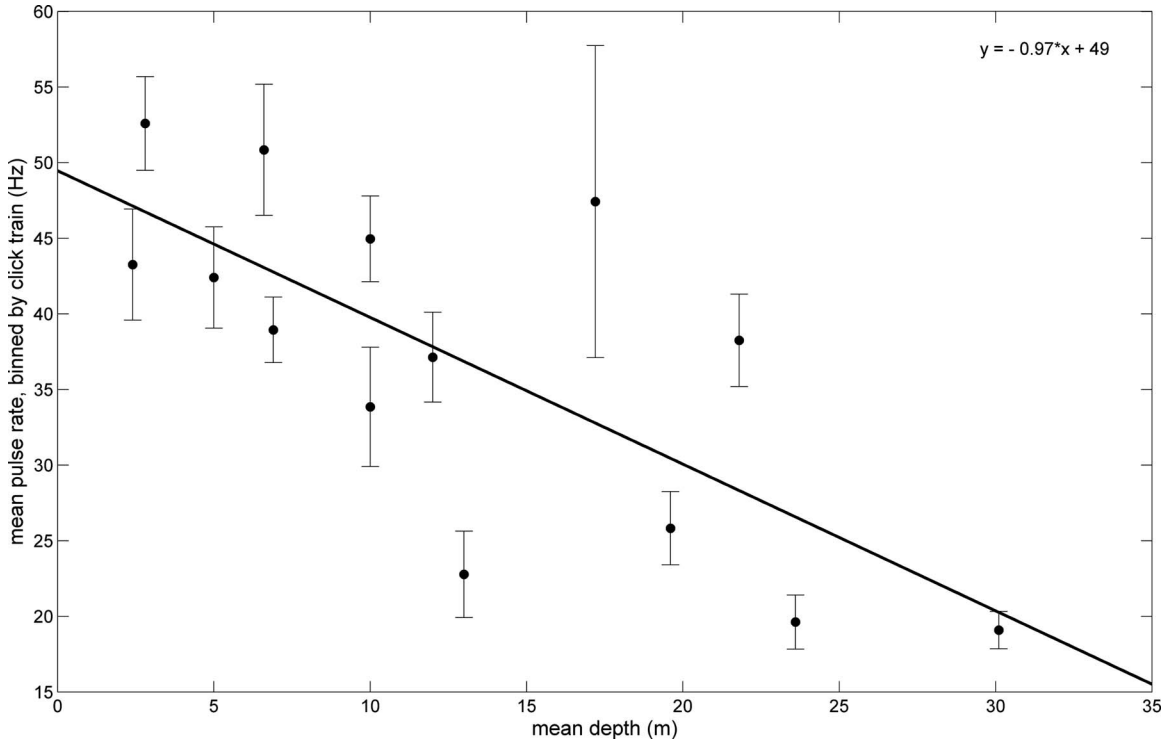


FIG. 4. Plot of mean pulse rate values (calculated using pulse rate values binned by echolocation click train) for 14 groups of bottlenose dolphins used in this study against mean depth of group with least-squares regression line. Error bars= ± 1 SE, r^2=0.528, and p=0.003.

tion click rates in relation to reasonably assumed foraging patterns (e.g., searching, final approach; Madsen *et al.*, 2002; Johnson *et al.*, 2004; Akamatsu *et al.*, 2005; Johnson *et al.*, 2008). Several studies have found that the initial pulse rates of sperm whales during foraging dives increased in relation to the depth in which they were foraging, suggesting a navigational function (Jaquet *et al.*, 2001; Thode *et al.*, 2002). However, this is the first analysis investigating the echolocation pulse rate of multiple groups of free-ranging delphinids in relation to depth, which could presumably be an important factor both in navigation and foraging.

Widely ranging pulse rates such as those found in this study (Table II) have previously been observed in free-ranging bottlenose dolphins. For example, Akamatsu *et al.*

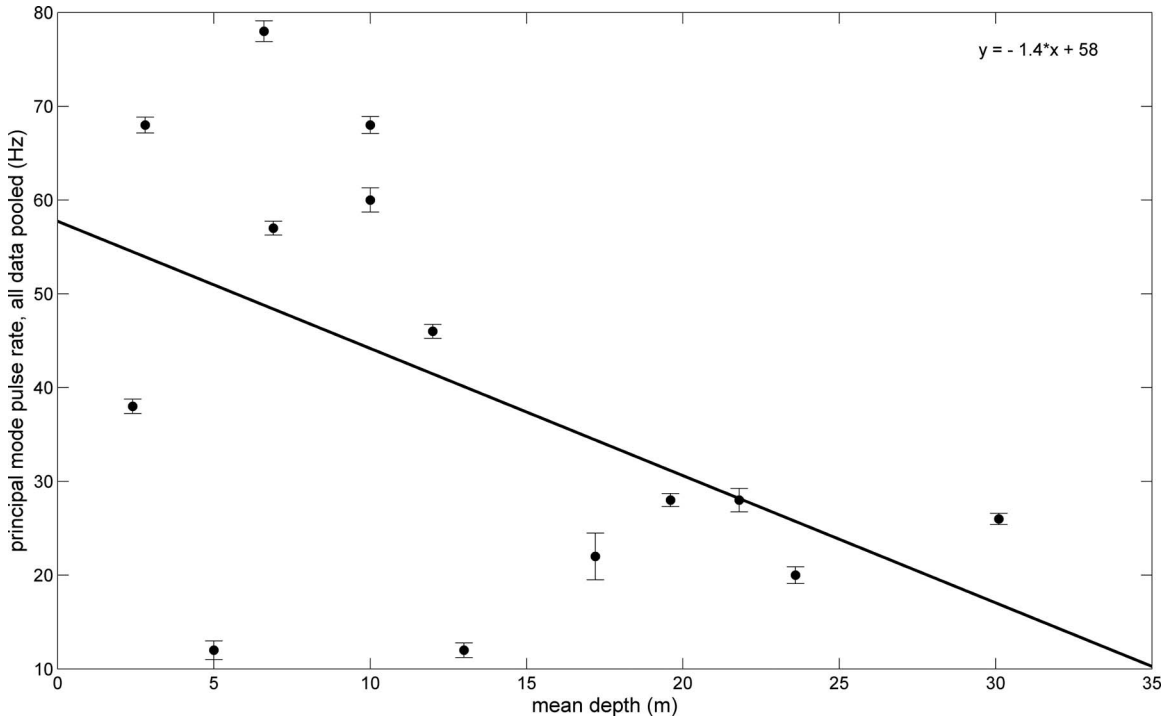


FIG. 5. Plot of principal (largest) pulse rate modes for 14 groups of bottlenose dolphins used in this study against mean depth of group with least-squares regression line. Error bars= ± 1 SE, r^2=0.260, and p=0.063.

(1998) found that echolocation inter-pulse intervals commonly ranged up to 200 ms (equivalent to 5-Hz pulse rate). Inter-pulse intervals ranged between 17 and 462 ms (equivalent to a pulse rate range of 59–2.2 Hz) for free-ranging bottlenose dolphins investigated by Jensen *et al.* (2009). Widely ranging pulse rates have also been observed in free-ranging baiji, finless porpoises (Akamatsu *et al.*, 1998), and Atlantic spotted dolphins (*Stenella frontalis*; Herzing, 1996). Group- and individual-specific differences in the use of echolocation will likely result in high variability of pulse rates (e.g., navigation and foraging on benthic versus pelagic prey) and will also likely confound the relationship between pulse rate and depth. This suggests that free-ranging dolphins are using overlapping pulse rates in a variety of depths and may indicate multiple purposes such as navigation, foraging on benthic prey, foraging on pelagic prey, and possibly communication (even though burst-pulses were omitted from the analysis).

Many of the groups in this study have several obvious modes in pulse rate occurrence (Fig. 2). When interpreting these modes and the associated distance to target estimates, we make three assumptions about dolphin echolocation and the echolocation observed in this study. First, we are assuming that these dolphins are timing their echolocation pulses based on the two-way travel time plus a lag time (e.g., Au *et al.*, 1974). Second, we are assuming that if a mode exists in the observed pulse rates, this pulse rate must be important to the animals (it is unlikely that pulse rates with little functionality will occur frequently). Thus, the peaks in these modes will be a representative value of important pulse rates. Third, we are assuming that all pulse rate modes are indeed used for target detection and ranging; by omitting burst-pulses from analysis we effectively omitted communicative pulsed signals. Most studies using trained animals estimate bottlenose dolphin lag times to be between 19 and 45 ms (e.g., Au *et al.*, 1974). A recent study by Jensen *et al.* (2009) demonstrated a lag time of 31 ms for free-ranging bottlenose dolphins, well within the ranges found in trained bottlenose dolphin studies. Therefore, it seems reasonable to use such lag time estimates for distance to target estimates. However, a 19 ms lag time resulted in reasonable (positive) distance to target estimates in just over half the pulse rate modes (12/23); the period between the clicks for the other modes was less than 19 ms. From the pulse rate modes presented here, a lag time of 5.29 ms is required to calculate distance to target estimates greater than or equal to 0.25 m. There are several possible explanations for this result. As most estimates of lag time are based on results obtained from trained animals who usually know the range to the target they must detect, it is possible that free-ranging odontocetes in natural foraging and navigation conditions have more flexible and potentially shorter lag times. Akamatsu *et al.* (2005) also calculated a lag time of 5 ms for free-ranging finless porpoises, while the lag time for trained harbor porpoises (a related species) was estimated to be 20–35 ms (Au *et al.*, 1999). There is some evidence for trained bottlenose dolphins to use very short lag times in short-range echolocation (Evans and Powell, 1967; Au, 1993), suggesting that the ability to process multiple clicks or use shorter lag times exists. In addition, Ivanov (2004) found that a trained bottlenose dolphin produced echolocation pulses at a faster rate than the two-way travel time when echolocating on objects at great distances (>600 m). These pulses were produced in bursts which did conform to the familiar two-way travel time plus lag time model (Ivanov, 2004). Similar click patterns have also been observed in a trained beluga during target detection (Turl and Penner, 1989) and in the echolocation of free-ranging false killer whales (*Pseudorca crassidens*) and Risso's dolphins (*Grampus griseus*; Madsen *et al.*, 2004). The results of studies such as Ivanov (2004) illustrate that our understanding of odontocete echolocation is not complete, and processes may vary greatly on a species, individual, and situational basis.

Distance estimates using the 5.29-ms lag time ranged from our minimum distance estimate of 0.25 m (group 041008_5, mean depth=17.2 m, pulse rate mode=178 Hz) to 91.8 m (group 072808_2, mean depth=10.0 m, pulse rate mode=8 Hz, Table II). To a first approximation this maximum estimated distance is similar to several other maximum effective range calculations for bottlenose dolphin echolocation. For example, the maximum range of echolocation in a noise limited environment for a natural prey item target using simulated dolphin pulses was determined to be 93 m (Au *et al.*, 2007), while a trained bottlenose dolphin was able to detect a metal sphere at a maximum distance of 113 m (Au and Snyder, 1980). Based on pulse rates, Akamatsu *et al.* (1998) estimated the maximum range of free-ranging bottlenose dolphin echolocation to be approximately 140 m. Our results also suggest that both long and short distance echolocations are used throughout the depth range in this study (Table II). For example, the shallow group 062008_2 (mean depth=2.8 m) had primary and secondary modes equivalent to 7.2 and 72.6 m, while the deep water group 041008_5 (mean depth=17.2 m) had modes equivalent to between the minimum of 0.25 and 30.8 m. Our distance estimates obviously must be interpreted with caution. These distance to target values are based on assumptions and are subject to errors in the calculation (e.g., inappropriate lag time). The distance estimates also may not reflect how these animals are actually using their echolocation. Pulse rate is not only determined by the distance to the target but by the difficulty of the echolocation task, the animal's expectations of finding a target, and whether a target is present or absent (Au, 1993). Animals within this study area are not exposed to dramatic bathymetry changes or deep depths, and some may be quite familiar with their habitats (see Nowacek, 2005); therefore there may be little need to echolocate at maximum distances. Penner (1988) found that dolphins used lower pulse rates in target absent tasks than when a known target was present. Therefore it is reasonable to postulate that free-ranging dolphins may echolocate at a slower rate than what would be expected from the two-way travel time plus lag time (which would inflate our distance estimates). Although target detection distances exceeding 100 m (e.g., Au and Snyder, 1980) to over 650 m (Ivanov, 2004) have been reported for bottlenose dolphins, the long distance values found in this study, especially in shallow water, should be regarded with additional caution. Shallow water is a particularly reverberant environment, and odontocete echolocation appears to be

negatively affected by reverberation (Au and Snyder, 1980; Au and Turl, 1983). Odontocetes likely use the information contained in the multiple returning echo highlights of the original pulse to discriminate objects (Au *et al.*, 1988, 2007). Echoes from reverberation return to an echolocating animal prior to echoes from potential targets of interest (Au *et al.*, 2007); therefore with increasing distance to target the loss of signal due to reverberation will increase (cf. Au and Snyder, 1980). At some distance the interference due to shallow water reverberation may therefore make the detection of targets impossible, and this distance is likely far less than the maximum range of dolphin echolocation. However, the maximum calculated distance in this study for shallow water groups (72.6-m distance in both 6.6- and 2.8-m mean water depths, Table II) is not unreasonable given the results of studies on trained animals in highly reverberant conditions (e.g., 113-m distance in water 5.8–6.1 m deep; Au and Snyder, 1980).

The results of the regression analysis indicate that mean pulse rate decreases with increasing depth. This supports the hypothesis that free-ranging bottlenose dolphins in this area are using echolocation for target detection and ranging, and the target distance is a function of water depth. This is an intuitively appealing result with regard to not only studies on trained odontocetes but also to an increasing number of free-ranging studies. The pulse rates of free ranging odontocetes were found to be slower than those of the same species in captive facilities, a relationship presumed to reflect target distance (Akamatsu *et al.*, 1998). The relationship between the pulse rates of on-axis echolocation clicks was found to increase with decreasing distance to the hydrophone target by Jensen *et al.* (2009). A similar relationship has been observed in the initial segments of sperm whale echolocation during descent to the foraging layer, where pulse rates were related to the depth of foraging or to the bottom (Jaquet *et al.*, 2001; Thode *et al.*, 2002; Zimmer *et al.*, 2003). The lack of significance in the relationship between the principal mode of pulse rate and mean depth may indicate that the outlier values in pulse rate are ecologically important. Therefore, although the mode is a more statistically robust measure of central tendency, the mean values (which are influenced by outliers to a greater degree) may be a better representation of bottlenose dolphin echolocation pulse rates. A potential bias which could complicate our analysis is the fact that as dolphin echolocation pulse rates increase, the amplitude tends to decrease (Au and Benoit-Bird, 2003; Jensen *et al.*, 2009). Therefore it is possible that in certain situations, such as dolphins being unusually distant from the hydrophone or sounds propagating poorly due to shallow water multipath effects, higher pulse rates could be undersampled. It is beyond the scope of this study to determine and quantify this bias.

The seasonal distribution of inshore resident bottlenose dolphins, which generally reside in restricted areas and those animals normally residing in the open Gulf of Mexico (analogous to "inshore" and "offshore/intermediate" types; Wells *et al.*, 1999), is generally well known for our study area. From April to September, resident dolphins in Sarasota Bay (immediately to the south of Tampa Bay) tend to remain within the core protected waters of the Bay, while from October to March, there is an increased tendency for the residents to range into the passes between the Bay and the open Gulf of Mexico and into the Gulf itself (Irvine *et al.*, 1981). This tendency has also been noted in Tampa Bay and Boca Ciega Bay (Weigle, 1987; S. Gowans, unpublished data). In addition, deeper water sightings (>15-m depth) often had noticeably larger animals with pseudo-stalked barnacles (possibly *Xenobalanus sp.*) attached to the caudal aspect of their dorsal fins (personal observation). Larger size and the presence of pseudo-stalked barnacles suggest that the animals in these groups were non-coastal animals (Barros, 1993; Wells *et al.*, 1999). Pseudo-stalked barnacles are only rarely seen attached to dolphins in Tampa Bay and Boca Ciega Bay and were not observed during any inshore sightings during this study (personal observation). Therefore, as this study was conducted from April to September, it is likely that groups recorded near Tampa Bay and Boca Ciega Bay are composed of inshore type dolphins, while groups recorded increasingly further into the open Gulf are composed of "intermediate" or "offshore" types (see Wells *et al.*, 1999). However, there does not appear to be an obvious division in the pulse rates between shallow water and deep water groups, instead the mean pulse rate values decrease in a linear manner with depth (Figs. 3 and 4). Likewise, both shallow and deep water groups produce pulse rate modes which suggest both short and long-range target detections (Table II). These results also suggest that the pulse rates of bottlenose dolphins in this area are primarily influenced by depth related processes, and the geographic differences observed in these dolphins may have little effect on their use of echolocation.

Although the results presented here indicate that echolocation clicks are produced with some function which is depth dependent, and that pulse rates potentially correspond to a wide variety of distances, we cannot determine the exact purposes of these echolocation clicks. Potential purposes include navigation or foraging for benthic, demersal, or pelagic prey. Nevertheless, this paper provides evidence that these dolphins alter the timing of their echolocation in relation to depth, which is likely related to navigation or foraging functions. The function of echolocation clicks in relation to depth will require additional attention in future studies of free-ranging dolphins.

ACKNOWLEDGMENTS

We wish to thank the captains of the *Eugenie Clark*, G. Byrd, D. Dougherty, and R. Welling, and of the *Fish Hawk*, A. Farmer, and our fearless pilot P. Gallizzi. C. Clark kindly allowed us to use the towed array aboard the *Eugenie Clark*. This paper was the result of many who helped in the field and took part in discussions, especially A. Davis, D. Greenhow, J. Locascio, K. Martin, M. McEachern, E. Montie, C. Murphy, M. Nelson, J. Powell, G. Stover, C. Wall, and S. Yagovane. P.S. was supported by the Paul L. Getting fellowship and the St. Petersburg Downtown Partnership fellowship. This project was supported by the National Oceanographic Partnership Program and Eckerd College. This manuscript was greatly improved by comments from W. Au,

P. Madsen, B. Würsig, and two anonymous reviewers. Field work was conducted under U.S. National Marine Fisheries Service LOC Contract Nos. 10038 and 1077-1794.

Akamatsu, T., Wang, D., Nakamura, K., and Wang, K. (**1998**). "Echolocation range of captive and free-ranging baiji (*Lipotes vexillifer*), finless porpoise (*Neophocaena phocaenoides*), and bottlenose dolphin (*Tursiops truncatus*)," J. Acoust. Soc. Am. **104**, 2511–2516.

Akamatsu, T., Wang, D., Wang, K., and Naito, Y. (**2005**). "Biosonar behaviour of free-ranging porpoises," Proc. R. Soc. London, Ser. B **272**, 797–801.

Allen, M. C., Read, A. J., Gaudet, J., and Sayigh, L. S. (**2001**). "Fine-scale habitat selection of foraging bottlenose dolphins *Tursiops truncatus* near Clearwater, Florida," Mar. Ecol.: Prog. Ser. **222**, 253–264.

Au, W. L. L., Floyd, R. W., Penner, R. H., and Murchison, A. E. (**1974**). "Measurement of echolocation signals of the Atlantic bottlenose dolphin, *Tursiops truncatus* Montagu, in open waters," J. Acoust. Soc. Am. **56**, 1280–1290.

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

Au, W. W. L., and Benoit-Bird, K. J. (**2003**). "Automatic gain control in the echolocation system of dolphins," Nature (London) **423**, 861–863.

Au, W. W. L., Benoit-Bird, K. J., and Kastelein, R. A. (**2007**). "Modeling the detection range of fish by echolocating bottlenose dolphins and harbor porpoises," J. Acoust. Soc. Am. **121**, 3954–3962.

Au, W. W. L., Floyd, R. W., and Haun, J. E. (**1978**). "Propagation of Atlantic bottlenose dolphin echolocation signals," J. Acoust. Soc. Am. **64**, 411–422.

Au, W. W. L., Kastelein, R. A., Rippe, T., and Schooneman, N. M. (**1999**). "Transmission beam pattern and echolocation signals of a harbor porpoise (*Phocoena phocoena*)," J. Acoust. Soc. Am. **106**, 3699–3705.

Au, W. W. L., Moore, P. W. B., and Pawloski, D. (**1986**). "Echolocation transmitting beam of the Atlantic bottlenose dolphin," J. Acoust. Soc. Am. **80**, 688–691.

Au, W. W. L., Moore, P. W. B., and Pawloski, D. (**1988**). "Detection of complex echoes in noise by an echolocating dolphin," J. Acoust. Soc. Am. **83**, 662–668.

Au, W. W. L., and Snyder, K. J. (**1980**). "Long-range target detection in open waters by an echolocating Atlantic bottlenose dolphin (*Tursiops truncatus*)," J. Acoust. Soc. Am. **68**, 1077–1084.

Au, W. W. L., and Turl, C. W. (**1983**). "Target detection in reverberation by an echolocating Atlantic bottlenose dolphin (*Tursiops truncatus*)," J. Acoust. Soc. Am. **73**, 1676–1681.

Barros, N. B. (**1993**). "Feeding ecology and foraging strategies of bottlenose dolphins on the central east coast of Florida," Ph.D. thesis, University of Miami, Miami, FL.

Barros, N. B., and Odell, D. K. (**1990**). "Food habits of bottlenose dolphins in the southeastern United States," in *The Bottlenose Dolphin*, edited by S. Leatherwood and R. R. Reeves (Academic, San Diego, CA), pp. 309–328.

Barros, N. B., and Wells, R. S. (**1998**). "Prey and feeding patterns of resident bottlenose dolphins (*Tursiops truncatus*) in Sarasota Bay, Florida," J. Mammal. **79**, 1045–1059.

Bradbury, J. W., and Vehrencamp, S. L. (**1998**). *Principles of Animal Communication* (Sinauer Associates, Sunderland, MA).

Divins, D. L., and Metzger, D. (**2009**). NGDC Coastal Relief Model, http://www.ngdc.noaa.gov/mgg/coastal/coastal.html (Last viewed January, 2009).

Evans, W. E., and Powell, B. A. (**1967**). "Discrimination of different metallic plates by an echolocating delphinid," in *Animal Sonar Systems, Biology and Bionics*, edited by R.-G. Busnel (NATO Advanced Study Institute, Frascati, Italy), Vol. **1**, pp. 363–383.

Fazioli, K. L., and Wells, R. S. (**1999**). "Stock structure of coastal bottlenose dolphins, *Tursiops truncatus*, near Sarasota, Florida," Report No. 40-WCNF701806, National Marine Fisheries Service, Southeast Fisheries Science Center, Miami, FL.

Gannon, D. P., Barros, N. B., Nowacek, D. P., Read, A. J., Waples, D. M., and Wells, R. S. (**2005**). "Prey detection by bottlenose dolphins, *Tursiops truncatus*: An experimental test of the passive listening hypothesis," Anim. Behav. **69**, 709–720.

Hersh, S. L., and Duffield, D. A. (**1990**). "Distinction between northwest Atlantic offshore and coastal bottlenose dolphins based on hemoglobin profile and morphometry," in *The Bottlenose Dolphin*, edited by S. Leatherwood and R. R. Reeves (Academic, San Diego, CA), pp. 129–139.

Herzing, D. L. (**1996**). "Vocalizations and associated underwater behaviour of free-ranging Atlantic spotted dolphins, *Stenella frontalis* and bottlenose dolphins, *Tursiops truncatus*," Aquat. Mamm. **22**, 61–79.

Irvine, A. B., Scott, M. D., Wells, R. S., and Kaufmann, J. H. (**1981**). "Movements and activities of the Atlantic bottlenose dolphin, *Tursiops truncatus*, near Sarasota, Florida," Fish. Bull. **79**, 671–688.

Ivanov, M. P. (**2004**). "Dolphin's echolocation signals in a complicated acoustic environment," Acoust. Phys. **50**, 550–561.

Jaquet, N., Dawson, S., and Douglas, L. (**2001**). "Vocal behavior of male sperm whales: Why do they click?," J. Acoust. Soc. Am. **109**, 2254–2259.

Jensen, F. H., Bejder, L., Wahlberg, M., and Madsen, P. T. (**2009**). "Biosonar adjustments to target range of echolocating bottlenose dolphins (*Tursiops* sp.) in the wild," J. Exp. Biol. **212**, 1078–1086.

Johnson, M., Hickmott, L. S., Aguilar Soto, N., and Madsen, P. T. (**2008**). "Echolocation behaviour adapted to prey in foraging Blainville's beaked whale (*Mesoplodon densirostris*)," Proc. R. Soc. London, Ser. B **275**, 133–139.

Johnson, M., Madsen, P. T., Zimmer, W. M. X., de Soto, N. A., and Tyack, P. L. (**2004**). "Beaked whales echolocate on prey," Proc. R. Soc. London, Ser. B **271**, S383–S386.

Lammers, M. O., Au, W. W. L., and Herzing, D. L. (**2003**). "The broadband social acoustic signaling behavior of spinner and spotted dolphins," J. Acoust. Soc. Am. **114**, 1629–1639.

MacKenzie, K. V. (**1981**). "Nine-term equation for sound speed in the ocean," J. Acoust. Soc. Am. **70**, 807–812.

Madsen, P. T., Johnson, M., Aguilar de Soto, N., Zimmer, W. M. X., and Tyack, P. (**2005**). "Biosonar performance of foraging beaked whales (*Mesoplodon densirostris*)," J. Exp. Biol. **208**, 181–194.

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. Exp. Biol. **207**, 1811–1823.

Madsen, P. T., Wahlberg, M., and Møhl, B. (**2002**). "Male sperm whale (*Physeter macrocephalus*) acoustics in a high-latitude habitat: Implications for echolocation and communication," Behav. Ecol. Sociobiol. **53**, 31–41.

Mate, B. R., Rossbach, K. A., Nieukirk, S. L., Wells, R. S., Irvine, A. B., Scott, M. D., and Read, A. J. (**1995**). "Satellite-monitored movements and dive behavior of a bottlenose dolphin (*Tursiops truncatus*) in Tampa Bay, Florida," Marine Mammal Sci. **11**, 452–463.

Miller, P. J. O., Johnson, M. P., and Tyack, P. L. (**2004**). "Sperm whale behaviour indicates the use of echolocation click buzzes 'creaks' in prey capture," Proc. R. Soc. London, Ser. B **271**, 2239–2247.

Moore, P. W., Dankiewicz, L. A., and Houser, D. S. (**2008**). "Beamwidth control and angular target detection in an echolocating bottlenose dolphin (*Tursiops truncatus*)," J. Acoust. Soc. Am. **124**, 3324–3332.

Nowacek, D. P. (**2005**). "Acoustic ecology of foraging bottlenose dolphins (*Tursiops truncatus*), habitat-specific use of three sound types," Marine Mammal Sci. **21**, 587–602.

Penner, R. H. (**1988**). "Attention and detection in dolphin echolocation," in *Animal Sonar: Processes and Performance*, edited by P. E. Nachtigall and P. W. B. Moore (Plenum, New York), pp. 709–713.

Sellas, A. B., Wells, R. S., and Rosel, P. E. (**2005**). "Mitochondrial and nuclear DNA analyses reveal fine scale geographic structure in bottlenose dolphins (*Tursiops truncatus*) in the Gulf of Mexico," Cons. Gen. **6**, 715–728.

Soldevilla, M. S., Henderson, E. E., Campbell, G. S., Wiggins, S. M., and Hildebrand, J. A. (**2008**). "Classification of Risso's and Pacific white-sided dolphins using spectral properties of echolocation clicks," J. Acoust. Soc. Am. **124**, 609–624.

Thode, A., Mellinger, D. K., Stienessen, S., Martinez, A., and Mullin, K. (**2002**). "Depth-dependent acoustic features of diving sperm whales (*Physeter macrocephalus*) in the Gulf of Mexico," J. Acoust. Soc. Am. **112**, 308–321.

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.

Turl, C. W., and Penner, R. H. (**1989**). "Differences in echolocation click pattern of the beluga (*Delphinapterus leucas*) and the bottlenose dolphin (*Tursiops truncatus*)," J. Acoust. Soc. Am. **86**, 497–502.

Verfuß, U. K., Miller, L. A., Pilz, P. K. D., and Schnitzler, H.-U. (**2009**). "Echolocation by two foraging harbour porpoises (*Phocoena phocoena*)," J. Exp. Biol. **212**, 823–834.

Watkins, W. A. (**1967**). "The harmonic interval: Fact or artifact in spectral analysis of pulse trains," in *Marine Bioacoustics*, edited by W. N. Tavolga

(Pergamon, Oxford), pp. 15–43.
Weigle, B. L. (**1987**). "Abundance, distribution, and movements of bottlenose dolphins, *Tursiops truncatus*, in lower Tampa Bay, Florida," MS thesis, University of South Florida, Tampa, FL.
Wells, R. S., Rhinehard, H. L., Cunningham, P., Whaley, J., Baran, M., Koberna, C., and Costa, D. P. (**1999**). "Long distance offshore movements of bottlenose dolphins," Marine Mammal Sci. **15**, 1098–1114.
Wells, R. S., and Scott, M. D. (**1990**). "Estimating bottlenose dolphin population parameters from individual identification and capture-release techniques," Rep. Int. Whal. Comm. **12**, 407–515.
Zimmer, W. M. X., Johnson, M. P., D'Amico, A., and Tyack, P. (**2003**). "Combining data from a multisensory tag and passive sonar to determine the diving behavior of a sperm whale (*Physeter macrocephalus*)," IEEE J. Ocean. Eng. **28**, 13–28.

Utilizing dual frequency band transmit pulse complexes in medical ultrasound imaging

Rune Hansen[a)]
Department of Medical Technology, SINTEF Technology and Society, N 7465 Trondheim, Norway

Svein-Erik Måsøy, Tonni F. Johansen, and Bjørn A. Angelsen
Department of Circulation and Medical Imaging, Norwegian University of Science and Technology, N 7489 Trondheim, Norway

(Received 30 March 2009; revised 29 July 2009; accepted 6 October 2009)

A method of acoustic imaging that potentially can improve the diagnostic capabilities of medical ultrasound is presented. The method, given the name SURF (Second order UltRasound Field) imaging, is achieved by processing the received signals from transmitted dual frequency band pulse complexes with at least partly overlapping high frequency (HF) and low frequency (LF) pulses. The transmitted HF pulses are used for image reconstruction, whereas the transmitted LF pulses are used to manipulate the elastic properties of the medium observed by the HF imaging pulses. The present paper discusses fundamental concepts in relation to the use of dual frequency band pulse complexes for medical ultrasound imaging. © *2010 Acoustical Society of America.* [DOI: 10.1121/1.3257584]

PACS number(s): 43.80.Qf, 43.25.Zx [CCC] Pages: 579–587

I. INTRODUCTION

The importance of nonlinear acoustics within the field of medical ultrasound imaging was largely neglected until 1980 when Muir and co-workers[1,2] focused the attention toward this topic. Two papers predicted and demonstrated nonlinear acoustical effects relevant for intensities and frequencies common in biomedical imaging. During the past 25 years, there has been an increasing interest with respect to nonlinear biomedical acoustics. This interest was further escalated by the introduction of ultrasound contrast agents and the study of these contrast agents was probably the main impetus for the introduction of the tissue harmonic imaging technique.

Regardless of the interest with respect to nonlinear acoustics, this field has found relatively limited applications within everyday clinical ultrasonic examinations. However, one application is, as indicated, tissue harmonic imaging which is used with good results for imaging of some organs.[3–9] This technique utilizes the accumulative nonlinear distortion of the forward propagating transmit pulse. With conventional transmit pulses, this distortion mainly results in the build-up of a second harmonic band which then is linearly back-scattered from tissue inhomogeneities. Second harmonic imaging typically facilitates reduced acoustical noise due to multiple scattering originating from structures close to the ultrasound transducer. The trade-off with this technique is reduced sensitivity and thus reduced penetration.

The other main application where nonlinear acoustics has been utilized is with respect to imaging of ultrasound contrast agents. Ultrasound contrast agents are typically made as solutions of gas microbubbles in a fluid that can be injected into a vein to increase the scattering from blood which is weak compared to the scattering from soft tissues[10] (Table 4.22).

The gas bubble has high compliance relative to the surrounding blood and scattering from the microbubble is resonant through an interaction between the co-oscillating fluid mass around the bubble and the bubble compression elasticity.[11] When driven by medical ultrasound pulses with frequencies in the vicinity of the gas microbubble resonance frequency, the local nonlinear scattering is typically much larger than from soft tissues.[12–14] This has resulted in several nonlinear ultrasound contrast agent detection techniques with the purpose to suppress the linear part of a received signal while maintaining as much as possible of the nonlinear part of a received signal.[15–21]

Reconstruction of medical ultrasound images from backscattered echoes are today based on two main acoustical assumptions that generally are not fulfilled. In many practical imaging situations, image quality suffers due to these assumptions.

The first assumption is that the ultrasound wave is propagating with a constant speed of sound which typically is set equal to 1540 m/s. Many organs show small variations in the speed of sound and the indicated assumption is then good. In some organs, e.g., the breast, and especially in the body wall, there are, however, larger variations in acoustic properties resulting in aberrations of the acoustic wavefronts. This variable acoustic velocity along the propagating wave-front destroys the transmit and receive beams resulting in reduced focusing of beam main-lobe and increase in beam side-lobes. The reduced focusing of the beam main-lobe reduces the spatial resolution in the acoustic imaging system. The increase in beam side-lobes introduces additive noise in the image, reducing the ratio of the strongest to the weakest

[a)]Author to whom correspondence should be addressed. Electronic mail: rune.hansen@sintef.no. Also at Department of Circulation and Medical Imaging, Norwegian University of Science and Technology.

scatterer that can be detected in the neighborhood of each other, defined as the contrast resolution in the image.

The second assumption is that the Born approximation is valid and multiple scattering is hence neglected. Interfaces between tissue components with large differences in acoustic properties can give so strong first order scattering from the transmitted acoustic pulse that higher order multiple scattering get large amplitudes. Such multiple scatterings are termed pulse reverberations. As with the increase in beam side-lobes, these reverberations reduce the contrast resolution in the ultrasound image. Reduced contrast resolution is, in particular, a problem when imaging hypo-echoic structures such as the heart chambers, the lumen of large blood vessels, some atherosclerotic lesions, cysts, some tumors, as well as in fetal imaging. The contact interface between the ultrasound transducer itself and the soft tissue is also a strong reflector enhancing this problem with multiple scattering.

Deviations from these two main assumptions will vary in extent both depending on the organ imaged on a given person and among different persons for a given organ. Discrepancies between the indicated assumptions in an ultrasound scanner and actual acoustical properties of the object being imaged are responsible for the significant patient variability in image quality typically seen with medical ultrasound imaging today. The most extreme example of discrepancies between indicated assumptions and actual acoustical properties is experienced when there is bone or gas in front of or within the object to be imaged. Image quality is in this case typically either totally destroyed or severely reduced unless the gas or bone is small in size relative to the wavelength applied, e.g., as with ultrasound contrast bubbles.

In some imaging situations, conventional ultrasound images do not have adequate contrast resolution between normal tissue and pathological tissue, even in the absence of image degrading effects from phase-front aberrations and pulse reverberations. This can be the case for imaging of various tumors (for example, prostate tumors), for robust imaging of micro-calcifications in breast pathology and for diagnosis of atherosclerotic lesions. Tissue contrast can in some situations be increased by adding ultrasound contrast agents in the form of microbubbles opening for assessment of tissue micro-circulation. For imaging of ultrasound contrast agents, in general, current techniques also have important limitations.[22,23]

The present paper analyzes a method of acoustic imaging, referred to as SURF (Second order UltRasound Field) imaging, where pulse complexes composed of a low frequency (LF) and a high frequency (HF) pulse are transmitted. The LF to HF ratio is, for example, $\sim 1:10$, and the LF pulse is used to manipulate the nonlinear material elasticity observed by the co-propagating HF pulse, which is used for image reconstruction. The SURF method aims to utilize nonlinear and resonant acoustical effects for improved image reconstructions of objects.

The use of dual band transmit pulses has previously been suggested and reported in the literature for nonlinear imaging of tissue and contrast agents.[24–29] The authors have also previously described the SURF method for imaging of ultrasound contrast agents.[22,23,30]

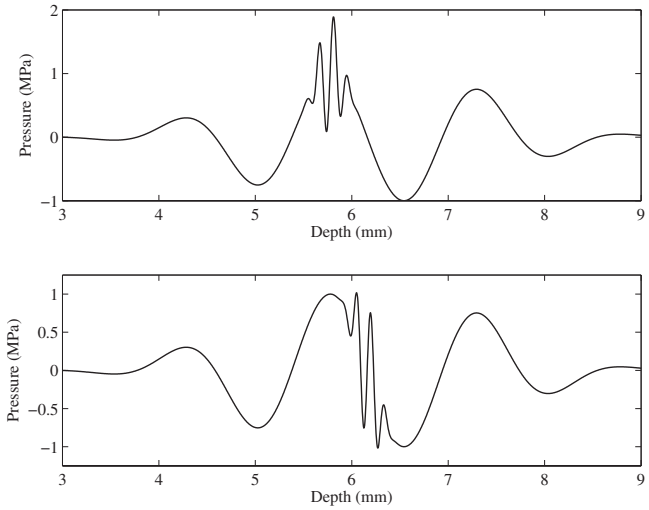


FIG. 1. Example of SURF transmit pulse complexes. The HF imaging pulse is placed on the positive pressure peak of the LF manipulation pulse (upper panel) and at the maximal negative spatial pressure gradient of the LF manipulation pulse (lower panel).

Figure 1 shows an example of two transmitted pulse complexes. In the upper panel, the HF pulse, i.e., the imaging pulse, is placed on the positive pressure peak of the LF pulse, i.e., the manipulation pulse. In the lower panel, the HF imaging pulse is placed at the maximal negative spatial pressure gradient of the LF manipulation pulse. For the pulse complexes shown here, the HF imaging pulse has a center frequency of 10 MHz, while the LF manipulation pulse has a center frequency of 1 MHz.

II. THEORY

The SURF method utilizes nonlinear acoustical effects for improved image reconstruction and to analyze this method, the nonlinear heterogeneous wave equation is first derived. Using a Lagrange coordinate description, the law for conservation of momentum can be written as[31] (Chap. 12)

$$\rho \frac{\partial \vec{u}}{\partial t} = -\nabla p(1 + \nabla \vec{\psi}) + (\nabla p \cdot \nabla)\vec{\psi}, \tag{1}$$

where ρ is mass density, p is pressure, and where the velocity, $\vec{u}$, is given by the time derivative of the particle displacement, $\vec{\psi}$,

$$\vec{u} = \frac{\partial \vec{\psi}}{\partial t}. \tag{2}$$

The force terms on the right-hand side of Eq. (1) are a nonlinear mixture of pressure gradient and displacement components. However, for plane waves, these terms reduce to $-\nabla p$. This is also a good approximation for ultrasound beams, where the radius of curvature of the wave-fronts are large relative to the wavelength, and the approximation is used in the following so that

$$\rho \frac{\partial \vec{u}}{\partial t} = -\nabla p. \tag{3}$$

For liquids and solids, including soft tissues, the nonlinear material elastic compressibility can typically be approximated to the second order in the acoustic pressure. An isentropic volume compression, δV, of a small isotropic volume element, ΔV, is then related to the pressure as[31,32]

$$\frac{\delta V}{\Delta V} = -\nabla\vec{\psi} = (1 - \beta_n \kappa p)\kappa p, \tag{4}$$

where κ is the linear bulk compressibility and β_n is a nonlinear elasticity parameter which can be found experimentally for various materials.

In ultrasound imaging of soft tissues, the volume compressions are, however, typically not isentropic. An additional term, taking care of conversion of acoustic energy to heat, must then be added to Eq. (4), and a constitutive nonlinear material equation for an isotropic material undergoing nonisentropic compression may be written as[31]

$$-\nabla\vec{\psi} = (1 - \beta_n \kappa p)\kappa p + h * \kappa p, \tag{5}$$

where h in the convolution term is a low-pass filter taking care of frequency dependent acoustic absorption. By combining Eqs. (3) and (5), one obtains

$$\nabla\left(\frac{1}{\rho}\nabla p\right) - \kappa\frac{\partial^2 p}{\partial t^2} + \beta_n \kappa^2 \frac{\partial^2 p^2}{\partial t^2} - h * \kappa\frac{\partial^2 p}{\partial t^2} = 0. \tag{6}$$

With compression waves only, i.e., assuming zero shear strain, it is common to introduce the momentum potential

$$\nabla\phi = -\rho\vec{u}. \tag{7}$$

From Eq. (3), one notes that

$$p = \frac{\partial\phi}{\partial t} \tag{8}$$

and Eq. (6) is rearranged to

$$\nabla\left(\frac{1}{\rho}\nabla\phi\right) - \kappa\left(1 - 2\beta_n \kappa\frac{\partial\phi}{\partial t}\right)\frac{\partial^2\phi}{\partial t^2} - h * \kappa\frac{\partial^2\phi}{\partial t^2} = 0. \tag{9}$$

Tissue is a mixture of different components such as muscle, fat, connective tissue, and parenchyma with different acoustic properties. Organs are typically groups of tissues performing a specific function or group of functions. On a smaller scale, the various tissue components are themselves heterogeneous consisting of different cells which again can be differentiated into various organelles. Biological tissue is hence heterogeneous on several scales, for example, ranging from several centimeters in some organs to less than a micrometer for some organelles. Variations in mass density and compressibility are responsible for back-scattering of acoustic energy.

By separating all material parameters into slow and rapid spatial variations relative to the wavelength applied, one can establish a heterogeneous wave equation where terms representing accumulative forward effects are grouped on the left-hand side of the equation, whereas terms accounting for local back-scattering are grouped on the right-hand side of the equation. For example, the mass density is then written $\rho = \rho_a + \rho_f$, where the subscript a denotes a slowly varying average and the subscript f denotes more rapidly varying fluctuations around this average. If it then is assumed that $\nabla\rho_a = 0$, one may write the inverse of the mass density as

$$\frac{1}{\rho} = \frac{\rho_a}{\rho\rho_a} = \frac{\rho - \rho_f}{\rho\rho_a} = \frac{1}{\rho_a} - \frac{\rho_f}{\rho\rho_a}. \tag{10}$$

From Eq. (9), the following equation is then derived:

$$\frac{1}{\rho_a}\nabla^2\phi - \kappa_a\frac{\partial^2\phi}{\partial t^2} - h * \kappa_a\frac{\partial^2\phi}{\partial t^2} + 2\beta_{na}\kappa_a^2\frac{\partial\phi}{\partial t}\frac{\partial^2\phi}{\partial t^2}$$
$$= \kappa_f\frac{\partial^2\phi}{\partial t^2} + h * \kappa_f\frac{\partial^2\phi}{\partial t^2} + \frac{1}{\rho_a}\nabla(\gamma\nabla\phi) - 2\sigma_n\beta_{na}\kappa_a^2\frac{\partial\phi}{\partial t}\frac{\partial^2\phi}{\partial t^2}, \tag{11}$$

where

$$\beta_n\kappa^2 = (\beta_{na} + \beta_{nf})(\kappa_a + \kappa_f)^2 = \beta_{na}\kappa_a^2(1 + \alpha_n)(1 + \sigma_l)^2$$
$$= \beta_{na}\kappa_a^2 + \sigma_n\beta_{na}\kappa_a^2, \tag{12}$$

and where

$$\alpha_n = \frac{\beta_{nf}}{\beta_{na}}, \quad \sigma_l = \frac{\kappa_f}{\kappa_a}, \quad \gamma = \frac{\rho_f}{\rho} \tag{13}$$

and

$$\sigma_n = \alpha_n + \sigma_l(1 + \alpha_n)(2 + \sigma_l). \tag{14}$$

It is now possible to rearrange Eq. (11) to obtain the heterogeneous nonlinear wave equation

$$\nabla^2\phi - \frac{1}{c_0^2}\frac{\partial^2\phi}{\partial t^2} - h * \frac{1}{c_0^2}\frac{\partial^2\phi}{\partial t^2} + 2\beta_{na}\kappa_a p\frac{1}{c_0^2}\frac{\partial^2\phi}{\partial t^2}$$
$$= \sigma_l\left(\frac{1}{c_0^2}\frac{\partial^2\phi}{\partial t^2} + h * \frac{1}{c_0^2}\frac{\partial^2\phi}{\partial t^2}\right)$$
$$+ \nabla(\gamma\nabla\phi) - 2\sigma_n\beta_{na}\kappa_a p\frac{1}{c_0^2}\frac{\partial^2\phi}{\partial t^2}, \tag{15}$$

where the linear propagation velocity, c_0, is defined as

$$c_0 = \frac{1}{\sqrt{\rho_a\kappa_a}}. \tag{16}$$

In Eq. (15), the terms grouped on the left-hand side now represent forward propagation effects, while the terms grouped on the right-hand side represent local back-scattering. The two first terms on the left-hand side represent the well-known d'Alembert operator accounting for linear diffraction effects, whereas the third term accounts for frequency dependent power absorption. The last term on the left-hand side represents a nonlinear propagation term accounting for accumulative nonlinear forward effects. By grouping the second and last terms on the left-hand side, it is noticed that a nonlinear pressure dependent propagation velocity, c, is given as

$$c = \frac{c_0}{\sqrt{1 - 2\beta_{na}\kappa_a \frac{\partial \phi}{\partial t}}} = \frac{c_0}{\sqrt{1 - 2\beta_{na}\kappa_a p}} \approx c_0(1 + \beta_{na}\kappa_a p) \tag{17}$$

where the last approximation is valid for the case when $\kappa_a p \ll 1$ which is typically the case for medical ultrasound imaging. The two first terms on the right-hand side of Eq. (15) account for linear scattering due to fluctuations in compressibility and mass density, respectively, whereas the last term on the right-hand side is a term representing local nonlinear back-scattering.

The LF manipulation pulses in Fig. 1 will affect the material elasticity and thus also the propagation velocity for the HF pulses along their propagation path. Using the approximation from Eq. (17), the modified propagation velocity of the HF pulse, due to the presence of the LF pulse, is

$$c(\vec{r}, p_{\mathrm{LF}}) = c_1(\vec{r})(1 + \beta_{na}(\vec{r})\kappa_a(\vec{r})p_{\mathrm{LF}}), \tag{18}$$

where $\vec{r}$ is the spatial position vector, c_1 is the propagation velocity observed by the HF pulse with no LF manipulation pulse, and p_{LF} is the LF manipulation pressure.

III. DISCUSSION

With transmit pulse complexes, as shown in Fig. 1, the LF manipulation pressure will affect both the forward propagation of the HF imaging pulse, through the last term on the left-hand side of Eq. (15), and the back-scattering of the HF imaging pulse, through the last term on the right-hand side of Eq. (15), for a medium with nonlinear elasticity. However, the LF manipulation pressure will typically drop so much in the back-scattered direction that nonlinear effects on the HF imaging pulse are negligible.

A. Nonlinear propagation

With dynamically focused HF receive beams, as used in medical ultrasound imaging, one does in back-scatter imaging only observe the object at a narrow region around the beam axis. The propagation lag of the HF pulse along the beam axis may in this situation be approximated as

$$t(r) = \int_{\Gamma(r)} \frac{ds}{c(s, p_{\mathrm{LF}}(s))} = t_0(r) + \tau_n(r), \tag{19}$$

where $\Gamma(r)$ is the propagation path of the HF pulse to a scatterer at depth r and the coordinate s denotes the ray location along the HF pulse at any time. Then

$$t_0(r) = \int_{\Gamma(r)} \frac{ds}{c_1(s)} \tag{20}$$

is the propagation lag without LF manipulation, i.e., for conventional imaging, and

$$\tau_n(r) = -\int_{\Gamma(r)} \frac{ds}{c_1(s)} \beta_{na}(s)\kappa_a(s)p_{\mathrm{LF}}(s) \tag{21}$$

is the added nonlinear propagation delay due to the presence of a LF manipulation pressure. For conventional back-scatter imaging, the unmodified propagation lag is given by the total propagation length of $2r$ so that

$$t_0(r) = 2\int_0^r \frac{ds}{c_1(s)}. \tag{22}$$

For the added nonlinear propagation lag, however, the LF manipulation pressure will as already indicated drop so much in the back-scattered direction that nonlinear effects are negligible and one thus only get a nonlinear delay contribution as long as the HF imaging pulse follows the high amplitude forward propagating LF manipulation pulse giving

$$\tau_n(r) = -\int_0^r \frac{ds}{c_1(s)} \beta_{na}(s)\kappa_a(s)p_{\mathrm{LF}}(s) \tag{23}$$

or alternatively

$$\tau_n(t) = -\int_0^t dt_0 \beta_{na}(s(t_0))\kappa_a(s(t_0))p_{\mathrm{LF}}(s(t_0)). \tag{24}$$

With the SURF method, it is common to transmit more than one pulse complex in each beam line direction. The resulting slow-time (or pulse number) coordinate is denoted by the subscript k in the following equation. Let the received HF signals, y, as a function of fast-time (or depth-time) t be given as

$$y_k(t) = \tilde{x}_{lk}(t - \tau_{nk}(t)) + \tilde{x}_{nk}(t - \tau_{nk}(t)), \tag{25}$$

where the linearly back-scattered signal x_l results from the two first terms on the right-hand side in Eq. (15) and the nonlinearly back-scattered signal x_n results from the last term on the right-hand side in Eq. (15). Here, $\tilde{x}$ indicates that the received HF signal has experienced pulse distortion effects, other than conventional harmonic distortion, due to LF pressure gradients over the HF oscillation, and the nonlinear propagation delay τ_n is due to a constant LF pressure over the HF oscillation. The HF/LF timing shown in the upper panel of Fig. 1 will, due to relatively small LF pressure gradients observed by the HF oscillation, mainly introduce a nonlinear propagation delay on the HF pulse. With the HF/LF timing shown in the lower panel of Fig. 1, mainly a pulse distortion of the HF pulse will, due to a relatively small constant LF pressure observed by the HF oscillation, be introduced. These nonlinear modifications arise from the last term on the left-hand side in Eq. (15), i.e., a result of forward propagation effects, and will be discussed in more detail below. With conventional imaging, i.e., with no LF manipulation pressure, $\tilde{x}=x$ and $\tau_n=0$, and nonlinear propagation effects with single band transmit pulses are observed as harmonic distortion only.

The resulting nonlinear effects on the forward propagating HF pulses are hence conveniently described by decomposing the LF manipulation pressure experienced by the HF oscillation into a constant LF pressure, typically the LF pressure experienced by the central part of the HF oscillation, and a remaining LF pressure variation around this constant LF pressure. In Fig. 2, the indicated LF pressure decomposition is done by subtracting the manipulation pressure experienced at the central HF zero-crossing for the pulse complexes shown in Fig. 1. For the pulse complex shown in the

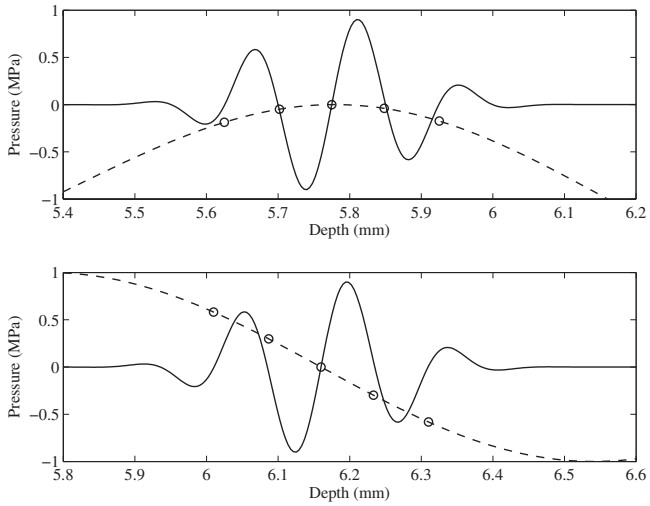


FIG. 2. LF manipulation pressure experienced by HF oscillation after LF manipulation pressure at central HF zero-crossing is subtracted. Upper and lower panels correspond to pulse complexes in upper and lower panels of Fig. 1, respectively.

upper panel of Fig. 1, 1 MPa is subtracted from the manipulation pressure and the remaining LF manipulation pressure experienced by the HF oscillation is displayed as the dashed line in the upper panel of Fig. 2. The position of the five central HF zero-crossings are indicated by the five circles on the dashed LF pressure line. The central zero-crossing hence experiences the highest LF pressure, i.e., 1 MPa. The two HF zero-crossings next to the central one experience a manipulation pressure which is about 50 kPa below the central one, i.e., 950 kPa, whereas the two outermost zero-crossings experience a manipulation pressure approximately 180 kPa below the central one, i.e., 820 kPa.

By doing the same LF manipulation pressure decomposition for the pulse complex in the lower panel of Fig. 1, one obtains the result shown in the lower panel of Fig. 2. The central HF zero-crossing in Fig. 1 coincides with a zero-crossing in the LF manipulation pressure so the constant LF pressure for this HF oscillation is zero. The dashed line again shows the manipulation pressure variation around this constant value and it is observed that for the five central HF zero-crossings, the manipulation pressure varies approximately from 600 kPa above to 600 kPa below the value obtained for the central HF zero-crossing.

In the LF pressure decomposition from the upper panel of Fig. 2, the constant 1 MPa LF pressure will make the whole HF oscillation propagate faster than a conventional transmit HF pulse without LF manipulation pressure, thus giving rise to a nonlinear propagation delay τ_n. The HF oscillation will then be time delayed by a given τ_n relative to a conventional imaging situation. From Eq. (23) or Eq. (24), it is seen that the nonlinear propagation delay is accumulative in nature and if, in the LF pressure decomposition indicated, the constant part of the LF pressure experienced by the HF pulse has the same sign for the entire propagation distance, $|\tau_n|$ will increase with depth.

Also, it is seen from Eq. (23) or Eq. (24) that τ_n depends on, and thus holds information about, the tissue parameter $\beta_{na}\kappa_a$ according to

$$\beta_{na}(r(t))\kappa_a(r(t)) = \frac{1}{p_{\mathrm{LF}}(r(t))}\frac{d\tau_n(t)}{dt}. \tag{26}$$

Estimation of nonlinear propagation delays is then important and cross-correlation techniques can, for example, be applied. A phase based approach for estimating nonlinear propagation delays with the SURF method has been described by Standal.[33] Estimation of the nonlinear compressibility parameter $\beta_{na}\kappa_a$ from τ_n hence facilitates tissue characterization.

From the LF pressure decomposition in the lower panel of Fig. 2, the constant LF pressure was zero and the HF oscillation will thus not obtain a nonlinear propagation delay compared to a conventional imaging situation without LF pressure, i.e., $\tau_n=0$. However, inspecting the LF pressure variation in the lower panel of Fig. 2, the tail of the HF oscillation experiences a higher LF pressure and, according to Eq. (18), a higher propagation velocity, than the front of the HF oscillation. The zero-crossings in the HF tail will then approach the zero-crossings in the front of the HF pulse increasing the HF center frequency as well as the HF bandwidth. This LF pressure variation over the HF oscillation will hence introduce a HF pulse distortion, indicated by $\tilde{x}$ in Eq. (25), which is different from conventional harmonic distortion. In addition to the indicated pulse distortion, the HF pulse will also experience conventional harmonic distortion.

Going back to the curves in the upper panel of Fig. 2, it is seen that the tail of the HF oscillation experiences a lower LF pressure and hence a lower propagation velocity than the central part of the HF oscillation. The distance between the zero-crossings in the HF tail and the central part of the HF oscillation will then be increased during propagation introducing lower HF components. The front of the HF oscillation also experiences a lower LF pressure, with a resulting lower propagation velocity than the central part of the HF oscillation. The central HF zero-crossing will approach zero-crossings in front of the HF oscillation during propagation introducing higher HF components. Hence, energy is during forward propagation pumped both up and down in frequency from the transmitted HF band with the HF/LF timing shown in the upper panel of Fig. 2, also giving a HF pulse distortion indicated by $\tilde{x}$ in Eq. (25). The LF manipulation pressure gradients responsible for this distortion are, however, relatively small compared to the LF manipulation pressure gradients seen in the lower panel of Fig. 2. Pulse distortion for the HF/LF timing shown in the upper panel of Fig. 2 will thus be less significant compared to what is obtained for the HF/LF timing shown in the lower panel of Fig. 2 as long as the difference in center frequency between the HF and LF pulses is as large as in the present example.

The pulse distortion discussed can be viewed as a variable delay, due to LF pressure gradients, over the HF oscillation. This variable delay is relative to the constant delay, τ_n, which for a given depth is constant for the whole HF oscillation. If, as in the present examples, the LF pressure experienced by the central HF zero-crossing is used as the constant LF pressure in the decomposition, then the variable delay will be zero for the central HF zero-crossing.

With SURF image reconstruction, one usually transmits

at least two pulse complexes in each beam line direction even if a few aspects of the SURF method can be performed utilizing only one transmit pulse complex in each beam line direction. Assume now that two pulse complexes are transmitted along the same beam line direction. In the first transmit pulse complex, the HF and LF pulses propagate substantially as shown in the upper panel of Fig. 1, whereas in the second transmit pulse complex, the LF pulse is removed so that the HF pulse propagates alone. Two HF echoes, y_1 and y_2, are then received, where y_2 is a conventional echo. From the above discussion, y_1 will be time delayed relative to y_2 and there will also be pulse distortion effects which, due to the indicated HF/LF timing, will be relatively small and will in the following example be neglected. The nonlinear part of the received signal is typically much weaker than the linear part of the received signal, i.e., $x_l \gg x_n$, and for several imaging situations, x_n may be neglected. By combining y_1 and y_2, one thus obtains the linearly back-scattered signal

$$s_l^1(t) = y_1(t) - y_2(t) = x_{l1}(t + \tau_{n1}(t)) - x_{l2}(t)$$
$$= x_l(t + \tau_n(t)) - x_l(t). \quad (27)$$

The nonlinear delay and pulse distortion due to the LF manipulation pressure are defined relative to a conventional situation with no LF manipulation pressure and for y_2, these nonlinear effects are thus zero by definition. The signal s_l^1 will be zero at $z=0$ and can then be made to increase gradually in magnitude as a function of depth due to the increase in nonlinear delay and pulse distortion for y_1. In many imaging situations, strong reflecting layers in the near-region between the ultrasound transducer and the region of interest will introduce additive noise due to multiple scattering. Hence, the signal s_l^1 is interesting with respect to suppression of such pulse reverberations.

For the signal y_1 in Eq. (27), assume that a 5 MHz HF pulse experiences a manipulation pressure of 0.4 MPa generated by a 0.5 MHz LF pulse within a depth range from 0 to 10 cm. From Eq. (24), the nonlinear propagation delay will with such a constant manipulation pressure increase linearly as a function of depth. At 10 cm, the nonlinear delay is 53 ns or approximately one-fourth of the HF period with numerical values of 5 and 400×10^{-12} Pa^{-1} for β_{na} and κ_a, respectively. The subtraction of two band-limited signals with a relative delay τ_n, as in Eq. (27), produces a depth variable gain factor of

$$G_\tau(t) \approx 2 \sin\left(\omega_h \frac{|\tau_n(t)|}{2}\right), \quad (28)$$

where ω_h is the angular center frequency of the HF imaging pulse and τ_n is the nonlinear propagation delay defined in Eq. (24). This gain factor expresses gain relative to a conventional fundamental field of an imaging pulse with no LF manipulation pulse present. Effects due to pulse compression and pulse expansion are not included in Eq. (28) but assuming that the HF pulse propagates on the crest or trough of the LF pulse, such nonlinear effects will be small.

Figure 3 displays the gain curve for s_l^1 in Eq. (27) obtained with the indicated numerical values as the solid line. The first order scattering, and hence also higher order scattering or reverberations, originating from reflectors very close to the transducer will be significantly suppressed. Around 6 cm, the signal s_l^1 has the same gain as a conventional fundamental signal and at 10 cm, the signal s_l^1 is approximately 3 dB stronger than the fundamental signal.

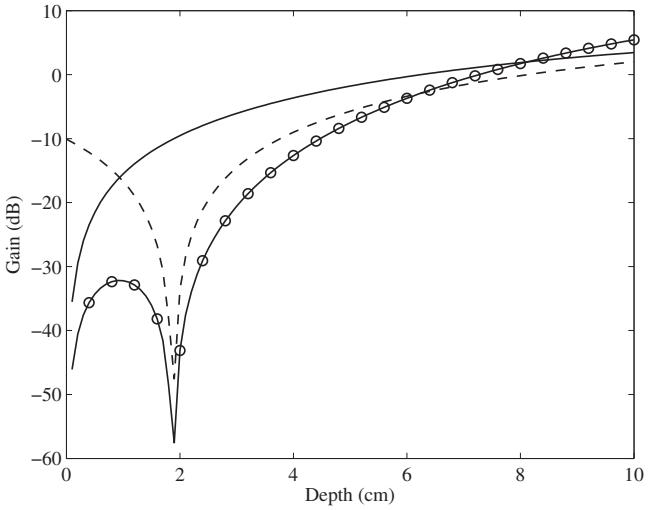


FIG. 3. Gain curves for nonlinear propagation delays with linear increase in delay as a function of depth. Solid line: The signal s_l^1 in Eq. (27). Dashed line: The signal s_l^2 in Eq. (29). Solid line with circles: The signal s_l^3 in Eq. (33).

With a strong reflecting layer around 2 cm, the signal suppression of 10 dB around this depth for s_l^1 is in some imaging situations potentially not enough for adequate reverberation suppression. Significant signal suppression around 2 cm may be obtained with the signal

$$s_l^2(t) = y_1(t - \hat{\tau}_n(t_2)) - y_2(t) = x_l(t + \tau_n(t) - \hat{\tau}_n(t_2)) - x_l(t), \quad (29)$$

where $\hat{\tau}_n(t_2)$ is the estimated nonlinear delay at 2 cm. The nonlinear propagation delay at the depth of the strong reflecting structure is hence estimated and corrected for before subtraction of the two signals. The gain curve for s_l^2 in Eq. (29) is depicted as the dashed line in Fig. 3. Strong suppression around 2 cm is now obtained but suppression very close to the transducer is less good because the two signals y_1 and y_2 are better aligned before delay correction with $\hat{\tau}_n(t_2)$ than after delay correction within this region.

To improve signal suppression within a larger region close to the transducer, a third pulse complex is transmitted along the same beam line direction as the two first transmit pulse complexes. This third transmit pulse complex differs from the first transmit pulse complex only in that the polarity of the LF manipulation pulse is inverted and the HF pulse will then experience a LF manipulation pressure of −0.4 MPa within the indicated depth range from 0 to 10 cm. For the third transmitted pulse complex, the nonlinear delay curve is then, according to Eq. (24), assumed to simply change sign relative to what obtained for the first transmitted pulse complex.

The linearly back-scattered signals resulting from the three indicated transmit pulse complexes may then be modeled as

$$y_1(t) = x_l(t + \tau_n(t)) + w_1(t + \tau_n(t_1)) + w_2(t + \tau_n(t_2)),$$

$$y_2(t) = x_l(t) + w_1(t) + w_2(t),$$

$$y_3(t) = x_l(t - \tau_n(t)) + w_1(t - \tau_n(t_1)) + w_2(t - \tau_n(t_2)), \quad (30)$$

where w_1 and w_2 are reverberation signals originating from strong reflectors located at t_1 and t_2, respectively. The two combination signals z_1 and z_2 will have a gain curve identical to the solid line in Fig. 3 and are hence useful for suppressing reverberation noise originating from structures very close to the transducer

$$\begin{aligned} z_1(t) &= y_1(t) - y_2(t) \\ &= x_l(t + \tau_n(t)) + w_1(t + \tau_n(t_1)) + w_2(t + \tau_n(t_2)) - x_l(t) \\ &\quad - w_1(t) - w_2(t) \\ &\approx x_l(t + \tau_n(t)) + w_2(t + \tau_n(t_2)) - x_l(t) - w_2(t), \end{aligned} \quad (31)$$

$$\begin{aligned} z_2(t) &= y_2(t) - y_3(t) \\ &= x_l(t) + w_1(t) + w_2(t) - x_l(t - \tau_n(t)) - w_1(t - \tau_n(t_1)) \\ &\quad - w_2(t - \tau_n(t_2)) \\ &\approx x_l(t) + w_2(t) - x_l(t - \tau_n(t)) - w_2(t - \tau_n(t_2)), \end{aligned} \quad (32)$$

where the last approximations are valid if t_1 is very close to the transducer so that $\omega_h \tau_n(t_1)$ is adequately small. To also get significant reverberation suppression for the strong reflector around t_2, the following signal is useful:

$$\begin{aligned} s_l^3(t) &= z_1\left(t - \hat{\tau}_n\left(\frac{t_2}{2}\right)\right) - z_2\left(t + \hat{\tau}_n\left(\frac{t_2}{2}\right)\right) \\ &= x_l\left(t + \tau_n(t) - \hat{\tau}_n\left(\frac{t_2}{2}\right)\right) - x_l\left(t - \hat{\tau}_n\left(\frac{t_2}{2}\right)\right) \\ &\quad + w_2\left(t + \tau_n(t_2) - \hat{\tau}_n\left(\frac{t_2}{2}\right)\right) - w_2\left(t - \hat{\tau}_n\left(\frac{t_2}{2}\right)\right) \\ &\quad - x_l\left(t + \hat{\tau}_n\left(\frac{t_2}{2}\right)\right) + x_l\left(t - \tau_n(t) + \hat{\tau}_n\left(\frac{t_2}{2}\right)\right) \\ &\quad - w_2\left(t + \hat{\tau}_n\left(\frac{t_2}{2}\right)\right) + w_2\left(t - \tau_n(t_2) + \hat{\tau}_n\left(\frac{t_2}{2}\right)\right) \\ &\approx x_l\left(t + \tau_n(t) - \hat{\tau}_n\left(\frac{t_2}{2}\right)\right) - x_l\left(t - \hat{\tau}_n\left(\frac{t_2}{2}\right)\right) \\ &\quad + x_l\left(t - \tau_n(t) + \hat{\tau}_n\left(\frac{t_2}{2}\right)\right) - x_l\left(t + \hat{\tau}_n\left(\frac{t_2}{2}\right)\right). \end{aligned} \quad (33)$$

Notice that to significantly suppress reverberations originating from t_2 in s_l^3, the nonlinear propagation delay at $t_2/2$ must be estimated and applied. The gain curve for s_l^3 is shown as the solid line with circles in Fig. 3. It is seen that the gain is below 20 dB for depths up to about 3 cm and at 10 cm the gain is approximately 6 dB above a conventional fundamental signal.

B. Nonlinear back-scattering

In other imaging situations, e.g., for imaging of ultrasound contrast bubbles, only the nonlinear part of the back-scattered signal is interesting. Scattering from contrast bubbles is not adequately described by the right-hand side terms of Eq. (15) and a separate nonlinear differential equation taking care of the strong nonlinear and resonant scattering from such microbubbles can be derived.[11] With the SURF method, the size and hence scattering properties of a contrast bubble can be manipulated by the LF pulse, and the HF imaging pulse can then be made to interrogate a bubble with varying scattering properties depending on the HF/LF timing in the transmit pulse complexes. However, to adequately suppress the strong linear tissue component of the received signal, one must estimate and correct nonlinear effects in the forward wave propagation, i.e., nonlinear propagation delays and possibly also pulse form distortions. For the nonlinear delays, one can then, for example, divide the received time/depth interval T into shorter time intervals T_i that possibly overlap, so that $T < \Sigma_i T_i$, and estimate nonlinear delay corrections for each subinterval T_i for adequate local suppression of the linear part of the received signal. One way of estimating the nonlinear delay corrections, $\hat{\tau}_{nk}$, for two consecutively received HF echoes, is by maximizing the power in the signal

$$\sigma(t) = \sum_{k=1}^{2} y_k(t + \hat{\tau}_{nk}(t)). \quad (34)$$

Cross-correlation techniques can, as already indicated, be applied and a phase based approach for estimating nonlinear propagation delays with the SURF method has been described by Standal.[33]

Local nonlinear back-scattering from tissue without contrast agents is, from Eq. (15), seen to vary with the LF manipulation pressure according to

$$x_n \sim 2\sigma_n \beta_{na} \kappa_a p_{\mathrm{LF}}. \quad (35)$$

The LF manipulation pressure does, by definition, not affect the linear back-scattering directly. From Eq. (15), linear back-scattering depends on σ_l and γ. Experimentally, one typically finds that $\gamma \approx -0.3\sigma_l$ for soft tissues, and linear back-scattering hence varies according to[31] (Chap. 7)

$$x_l \sim \sigma_l - \gamma \approx 1.3\sigma_l. \quad (36)$$

For soft tissues, typical numerical values for variations in the nonlinear parameter and bulk compressibility are[10,31] $\alpha_n \approx \sigma_l \approx \pm 0.1$ so that in Eq. (14), $\sigma_n \approx 3\sigma_l$ and from Eq. (35), the nonlinear back-scattering is

$$x_n \sim 6\sigma_l \beta_{na} \kappa_a p_{\mathrm{LF}}. \quad (37)$$

The ratio of the nonlinearly to the linearly back-scattered signal is hence

$$\frac{x_n}{x_l} \approx \frac{6\sigma_l \beta_{na} \kappa_a p_{\mathrm{LF}}}{1.3\sigma_l}. \quad (38)$$

With a LF manipulation pressure of 1 MPa and with numerical values of 5 and 400×10^{-12} Pa^{-1} for β_{na} and κ_a, respectively, $x_n/x_l\approx-40$ dB.

With conventional single band transmit pulses, the last term on the left-hand side of Eq. (15) produce distortion of the forward propagating pulses, resulting in incident pulses with harmonic components. These incident pulses will then be linearly back-scattered from tissue inhomogeneities, according to the two first terms on the right-hand side of Eq. (15), so that echoes received by linear back-scattering will contain harmonic components. Nonlinear back-scattering, according to the last term on the right-hand side of Eq. (15), will also introduce harmonic components in the received signal similar to the nonlinear propagation distortion. For soft tissues, local nonlinear back-scattering is as indicated typically low compared to linear back-scattering, and the nonlinearly back-scattered signal is not differentiable from the propagation distortion. The last term on the right-hand side of Eq. (15) is hence typically not seen with conventional harmonic imaging methods.

The SURF method shows a potential for separating the nonlinear propagation effects from the local nonlinear back-scattering. As already described, nonlinear propagation effects due to the LF manipulation pressure will show up as nonlinear HF propagation delays and as HF pulse distortion depending on the HF/LF timing. The resulting pulse distortion will, as discussed above, not be a simple harmonic distortion as seen with conventional single band transmit pulses but will be observed as pulse compression and pulse expansion effects.

Assume now that two pulse complexes are transmitted along the same beam line direction. In the first pulse complex, the HF and LF pulses propagate substantially, as shown in the upper panel of Fig. 1, where the HF pulse experiences a high LF manipulation pressure. In the second transmitted pulse complex, the polarity of the LF pulse is inverted while the HF pulse is unchanged. The HF pulse will then experience a low pressure generated by the LF manipulation pulse. Due to relatively small LF pressure gradients over the HF oscillation with the indicated HF/LF timing, pulse distortion effects from the LF pulse on the HF pulse are neglected in the present example. Two HF echoes, y_1 and y_2, are then received where the main nonlinear propagation effect on the HF pulse by the LF pulse is a nonlinear propagation delay. Since the LF pulse is inverted for the two transmit pulse complexes, based on Eq. (23), it is assumed that $\tau_{n1}=-\tau_{n2}$ and from Eq. (37), $x_{n1}=-x_{n2}$. To obtain a local nonlinear back-scattering signal, the strong linear echoes must first be suppressed. In the present example, this is done by estimating and correcting for the resulting nonlinear propagation delays giving the nonlinear signal

$$\begin{aligned} s_n(t) &= y_1(t+\hat{\tau}_n(t)) - y_2(t-\hat{\tau}_n(t)) \\ &= x_l(t-\tau_n(t)+\hat{\tau}_n(t)) + x_n(t-\tau_n(t)+\hat{\tau}_n(t)) \\ &\quad - x_l(t+\tau_n(t)-\hat{\tau}_n(t)) + x_n(t+\tau_n(t)-\hat{\tau}_n(t)) \\ &\approx 2x_n(t). \end{aligned} \tag{39}$$

The last approximation in the above equation is valid if the estimation of nonlinear propagation delay, $\hat{\tau}_n$, is adequately good.

IV. CONCLUSION

The present paper has described a method for acoustic imaging utilizing transmit pulse complexes consisting of overlapping HF and LF pulses. The LF pulse is used to manipulate the nonlinear material elasticity observed by the co-propagating HF pulse, which is used for image reconstruction. This method aims to further utilize nonlinear and resonant acoustical effects for improved image reconstructions of objects.

ACKNOWLEDGMENT

This work was supported by the Research Council of Norway.

[1]T. Muir and E. Carstensen, "Prediction of nonlinear acoustic effects at biomedical frequencies and intensities," Ultrasound Med. Biol. **6**, 345–357 (1980).
[2]E. Carstensen, W. Law, N. McKay, and T. Muir, "Demonstration of nonlinear acoustical effects at biomedical frequencies and intensities," Ultrasound Med. Biol. **6**, 359–368 (1980).
[3]K. Caidal, E. Kazzam, J. Lidberg, G. Neumann Andersen, J. Nordanstig, S. Dahlqvist, A. Waldenstrom, and R. Wikh, "New concepts in echocardiography: Harmonic imaging of tissue without use of contrast agents," Lancet **352**, 1264–1270 (1998).
[4]K. Spencer, J. Bednarz, P. Rafter, C. Korcarz, and R. Lang, "Use of harmonic imaging without echocardoigraphic contrast to improve two-dimensional image quality," Am. J. Cardiol. **82**, 794–799 (1998).
[5]M. Averkiou, D. Roundhill, and J. Powers, "A new imaging technique based on the nonlinear properties of tissue," in Ultrasonics Symposium Proceedings (1997), pp. 1561–1566.
[6]M. Averkiou, "Tissue harmonic imaging," in Ultrasonics Symposium Proceedings (2000), pp. 1563–1572.
[7]J. Thomas and D. Rubin, "Tissue harmonic imaging: Why does it work?," J. Am. Soc. Echocardiogr **11**, 803–808 (1998).
[8]F. Duck, "Nonlinear acoustics in diagnostic ultrasound," Ultrasound Med. Biol. **28**, 1–18 (2002).
[9]S. Choudhry, B. Gorman, J. Charboneau, D. Tradup, R. Beck, J. Kofler, and D. Groth, "Comparison of tissue harmonic imaging with conventional us in abdominal disease," Radiographics **20**, 1127–1135 (2000).
[10]F. Duck, *Physical Properties of Tissue* (Academic, London, 1990).
[11]T. G. Leighton, *The Acoustic Bubble* (Academic, San Diego, 1994).
[12]D. Miller, "Ultrasonic detection of resonant cavitation bubbles in a flow tube by their second harmonic emissions," Ultrasonics **22**, 217–224 (1981).
[13]N. de Jong and R. Cornet, "Higher harmonics of vibrating gas-filled microspheres. Part one: Simulations," Ultrasonics **32**, 447–453 (1994).
[14]N. de Jong and R. Cornet, "Higher harmonics of vibrating gas-filled microspheres. Part two: Measurements," Ultrasonics **32**, 455–459 (1994).
[15]J. Hwang and D. Simpson, "Two pulse technique for ultrasonic harmonic imaging," U.S. Patent No. 5,951,478 (1999).
[16]D. Simpson, C. Chin, and P. Burns, "Pulse inversion Doppler: A new method for detecting nonlinear echoes from microbubble contrast agents," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **46**, 372–382 (1999).
[17]P. Burns, S. Wilson, and D. Simpson, "Pulse inversion imaging of liver blood flow: Improved method for characterizing focal masses with microbubble contrast," Invest. Radiol. **35**, 58–71 (2000).
[18]V. Mor-Avi, E. Caiani, K. Collins, C. Korcarz, J. Bednarz, and R. Lang, "Combined assessment of myocardial perfusion and regional left ventricular function by analysis of contrast-enhanced power modulation images," Circulation **104**, 352–357 (2001).
[19]B. Haider and R. Chiao, "Higher order nonlinear ultrasonic imaging," in Ultrasonics Symposium Proceedings (1999), pp.1527–1531.
[20]P. Phillips, "Contrast pulse sequence (CPS): Imaging nonlinear microbubbles," in Ultrasonics Symposium Proceedings (2001), pp.1739–1745.
[21]R. Eckersley, C. Chin, and P. Burns, "Optimising phase and amplitude

modulation schemes for imaging microbubble contrast agents at low acoustic power," Ultrasound Med. Biol. **31**, 213–219 (2005).
[22]B. Angelsen and R. Hansen, "SURF imaging—A new method for ultrasound contrast agent imaging," in Ultrasonics Symposium Proceedings (2007), pp. 531–541.
[23]R. Hansen and B. Angelsen, "SURF imaging for contrast agent detection," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **56**, 280–290 (2009).
[24]T. Sato, A. Fukusima, N. Ichida, H. Ishikawa, H. Miwa, Y. Igarashi, T. Shimura, and K. Murakami, "Nonlinear parameter tomography system using counterpropagating probe and pump waves," Ultrason. Imaging **7**, 49–59 (1985).
[25]T. Sato, Y. Yamakoshi, and T. Nakamura, "Nonlinear tissue imaging," in Ultrasonics Symposium Proceedings (1986), pp. 889–900.
[26]C. Cain, "Ultrasonic reflection mode imaging of the nonlinear parameter B/A: I. A theoretical basis," J. Acoust. Soc. Am. **80**, 28–32 (1986).
[27]H. Fukukita, S.-i. Ueno, and T. Yano, "Ultrasound pulse reflection mode measurements of nonlinearity parameter B/A and attenuation coefficient," J. Acoust. Soc. Am. **99**, 2775—2782 (1996).
[28]C. Deng, F. Lizzi, A. Kalisz, A. Rosado, R. Silverman, and D. Coleman, "Study of ultrasonic contrast agents using a dual-frequency band technique," Ultrasound Med. Biol. **26**, 819–831 (2000).
[29]A. Bouakaz, M. Versluis, J. Borsboom, and N. de Jong, "Radial modulation of microbubbles for ultrasound contrast imaging," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **54**, 2283–2290 (2007).
[30]S.-E. Måsøy, Ø. Standal, P. Nasholm, T. Johansen, B. Angelsen, and R. Hansen, "SURF imaging: In vivo demonstration of an ultrasound contrast agent detection technique," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **55**, 1112–1121 (2008).
[31]B. Angelsen, *Ultrasound Imaging* (Emantec, Trondheim, Norway, 2000), Vol. **II**, www.ultrasoundbook.com (Last viewed 3/28/2009).
[32]M. Hamilton and D. Blackstock, *Nonlinear Acoustics* (Academic, San Diego, 1998).
[33]Ø. Standal, T. Tangen, and B. Angelsen, "A phase based approach for estimation and tracking of locally variable delays," in Ultrasonics Symposium Proceedings (2007), pp. 1583–1585.

ACOUSTICAL NEWS

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

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

New Fellows of the Acoustical Society of America

Ruth Y. Litovsky—For contributions to spatial hearing

Patrick J. Loughlin—For contributions to time-frequency signal processing

Oleg A Sapozhnikov—For contributions to nonlinear ultrasonics and medical acoustics

Keith A. Wear—For contributions to ultrasonic bone-quality characterization

Reports from Frederick V. Hunt Postdoctoral Research Fellows

Report of the 29th F. V. Hunt Post-Doctoral Fellow, 2006–2007

Erica E. (Bowden) Ryherd, Woodruff School of Mechanical Engineering, Georgia Institute of Technology, Atlanta, Georgia 30332-0405 (erica.ryherd@me.gatech.edu)

I was honored to receive the F.V. Hunt Postdoctoral Fellowship to pursue acoustics research at the Sahlgrenska Academy of Medicine of Gothenburg University in Sweden. Dr. Kerstin Persson Waye, a professor of Occupational and Environmental Medicine, was my advisor. The topic of our research was "Establishing a Hospital Soundscape through Qualitative and Quantitative Observation." The hospital soundscape was of particular interest to us because of the variety of complex noise sources, the unique sensitivity of occupants, and the limitations in acoustical treatments. We also saw this topic as an exciting opportunity to combine my background in building systems engineering with Dr. Persson Waye's medical expertise. Our first phase focused on the acoustics and staff response in a Swedish neurological intensive care unit (Neuro-ICU). We conducted noise measurements and administered a questionnaire to nurses. We found that the nurses perceived noise in the Neuro-ICU as contributing to stress symptoms such as irritation, fatigue, concentration problems, and tension headaches. We measured average A-weighted equivalent level (L_{Aeq}) of 53–58 dBA near patients. Dosimeters (body-mounted microphones) worn by nurses revealed even higher values. This led to separate spin-off collaboration with the Chalmers University Applied Acoustics Program to quantify the effect of the wearer's voice on noise dosimeter measurements. As part of the Neuro-ICU

 0001-4966/2010/127(1)/588/6/$25.00

study, we also investigated more detailed methods to quantify the background noise including statistical level distributions, restorative periods, and indoor noise criteria evaluations of spectral content. These detailed metrics were found to give a much more complete description of the sound environment than the overall average, minimum, maximum, and peak levels. For example, we found that the C-weighted peak sound pressure level (L_{CPeak}) exceeded 70 dBC more than 90% of the time and 100 dBC less than 5% of the time.

I am continuing to collaborate with Dr. Persson Waye. In a phase currently underway, we are focusing on the acoustics, staff, and patient response in a Swedish medical-surgical intensive care unit (Med-Surg ICU). An expanded version of the questionnaire used in the previous phase has been administered to nurses. We have found that excess sound stimulation is a highly significant explanatory factor for noise annoyance even after correcting for age, education, experience, and psycho-social factors including salary and leadership. Patients in this phase have also been monitored for 24 hours during their stay in the Med-Surg ICU. During that time, continuous noise data and physiological data have been collected. Additionally, evaluations of ICU Syndrome have been conducted for each patient, using a log developed by Swedish collaborators in nursing science. ICU Syndrome is a condition in which ICU patients may exhibit a variety of negative psychological reactions such as distress, hallucinations, and fear. Data from the patient portion of the project are currently under analysis.

My tenure as the F.V. Hunt Postdoctoral Fellow was extremely rewarding in both intellectual and personal capacities. One of the most fruitful parts of my postdoc was identifying many areas of hospital acoustics that still need work. In addition to continued collaborations with Dr. Persson Waye, I have begun formal collaborations with U.S. colleagues in engineering, medicine, architecture, and psychology. My long-term career goal is to revolutionize the acoustic climate in hospitals through research that enables transformative changes to current hospital acoustic policies and ordinances, measurement techniques, design practices, and noise control technologies. I want to thank the ASA and the Hunt family for such an incredible opportunity. I also want to express my deepest and most sincere gratitude to my wonderful collaborators in this area of research, particularly Dr. Kerstin Persson Waye, Dr. Jim West, and Dr. Ilene Busch-Vishniac for their mentorship during my Hunt Fellowship.

The following publications and presentations resulted or followed from my Hunt Fellowship:

Publications & Proceedings

- H. Pelton, E. Ryherd, and M. Martin, "Acoustical design of a burn acute care unit for enhanced patient comfort," Noise Control Eng. J. 57(1), 32–41 (2009).
- E. Ryherd, K. Persson Waye, and Linda Ljungkvist, "Characterizing noise and perceived work environment in a neurological intensive care unit," J. Acoust. Soc. Am. 123(2), 747–756 (2008).
- E. Ryherd, J. West, I. Busch-Vishniac, and K. Persson Waye, "Evaluating the hospital soundscape," Acoustics Today 4(4), 22–29 (2008).
- E. Ryherd and K. Persson Waye, "Soundscape evaluations in adult intensive care units," Noise-Con Conf. Proc., Reno, NV (2007).

Conference Presentations

- T. Hsu, E. Ryherd, and K. Persson Waye, "Evaluating the intensive care unit soundscape," 157th meeting of the Acoust. Soc. Am., Portland, OR (2009).
- S. Okcu, E. Ryherd, and C. Zimring, "Comparing the sound environments in two critical care settings," 157th meeting of the Acoust. Soc. Am., Portland, OR (2009).
- E. Ryherd, "Characterizing the hospital acoustic environment and occupant response," 2009 Amer. Soc. Heating, Refrig. & Air-Cond. Eng. (ASHRAE) Winter Meeting, Chicago, IL (2009).
- E. Ryherd, "Psychological and physiological effects of sound," Joint Conf. of Natl. Soc. Black Phys. (NSBP) & Natl. Soc. Hispanic Phys. (NSHP), Nashville, TN (2009).
- S. Okcu, C. Zimring, and E. Ryherd, "Developments in aural connectivity: Enhancing sound environments in critical care settings for effective nurse auditory monitoring," 155th meeting of the Acoust. Soc. Am., Paris, France (2008).
- E. Ryherd and K. Persson Waye, "Describing the sound environment of the neurological ICU and personnel response," 155th meeting of the Acoust. Soc. Am., Paris, France (2008).
- K. Persson Waye, E. Ryherd, B. Lindahl, and I. Bergbom, "Relating the hospital sound environment to occupant psychological and physiological response," 155th meeting of the Acoust. Soc. Am., Paris, France (2008).
- S. Ryherd, K. Persson Waye, M. Kleiner, and E. Ryherd, "Quantifying the noise environment: Effects of the wearer's voice on body-mounted noise dosimeter measurements," 155th meeting of the Acoust. Soc. Am., Paris, France (2008).

Seminars

- E. Ryherd, "Impact of the indoor soundscape on occupants," Vanderbilt Univ. Dept. of Hearing and Speech Sciences, Nashville, TN (2009).
- E. Ryherd, "Research in healthcare acoustics: Past—present—future," Ecophon HealthNet Mtg., Helsingborg, Sweden (2006).
- E. Ryherd, "Quantifying indoor noise based on human performance and perception," Chalmers Univ. of Tech. Applied Acoustics Program, Gothenburg, Sweden (2006).
- E. Ryherd, "Research topics in architectural acoustics and noise control," Gothenburg Univ. Dept. of Occupational and Environmental Medicine Seminar Series, Gothenburg, Sweden (2006).

Report of the 30th F. V. Hunt Postdoctoral Fellow, 2007–2008

Julie N Oswald, Univ of California San Diego, Scripps Inst of Oceanography, 9500 Gilman Dr., La Jolla, CA 92093-0225

During my tenure as the 2007–2008 F.V. Hunt Postdoctoral Fellow, I had the privilege of working under the mentorship of Dr. Whitlow W.L. Au of the Marine Mammal Research Program (MMRP) at the University of Hawaii. The primary objective of my research was to develop methods to acoustically monitor the presence of marine mammals in Hawaiian waters by 1) increasing the accuracy of acoustic species identification software (ROCCA), and 2) using data collected at the Station ALOHA cabled observatory (ACO) to study the occurrence of minke whales (*Balaenoptera acutorostrata*).

Monitoring marine mammal occurrence, distribution, and abundance has traditionally depended on visual observations made from a shipboard or aerial platform. These methods are limited to daytime hours and must be suspended when poor weather conditions prohibit reasonable visual detection of animals. Because many marine mammals produce sounds that propagate well under water, acoustic techniques can be used to monitor their presence. For example, acoustic recorders can be mounted on the sea floor and left to record for extended periods, providing long time-series data that would not be possible to obtain using shipboard visual surveys alone. However, in order for acoustic techniques to be useful, it is necessary to have the ability to identify species based on the sounds that they make. My dissertation work culminated in the creation of software (Real-time Odontocete Call Classification Algorithm, ROCCA) for the acoustic identification of dolphin species in the eastern tropical Pacific Ocean (ETP) (Oswald 2006, Oswald *et al.*, 2007). During my year at the University of Hawaii, I worked to increase the accuracy of this software by exploring geographic variation in dolphin whistles within species. I found significant variation both within and between ocean basins (ETP vs. western Mediterranean Sea) for short-beaked common dolphins (*Delphinus delphis*) and striped dolphins (*Stenella coeruleoalba*) and significant variation between subspecies of spotted dolphins (*Stenella attenuata*) in the ETP. Taking this variation into account and creating different species identification algorithms for different regions within the ETP leads to higher correct classification scores when identifying whistles to species. This work was presented at the 2nd International Conference on Acoustic Communication by Animals (Oswald *et al.*, 2008) and the 4th International Workshop on Detection, Classification, and Localization of Marine Mammals Using Passive Acoustics (Oswald *et al.*, 2009). A manuscript is in preparation for submission to the *Journal of the Acoustical Society of America* (Oswald *et al.*, *in prep*).

Some species, such as fin whales (*Balaenoptera physalus*), sperm whales (*Physeter macrocephalus*), and minke whales produce relatively stereotyped sounds that can be easily identified to species. In these cases, an automated detector can also act as a classifier. To fulfill the second objective of my research, I used an automated spectrogram correlation detector

(XBAT) to locate minke whale boing sounds in recordings made at the ACO. The ACO is a cabled ocean bottom observatory located at a depth of 4.7 km, 100 km north of Oahu, HI. It is funded by the National Science Foundation and maintained by Fred Duennebier, Roger Lukas and Bruce Howe of the School of Ocean and Earth Science and Technology at the University of Hawaii. It has been the site of near monthly shipboard oceanographic observations for over 18 years as part of the Hawaii Ocean Time Series Program, making it a perfect site for an examination of the seasonal and diurnal occurrence of marine mammals. Although the ACO hydrophone was originally installed to listen for seismic events, it was able to detect many marine mammal species, including fin, minke, sperm and humpback (*Megaptera novaeangliae*) whales among others. The ACO hydrophone was operational and recording continuously from February 2007 to October 2008. Over 15,000 boings were detected in these recordings during the months of October to June, suggesting that minke whales are seasonally abundant in Hawaiian waters. These results contrast with the traditionally held notion, based on visual observations, that minke whales are rare in these waters. In addition to the seasonal occurrence of boings at the ACO, my research showed that there is no diurnal pattern in the production of boings. We are currently exploring factors related to the seasonal presence of minke whale boings at the ACO by comparing the presence of boings to oceanographic data (temperature, chlorophyll a, zooplankton, etc.) collected during Hawaii Ocean Time Series Program cruises. Some of the results of this research were presented at the 156th meeting of the Acoustical Society of America (Oswald *et al.*, 2008) and a manuscript is currently being prepared for submission to the *Journal of the Acoustical Society of America* (Oswald *et al.*, *in prep*). Plans are also in place to create automated detectors for other species such as fin and blue (*Balaenoptera musculus*) whales in order to examine their occurrence in the ACO area as well as their co-occurrence with other species.

My time as an F.V. Hunt Fellow at the University of Hawaii (UH) was incredibly enriching and rewarding. I learned a great deal and had the opportunity to work on several interesting projects with many exceptional people. I would like to thank the Acoustical Society of America and the Hunt family for an experience that has had a significant positive impact on my education and my career. I would like to thank Dr. Au for his unending patience, guidance and ideas, and Dr. Marc Lammers for his advice and support. During my time at UH, I greatly benefited from interactions with Hawaii Institute of Marine Biology graduate students, including: Laura Kloepper, Aran Mooney, Aude Pacini, Michael Richlen, Alexis Rudd, and Alison Stimpert. I am grateful also to Dr. Fred Duennebier, Dr. Paul Nachtigall, Dr. Joe Mobley, Dr. Sofie Van Parijs, Linda Choy, Tom Norris, and Charlotte Fabjanczyk for their contributions to my time in Hawaii.

Literature Cited

Oswald, J.N., Au, W.W.L., Duennebier, F., and Fabjanczyk, C. (**In prep**). "Things that go 'boing' in the night: using the Station Aloha Cabled Observatory to study the occurrence of minke whales in Hawaiian waters," to be submitted to the Journal of the Acoustical Society of America.

Oswald, J.N., Rankin, S., Gannier, A., Barlow, J., Fuchs, S., Rudd, A., and Au, W.W.L. (**In prep**). "Geographic variation in the whistles produced by delphinid species and implications for acoustic species identification," to be submitted to Journal of the Acoustical Society of America.

Oswald, J.N., Rankin, S., Gannier, A., Barlow, J., Fuchs, S., Rudd, A., and Au, W.W.L. (**2009**). "Geographic variation in the whistles produced by four delphinid species in the Pacific Ocean and Mediterranean Sea," 4th International Workshop on Detection, Classification, and Localization of Marine Mammals Using Passive Acoustics, September 10-12, 2009, Pavia, Italy.

Oswald, J.N., Gannier, A., Fuchs, S., Rankin, S., and Barlow, J. (**2008**). "Differences in whistle characteristics of two delphinid species in the eastern tropical Pacific Ocean and the western Mediterranean Sea," 2nd International Conference on Acoustic Communication by Animals, August 12-15, 2008, Oregon State University, Corvallis, Oregon.

Oswald, J.N., Norris, T., Au, W.W.L., and Duennebier, F. (**2008**). "Temporal and geographic patterns in the occurrence and distribution of minke whale (*Balaenoptera acutorostrata*) boings in the central and western north Pacific," Journal of the Acoustical Society of America **124**:2484.

Oswald, J.N. (**2006**). "An examination of the whistling behavior of small odontocetes and the development of methods for species identification of delphinid whistles," PhD dissertation, University of California, San Diego, Scripps Institution of Oceanography.

Oswald, J.N., Rankin, S., Barlow, J., and Lammers, M.O. (**2007**). "ROCCA: a new tool For real-time acoustic species identification of delphinid whistles," Journal of the Acoustical Society of America 122:587-595.

Report of the 31st F. V. Hunt Postdoctoral Fellow, 2008–2009

Todd A. Hay, Applied Research Labs., Univ. of Texas at Austin, Austin, TX 78713-8029

As the 2008-2009 F.V. Hunt Postdoctoral Fellow I had the privilege of working under the mentorship of Dr. Michel Versluis and Prof. Detlef Lohse in the Physics of Fluids (PoF) group at the University of Twente in the Netherlands. My work was concerned with dynamics of micron-sized ultrasound contrast agent (UCA) bubbles in confining environments, a topic currently of interest in the biomedical acoustics community. My graduate research activities focused on the development of theoretical models for bubble-particle interaction and bubble behavior in compliant channels, and the facilities and expertise available in the PoF group provided a unique opportunity to test these models through two sets of experiments.

The first set of experiments involved the interaction of acoustically driven bubbles and solid particles, an investigation motivated by a desire to understand the interaction of bubble clusters with kidney stone fragments during shock-wave lithotripsy. In my dissertation I developed a mathematical model of the dynamics of clustered bubbles and solid particles, and so I was excited to test this model in the lab. In particular, I wanted to observe the translational force predicted to act between a pulsating bubble and rigid particle. Using optical tweezers, we positioned several UCA bubbles close to larger glass spheres while exciting the system acoustically. We were then able to measure the radial and translational motion of the bubble with high speed imaging. Our model simulations agree with our measurements if the drag force acting on the bubble is properly included. These results will be presented at a future meeting of the Acoustical Society of America.

The second set of experiments focused on UCA behavior in compliant channels. While the ultimate goal is to understand UCA interaction with biological tissue, we limited ourselves to situations which we could model analytically. The geometries we selected were that of a bubble oscillating near a single plate, between parallel plates, or within microchannels of rectangular cross section. In addition, we constructed the plates and channels from several plastics and polymers which have well defined and measurable acoustic properties. Over the course of several weeks, we measured the radial motion of several UCA bubbles over a range of insonation frequencies and at various offset distances from the plates or channel walls. I expect that analyzing the data and modeling these experiments will keep me busy for quite some time.

In addition to collecting data I also began to develop theoretical models to explain some of the trends we observed. For example, we found that the oscillation amplitude and natural frequency of the bubbles depend on their position relative to the surfaces, as well as the material properties and geometry of the confinement. In the single-plate geometry we found that both the resonance frequency and maximum oscillation amplitude of the bubble tend to decrease as the UCA is moved closer to a plastic plate. These results are unexpected if the plate is assumed to be rigid. However, through our modeling efforts we found that if we allow the plate to have finite stiffness (so that it supports elastic waves), we can match the measured behavior with our model. I presented our theory [Hay, T.A., *et al.*, "Model for bubble dynamics in liquid near an elastic tissue interface"] along with our measurements [Hay, T.A., *et al.*, "Influence of finite wall impedance on contrast agent bubble behavior near a membrane"] at the Spring 2009 Meeting of the ASA in Portland. Two related papers are in preparation for publication in JASA. In the meantime, I have extended our single viscoelastic plate model to the parallel plate geometry. This was presented along with our measurements at the Fall 2009 Meeting in San Antonio.

My tenure as a Hunt Fellow was rewarding, both scientifically and personally, and I am most grateful to the Hunt family and the Acoustical Society of America for this opportunity to begin what I hope will become a long-term collaboration with the Physics of Fluids group. In addition to the excellent guidance of Dr. Versluis and Prof. Lohse I am extremely grateful to Dr. Valeria Garbin, who was instrumental in designing and executing each experiment. I am also indebted to the graduate students and support staff,

including Marlies Overvelde, Jeroen Sijl, Erik Gelderblom, Joanita Leferink, and Gert-Wim Bruggert, who helped make my year in the Netherlands one of the most productive and enjoyable times of my life.

Calendar of Meetings and Congresses

Compiled by the Information Service of the International Commission for Acoustics

2010

6–9 January	Sanya, China. 2nd International Conference on Vibro-Impact Systems. http://www.neu.edu.cn
8–10 March	Chofu, Japan. Spring Meeting of the Acoustical Society of Japan.
15–18 March	Berlin, Germany. Meeting of the German Association for Acoustics DAGA 2010. http://www.daga-tagung.de/2010
15–19 March	Dallas, TX, USA. International Conference on Acoustics, Speech, and Signal Processing. http://icassp2010.org
7–9 April	Cambridge, UK. David Weston Sonar Performance Assessment Symposium
19–23 April	Baltimore, MD, USA. Joint meeting: 159th Meeting of the Acoustical Society of America and Noise-Con 2010. http://asa.aip.org/meetings.html
27–30 April	Ghent, Belgium. Institute of Acoustics/Belgian Acoustical Association Joint Meeting. http://www.ioa.org.uk/viewupcoming.asp
6–7 May	Paris, France. 2nd International Symposium on Ambisonics and Spherical Acoustics. http://ambisonics10.ircam.fr
10–12 May	Bergen, Norway. Baltic-Nordic Acoustics Meeting2010. E-mail: sanordby@norsonic.com
9–11 June	Aalborg, Denmark. 14th Conference on Low Frequency Noise and Vibration. http://lowfrequency2010.org
13–16 June	Lisbon, Portugal. INTERNOISE2010. http://www.internoise2010.org
5–9 July	Istanbul, Turkey. 10th European Conference on Underwater Acoustics. http://ecua-2010-istanbul.org
18–22 July	Cairo, Egypt. 17th International Congress on Sound and Vibration (ICSV17). http://www.icsv17.org
23–27 August	Sydney, Australia. International Congress on Acoustics 2010. http://www.ica2010sydney.org
23–27 August	Seattle, USA. 11th International Conference on Music Perception and Cognition.
29–31 August	Melbourne, Australia. International Symposium on Room Acoustics (ISRA2010). http://www.isra2010.org
14–18 September	Kyoto, Japan. 5th Animal Sonar Symposium. http://cse.fra.affrc.go.jp/akamatsu/AnimalSonar.html
15–18 September	Ljubljana, Slovenia. Alps-Adria-Acoustics Association Meeting joint with EAA. E-mail: mirko.cudina@fs.uni-lj.si
26–30 September	Makuhari, Japan. Interspeech 2010—ICSLP. http://www.interspeech2010.org
11–14 October	San Diego, Cal. USA. IEEE 2010 Ultrasonics Symposium. E-mail: b.potter@vectron.com
14–16 October	Niagara-on-the Lake, Ont., Canada. Acoustics Week in Canada. http://caa-aca.ca/E/index.html
11–14 October	San Diego, California, USA. IEEE 2010 Ultrasonics Symposium. E-mail: bpotter@vectron.com
13–15 October	Leon, Spain. 41st Spanish Congress of Acoustics and 6th IberianAcoustics Congress. http:// www.sea-acustica.es
18–22 October	Nagahama, Japan. 10th International Workshop on Railway Noise (IWRN10). http://www.rtri.or.jp/IWRN10/first.announcement.html
15–19 November	Cancún, Mexico. 2nd Pan-American Meeting on Acoustics http://asa.aip.org/meetings.html

2011

22–27 May	Prague, Czech Republic. International Conference on Acoustics, Speech, and Signal Processing (IEEE ICASSP 2011). http://www.icassp2011.com
23–27 May	Seattle, WA, USA. 160th meeting of the Acoustical Society of America. http://asa.aip.org/meetings.html
27 June–1 July	Aalborg, Denmark. Forum Acusticum 2011. http://www.fa2011.org
24–28 July	Tokyo, Japan. 19th International Symposium on Nonlinear Acoustics (ISNA 19). http://www.isna19.com
27–31 August	Florence, Italy. Interspeech 2011. http://www.interspeech2011.org
4–7 September	Osaka, Japan. Internoise 2011. http://www.internoise2011.com
5–8 September	Gdansk, Poland. International Congress on Ultrasonics. http://icu2011.ug.edu.pl/index.html

2012

20–25 March	Kyoto, Japan. IEEE International Conference on Acoustics, Speech, and Signal Processing. http:// www.icssp2012.com

2013

26–31 March	Vancouver, Canada. 2013 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP). http://www.icassp2013.com
2–7 June	Montréal, Canada. 21st International Congress on Acoustics (ICA 2013). http://www.ica2013montreal.org

Revision List

New Associates

Anthony, Samuel, 1415 Blairwood Ln., Chesapeake, VA 23320
Athaide, Christopher, 6030 Charlotte St., Houston, TX 77005
Avery, Christopher, 234 Causeway St., Unit 1210, Boston, MA 02114
Avsic, Tom, ELAC Nautik GmbH, Neufeldstr. 10, 24118 Kiel, Germany
Bayram, Baris, Middle East Technical Univ., Electrical and Electronics Eng, Rm. C-105,
Cankaya, 06531 Ankara, Turkey
Bingham, Jill, 1332 East Capitol St., #T-1, Washington, DC 20003
Brown, Andrew, Speech and Hearing Sci., Univ. of Washington, 1417 NE 42nd St., Seattle, WA 98105
Burnfield, Jason, 1907 Oakhurst Cir., Norman, OK 73071
Caresta, Mauro, UNSW, Mechanical Eng., Gate 14, Barker St., Sydney 2052, Australia
Caute, Didier, Marport, 151 rue de larmor, 56100 Lorient, France
Confrey, John, 2298 La Playa Dr., N., Costa Mesa, CA 92627
Digweed, Shannon, Grant MacEwan College, Dept. of Psychology, City Center Campus, Edmonton, AB T5J 2P2, Canada
Drager, Katie, Univ. of Hawaii at Manoa, Linguistics, 569 Moore Hall, 1890 East-West Rd., Honolulu, HI 96822
Duboc, Pedro, Rua Glaucio Gil 777 BL1 Ap 101, Rio de Janeiro 28950-000, Brazil
Dugan, Peter, Cornell Univ., Bioacoustics Res. Program, 159 Sapsucker Woods Rd., Ithaca, NY 14850
East, James, 9917 Bellflower Way, Knoxville, TN 37932
Fingerhuth, Sebastian, Hubertusstr. 22, 52064 Aachen, Germany
Fischer, Mark, AguaSonic Acoustics, P.O. Box 1073, Rio Vista, CA 94571

Fisher, John, John Sergio Fisher & Assoc, Inc., 5567 Reseda Blvd., Ste. # 209, Tarzana, CA 91356
Foster, Kathryn, Lincoln Public Schools, SPED-Audiology, 5901 O St., Lincoln, NE 68510
Fox, Jeffrey, 413 Hector St., Ithaca, NY 14850
Franks, John, LytleSound, 1 Lytle Pl., #809, Cincinnati, OH 45202-3528
Geluso, Paul, NYU, 65 Delaware Ave., Delhi, NY 13753
Glaser, Kevin, Mayo Clinic, Radiology, 200 First St. SW, Rochester, MN 55905
Greff, Raphak L., A-Volute, 2 rue du Transit, 59650 Villeneuve D″Ascq, France
Griffin, Benjamin, Univ. of Florida, Mechanical and Aerospace Eng., 125 NEB, Gainesville, FL 32611
Gros, Jaime, Cmo Fuente la Junquera No 90, C1, Zaragoza, Spain
Hanes, Brett, 6533 Huckleberry Ln., Lula, GA 30554
He, Bingrong, 2260 Westchester Dr., #7, Manhattan, KS 66503
Heller, Aaron, SRI International, AI Center, 333 Ravenswood Ave., Menlo Park, CA 94002
Hesford, Andrew, Univ. of Rochester, Electrical and Computer Eng., 601 Elmwood Ave., Box 648, Rochester, NY 14642-8648
Hiryu, Shizuko, Doshisha Univ., 1-3 Tatara Miyakodani, Kyotanabe, 610-0321, Japan
Hu, Fang, Dept. of Math. and Stats., Old Dominion Univ., Norfolk, VA 23508
Hu, Yi, Univ. of Wisconsin-Milwaukee, Electrical Eng. and Computer Science, 2200 E. Kenwood Blvd., Milwaukee, WI 53201
Hunter, Kenneth, 32 Berkley St., Merrimack, NH 03054
Jin, Guoyong, Harbin Eng. Univ., College of Power and Energy, Nantong St. 145, Harbin 150001, P.R. China
Johnson, Robert, 16 Chatterly Ct., Perry Hall, MD 21128
Kenehan, Garrett, 232 Stockbridge Rd., Scituate, MA 02066
Kim, Jin-Yeon, Georgia Inst. of Technology, Mechanical Eng., 801 Ferst Dr., Atlanta, GA 30332
Kim, Gibak, Univ. of Texas at Dallas, Electrical Eng. (EC33), 800 West Campbell Rd., Richardson, TX 75080
Kristoffersen, Jacob, Terma A/S, Hovmarken 4,8520 Lystrup, Denmark
Lahvis, Garet, Oregon Health and Science Univ., Mackenzie Hall, Rm. 2182, 3181 W. Sam
Jackson Park Rd., Mail Code L470, Portland, OR 97239-3011
Le Moyne, Sylvie, 25 rue Marcel Gaujour, F-58130 Guerigny, France
Leasure, Joshua, JEAcoustics, 1705 W. Koenig Ln., Austin, TX 78756
Lee, Kevin, Applied Res. Laboratories, UT-Austin, PO Box 8029, Austin, TX 78713
Lonzaga, Joel, Univ. of Mississippi, NCPA, 1 Coliseum Dr., University, MS 38677
Lovgren, Bo, Saab Underwater Systems AB, Box 910, SE-59129 Motala, Sweden
Luo, Wenbo, 11338 3rd Ave., NE #B, Seattle, WA 98125
Mackey, Brian, Naval Surface Warfare Ctr., Signatures Dept., 9500 MacArthur Blvd., Bldg. 15, West Bethesda, MD 20817
Maki, Garret, 1 Hermann Museum Cir. Dr., Apt. 2081, Houston, TX 77004
Matsuo, Ikuo, Dept. of Information Science, Tohoku Gakuin Univ., 2-1-1 Tenjinzawa, Izumi-ku, Sendai 981-3193 Japan
May-Collado, Laura, Univ. of Puerto Rico-RP, Biology, P.O. Box 23360, San Juan, PR 00931
McGuire, Patrick, McGuire-Wolcott, PO Box 16130, Oakland, CA 94610
Mitchell, Andrew 1817 Glen Wood Dr., Grapevine, TX 76051
Munjak, David, Resound Eng. Corp., 8818—113 Ave., Grande Prairie, AB T8X 1S2, Canada
Murray, James, OASIS, 8904 Moreland Ln., Annandale, VA 22003
Nanayakkara, Asiri, Inst. of Fundamental Studies, Design Eng., Hanthana Rd., Kandy Central 0010, Sir Lanka
Neitzel, Edward, W.L. Gore & Associates, Inc., 201 Airport Rd., P.O. Box 1488, Elkton, MD 21921
Nguyen, Trang, 4209 Marathon Blvd., #203, Austin, TX 78756
Nogawa, Masafumi, 26-23, Motonoyama, Oshima-cho, Toyota, Aichi, Toyota, 473-0935, Japan
Norris, Jeff, HDR, 1907 Beeches Glory Path, Annapolis, MD 21401
Norris, Kristin, 5085 Keysville Ave., Spring Hill, FL 34608
Ohadi, Abdolreza, Amirkabir Univ. of Tech., Mech. Eng. Dept., Amirkabir, Hafez Ave., 1591634311 Tehran, Iran
Orellana, Douglas, Northrop Grumman Corp—US, Systems Eng. Dept, 895 Oceanic Dr, MS 9910, Annapolis, MD 21401
Panahi, Issa, Univ. of Texas at Dallas, EE, 800 W. Campbell Rd., EC33, Richardson, TX 75080
Patin, Arno, Arno Patin Studios LLC, 918 Pomona Rd., Ann Arbor, MI 48103
Pickerel, Craig, SSOE, Inc., Healthcare, 1001 Madison Ave., Toledo, OH 43604
Pinstrup, Thomas, Larsen Strings A/S, Egelykke 3, 6400 Soenderborg, Denmark
Pinyard, Scott, Investment Eng., 358 Main St., Yarmouth, ME 04096
Reed, Sierra, 11120 Allison Parke Trail, Austin, TX 78750
Reinhardt, William, 12025 Sherwood Dr., Hagerstown, MD 21742,
Robertson, Charles, Adaptive Methods Inc., Advanced Development, 15825 Shady Grove Rd., Suite 135, Rockville, MD 20850
Schevciw, Andre, Qualcomm, Inc., Multimedia R&D and Standards, 5775 Morehouse Dr., San Diego, CA 92122
Seiner, John, Univ. of Mississippi, National Center for Physical Acoustics, 1 Coliseum Dr., University, MS 38677
Solberg, Marianne, Erdalsvegen 17 A, 5306 Erdal, Norway
Stoler, David, 10 Mountain Ave., Wanaque, NJ 07465
Stradal, Octavian, Univ. of Pardubice, Informatics in Transport, Studentska 95,532 10 Pardubice, Czech Republic
Thibeault, Michael, Ultra Electronics Ocean System, Eng., 115 Bay State Dr., Braintree, MA 02184
Thomas, Brad, Wenger Corporation, Design Eng., 555 Park Dr., Owatonna, MN 55060
Tillett, Jason, 278 Valley Green Dr, Penfield, NY 14526
Trainor, Laurel, McMaster Univ., Psychology, Neuroscience & Behavior, 1280 Main St. West, Hamilton, ON L8S 4K1, Canada
Turnbull, Robert, Gentex Corporation, 600 N. Centennial St., Zeeland, MI 49464
Valente, Daniel, Boys Town Natl. Res. Hospital, Res., 555 N. 30th St., Omaha, NE 68131
Vedvik, Richard, KJWW Eng. Consultants, 623 26th Ave., Rock Island, IL 61201
Wang, John, 7 Kerry Ct., Pearl River, NY 10965
Wenzel, Stuart, 39660 Eureka Dr., Newark, CA 94560
Weyker, John, Flex-Tek Group, AeroSonics, 10715 Glenway Dr, Houston, TX 77070
Willsallen, Scott, Auditoria Pty. Ltd., PO Box 103, Camperdown 1450, Australia
Witchard, Clyde, 19 Royal Victoria Park, Bristol, BS10 6TD, UK
Wollstrom, Magnus, Martec Ltd., 1888 Brunswick St., Halifax, NS B3 J 3J8, Canada
Workman, Dennis, Trimble Navigation, Advanced Devices, 15940 Jackson Oaks Dr., Morgan Hill, CA 95037
Wurtz, Michael, Lightspeed Aviation, Eng., 6135 SW Jean Rd., Lake Oswego, OR 97035
Zhao, Dan, Silixa House, 230 Centennial Park, Elstree, Hertfordshire WD6 3SN, UK

New Students

Adams, Tyler, 1 Irving Pl., #3F, Troy, NY 12180
Aldoori, Ghassan, Univ. of South, Mechanical Eng., Student, Toowoomba, QLD, 4350, Australia
Amponsah, Robert, 95 Crittenden Blvd., Rochester, NY 14620
Anastasiadis, Pavlos, Univ. of Hawaii at Manoa, Mechanical Eng., 2540 Dole St., Honolulu, HI 96822
Angle, Glenda, 303 Wildwood Ln., Rockwall, TX 75087
Arctander, Kaitlyn, Macalester College, 1600 Grand Ave., St. Paul, MN 55105
Bacon, Cathy, 1616 Quails Nest Dr., Fort Worth, UT 76177
Bardhan, Neil, Univ. of Rochester, Brain & Cognitive Sci., Meliora Hall Box 270268, Rochester, NY 14627
Betancourt, Vanessa, 260 Crandon Blvd., Suite #32, Key Biscayne, FL 33149
Blair, Ballard, 410 Memorial Dr., Apt., 154, Cambridge, MA 02139
Blevins, Matthew, 429 South Chauncey, West Lafayette, IN 47906
Bonomo, Anthony L., 8000 Boteler Ln., Apt. 247-B2, College Park, MD 20740

Burton, Lisa, 410 Memorial Dr., Apt. 242, Cambridge, MA 02139
Chabassier, Juliette, INRIA Rocquencourt, Equipe POEMS—Bat 13, Domaine de Voluceau,
Rocquencourt BP 105, 78153 Le Chesnay Cedex, France
Chen, Jer-Ming, Univ. of New South Wales, Physics, School of Physics, UNSW, Sydney, NSW 2052, Australia
Choi, Wonjae, Univ. of Cambridge, Eng. Dept., Trumpington St., Cambridge CB2 1PZ, UK
Cuadrado, Osvaldo, Cristobal Colon Ave., #6020, Apt. 202, Las Condes, Santiago, Chile
Daniels, Eric, 18810 Chandler St., Omaha, NE 68136
Davis, K. Allen, 626 S. Pugh St., Apt. 17, State College, PA 16801
Doutres, Olivier, 2596 Le Moyne, Sherbrooke, QC J1K1S5, Canada
Draudt, Andrew, 68 Egerton Rd., Arlington, MA 02474
Dziklinski, Richard, 2551 Wenona Dr., Wixom, MI 48393
Embleton, Steve, 3301 Denver Ave., Austin, TX 78723
Hansen, Hans, Steinrader Weg 89, 23558 Lubeck, Germany
Hasanov, Azar, 1800 Beaumont Dr., Apt. 132, Norman, OK 73071
Hegland, Erica, 146 E. Oak St., Villa Park, IL 60181
Hsieh, Michelle, 9525 N. Capital of TX Hwy #224, Austin, TX 78759
Hsieh, Yu-Hao, Georgia Inst. of Tech., Mechanical Eng., 771 Ferst Dr., Atlanta, GA 30332
Jenkin, Lance, 49 Bell Rd., Cape Town, Republic South Africa
Jolibois, Alexandre, 762 Whitehall Rd., State College, PA 16801
Joshi, Sumedh, Applied Res. Labs., 10000 Burnet Rd., Austin, TX 78758
Kang, Hijo, Stony Brook Univ., Linguistics, S201, SBS Bldg., Stony Brook, NY 11794-4376
Kashdan, Lia, 4700 W. Guadalupe St., Apt. A326, Austin, TX 78751
Kirk, Karen, 401 W. 39th St., Apt. 208, Austin, TX 78751
Kuntzman, Michael, 2601 Scofield Ridge Pkwy., Apt. 826, Austin, TX 78727
Lacross, Amy, Univ. of Arizona, 200 E. Douglass, Tucson, AZ 85721
Lai, Puxiang, Boston Univ., Mechanical Eng., 110 Cummington St., Boston, MA 02215
Lewis, George, Cornell Univ., Biomedical Eng., 2250 N. Triphammer Rd., T8B, Ithaca, NY 14850
Macaulay, Eric, 14805 Abbey Ln., Apt. A3, Bath, MI 48808
Mallory, Melanie, 1906 18th St., Lubbock, TX 79401
Mares, David, 2350 E. Water St., Apt. B214, Tucson, AZ 85719
Mathias, Delphine K., Scripps Inst. of Oceanography, MPL, 9500 Gilman Dr., La Jolla, CA 97037-0238
McGowan, Kevin, Univ. of Michigan, Linguistics, 440 Lorch Hall, 611 Tappan St., Ann Arbor, MI 48105
McNeese, Andrew, 8615 Rockwood Ln., Apt. 225, Austin, TX 78705
Moss, Brian, 32 Ard Aulin, Skehacreggaun, Mungret, Limerick LIMK, Ireland
Mylavarapu, Sandeep, 5200 Anthony Wayne Dr., Apt. 614, Detroit, MI 48202
Naify, Christina, Univ of Southern California, Materials Science, 3651 Watt Way, VHE 416, Los Angeles, CA 90089
Nenadic, Ivan, 1523 21st Ave., NE, Rochester, MN 55906
Nguyen, Kenneth, 3206 Bonnebridge Way Blvd., Houston, TX 77082
Niculescu, Anca Maria, Univ. of Louisiana-Lafayette, Physics, PO Box 44210, Lafayette, LA 70504
Obrien, Kathleen, 500 W. 55th St., Austin, TX 78751
Papp, Joseph, 12 Northgate Ct., North Grafton, MA 01536-1589
Parakkal, Santosh, 111 Fairway Dr., Senatobia, MS 38668
Park, Sang Bum, Univ. of Florida, School of Architecture, 4440 SW Archer Rd., #1705, Gainesville, FL 32608
Perng, Chao-Yu, 698 Menlo Oaks Dr., Menlo Park, CA 94025
Pierzchala, Timothy, 6055 N. Nassau, Chicago, IL 60631
Pineda, Marisa, 74 Barnes Ct., #401, Stanford, CA 94305
Pizarek, Tom, 2606 Enfield Rd., Apt. 212, Austin, TX 78703
Poulain, Kieran, 1177 S. Atherton St., State College, PA 16801
Radhakrishnan, Sreedivya, 496 County Rt. 49, Middletown, NY 10940
Randin, David, 520 E. Highland, Villa Park, IL 60181
Reis, Callie, 2013 Arbor Forest Dr., Marietta, GA 30064
Rife, Brian, 101 Chowning Pl., Blacksburg, VA 24060
Round, Jillian, 2523 Ohio Dr., Apt. #501, Plano, TX 75093
Samlan, Robin, 1681 N. Pima Ct., Tucson, AZ 85716
Schexnayder, Philip, 413 Saint Louis St., Lafayette, LA 70506
Sieck, Caleb, 7112 Jones Cir., Apt. 7, Omaha, NE 68106
Siedenburg, Kai, An der Allee 22a, 27568 Bremerhaven, Germany
Syrika, Asimina, Univ. of Wisconsin-Madison, Communicative Disorders, 1975 Willow Dr., Madison, WI 53706
Taghavi, Hamidreza, Chalmers Univ., Signals & Systems, Horsalsvagen 9, SE-412 96, Goteborg, Sweden
Tarr, Eric, 885 Copeland Rd., Columbus, OH 43212
Tatarchuk, John, 11500 Jollyville Rd., Apt. 3623, Austin, TX 78759
Tengelsen, Daniel, Brigham Young Univ., Physics and Astronomy, N-283 ESC, Provo, UT 84602
Valtierra, Robert, Mechanical Eng., Boston Univ., 110 Cummington St., Boston, MA 02215
Wang, Qi, Boston Univ., Mechanical Eng., 110 Cummington St., Boston, MA 02215
Xu, Hongan, 4417 2nd Ave., Detroit, MI 48201
Zeman, John, 1554 East Marconi Ave., Phoenix, AZ 85022
Zhang, Xuefeng, 630 Merrick St., Apt. 809, Detroit, MI 48202

New Corresponding Electronic Associates

A. Fatah, Moataz, 75G Pyramids Valley, Cairo, Egypt
De Jager, Morni, MENCO, Environment, 642B Corinne St., Garsfontein-east, Pretoria 0060, Swaziland
Gu, Myeong Gweon, MEMS Solution Inc, 4F, IAE, 633-2 Goanri Baekam, Yongin Kyunggi-Do, 449-863, Korea
Havranek, Zdenek, Brno Univ. of Technology, UAMT-FEKT, Kolejni 4, 61200 Brno, Czech Republic
Moon, Gun, Sovico Corporation, Res. & Development Ctr., 1027-5, Bangbae 3-dong, Seocho-gu, Seoul 137-851, Korea,
Nadgir, Madhusudhan, #52 Thimmayah Rd., W.C.R, Manjunathnagar Rajajinagar, Bangalore 560010, India
Ortega Romero, Luis Miguel, Sedam, Bioarchitec & Acustical, Abasolo, #503, 42000 Pachuca, Estado De FDM, Mexico
Sahin, Zekeriyya, Aselsan Inc., Sonar Systems Dept., Mehmet Akif Ersoy Mah. 296., Cad. no:16, 06370 Ankara, Turkey
Tay, Boh Tsun, Arco Strings Music, Lot 288 Ground Floor, Westwood, Tabuan Rd., 93100 Kuching, Sarawak, Malaysia

Reinstated

B. J. Sotirin, *Fellow*

Associates Elected Members

M. Garai, W. Hammon, L. M. Heller, M. E. Hermes, E. W. Large, S-K. Lau, K. Lee, J. Lentz, J. Liu, G. S. Madaras, G. S. Morrison, M. J. Pettersson, I. A. Rodriguez, G. Rosenhouse, E. C. Salter, P. Smaragdis, B. V. Tucker, N. L. Warner

Student to Associate

J. Brannstromg, M. Canney, E. Echols, M. Hornikx, Y. Jing, P. J. Lee, A. Marui, C. M. Siciliano, R. Winkler, J. A. Zybura

Associate to Student

B. A. Jones

Resigned

M. M. Boone, Y. Higurashi, H. G. Schneider, *Members*
G. Burghardt, R. Drullman, R. K. Handa, G. Heinrich, G. Kawai, *Associates*
T. De Ryck, J. Herman, A. R. Mayr, T. Oesterlein, *Students*

Deceased

M. A. Breazeale, G. Fant, G. Raisbeck, E. A. G. Shaw, C. H. Sherman, *Fellows*
H. A. Conklin, B. Hosten, D. A. Thomas, *Members*
B. Larkin, *Associates*
G. Overson, *Student*

Fellows	920
Members	2091
Associates	3526
Students	1165
	7702

ACOUSTICAL STANDARDS NEWS

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

Paul D. Schomer, Standards Director
Schomer and Associates, 2117 Robert Drive, Champaign, IL 61821 [Tel.: (217) 359-6602; Fax: (217) 359-3303; e-mail: Schomer@SchomerAndAssociates.com]

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

Comments are welcomed on all material in Acoustical Standards News.

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

Standards Meetings Calendar—National

Accredited Standards Committees S1, Acoustics; S2, Mechanical Vibration and Shock; S3, Bioacoustics; S3/SC 1, Animal Bioacoustics; and S12, Noise, along with the U.S. Technical Advisory Groups to ISO/TC 43, ISO/TC 43/SC 1, ISO/TC 108 and its five Subcommittees, and IEC/TC 29, ASACOS and the Standards Plenary Group will meet in conjunction with the Joint 159th ASA Meeting and Noise-Con 2010 to be held in Baltimore, Maryland, **19-23 April 2010**. Additional details will be provided when available.

Standards Meetings Calendar—International

20–24 September 2010

- **ISO/TC 108/SC 4** and **ISO/TC 108/SC 5**, London, UK.

27 September–1 October 2010

- **ISO/TC 108** and its subcommittees **ISO/TC 108/SC 2**, **SC 3** and **SC 6**, London, UK.

For additional information on these meetings, please contact the Standards Secretariat.

Accredited Standards Committee on Acoustics, S1

(P. Battenberg, Chair; R.J. Peppin, Vice Chair)

Scope: Standards, specifications, methods of measurement and test, and terminology in the field of physical acoustics including architectural acoustics, electroacoustics, sonics and ultrasonics, and underwater sound, but excluding those aspects which pertain to biological safety, tolerances and comfort.

S1 WORKING GROUPS

S1/Advisory—Advisory Planning Committee to S1 (P. Battenberg, Chair; R.J. Peppin, Vice Chair);

S1/WG1—Standard Microphones and their Calibration (V. Nedzelnitsky);

S1/WG2—Attenuation of Sound in the Atmosphere (A.H. Marsh);

S1/WG4—Measurement of Sound Pressure Levels in Air (VACANT, Chair; E. Dunens, Vice Chair);

S1/WG5—Band Filter Sets (A.H. Marsh);

S1/WG9—Calibration of Underwater Electroacoustic Transducers (R.M. Drake);

S1/WG16—FFT Acoustical Analyzers (R.L. McKinley);

S1/WG17—Sound Level Meters and Integrating Sound Level Meters (G.R. Stephany);

S1/WG19—Insertion Loss of Windscreens (A.J. Campanella);

S1/WG20—Ground Impedance (Measurement of Ground Impedance and Attenuation of Sound Due to the Ground) (K. Attenborough, Chair; J. Sabatier, Vice Chair);

S1/WG22—Bubble Detection and Cavitation Monitoring (Vacant);

S1/WG27—Acoustical Terminology (J.S. Vipperman).

S1 STANDARDS ON ACOUSTICS

ANSI S1.1-1994 (R 2004) American National Standard Acoustical Terminology.

ANSI S1.4-1983 (R 2006) American National Standard Specification for Sound Level Meters. This Standard includes **ANSI S1.4A-1985 (R 2006)** Amendment to ANSI S1.4-1983.

ANSI S1.6-1984 (R 2006) American National Standard Preferred Frequencies, Frequency Levels, and Band Numbers for Acoustical Measurements.

ANSI S1.8-1989 (R 2006) American National Standard Reference Quantities for Acoustical Levels.

ANSI S1.9-1996 (R 2006) American National Standard Instruments for the Measurement of Sound Intensity.

ANSI/ASA S1.11-2004 (R 2009) American National Standard Specification for Octave-Band and Fractional-Octave-Band Analog and Digital Filters.

ANSI S1.13-2005 American National Standard Measurement of Sound Pressure Levels in Air.

ANSI/ASA S1.14-1998 (R 2008) American National Standard Recommendations for Specifying and Testing the Susceptibility of Acoustical Instruments to Radiated Radio-Frequency Electromagnetic Fields, 25 MHz to 1 GHz.

 0001-4966/2010/127(1)/594/9/$25.00

ANSI S1.15/Part 1-1997 (R 2006) American National Standard Measurement Microphones, Part 1: Specifications for Laboratory Standard Microphones.

ANSI S1.15/Part 2-2005 American National Standard Measurement Microphones, Part 2: Primary Method for Pressure Calibration of Laboratory Standard Microphones by the Reciprocity Technique.

ANSI S1.16-2000 (R 2005) American National Standard Method for Measuring the Performance of Noise Discriminating and Noise Canceling Microphones.

ANSI S1.17/Part 1-2004 American National Standard Microphone Windscreens—Part 1: Measurements and Specification of Insertion Loss in Still or Slightly Moving Air.

ANSI S1.18-1999 (R 2004) American National Standard Template Method for Ground Impedance.

ANSI S1.20-1988 (R 2003) American National Standard Procedures for Calibration of Underwater Electroacoustic Transducers.

ANSI S1.22-1992 (R 2007) American National Standard Scales and Sizes for Frequency Characteristics and Polar Diagrams in Acoustics.

ANSI S1.24 TR-2002 (R 2007) ANSI Technical Report Bubble Detection and Cavitation Monitoring.

ANSI S1.25-1991 (R 2007) American National Standard Specification for Personal Noise Dosimeters.

ANSI/ASA S1.26-1995 (R 2009) American National Standard Method for Calculation of the Absorption of Sound by the Atmosphere.

ANSI S1.40-2006 American National Standard Specifications and Verification Procedures for Sound Calibrators.

ANSI S1.42-2001 (R 2006) American National Standard Design Response of Weighting Networks for Acoustical Measurements.

ANSI S1.43-1997 (R 2007) American National Standard Specifications for Integrating-Averaging Sound Level Meters.

Accredited Standards Committee on Mechanical Vibration and Shock, S2

(A.T. Herfat, Chair; C.F. Gaumond, Vice Chair)

Scope: Standards, specifications, methods of measurement and test, and terminology in the field of mechanical vibration and shock, and condition monitoring and diagnostics of machines, including the effects of exposure to mechanical vibration and shock on humans, including those aspects which pertain to biological safety, tolerance and comfort.

S2 WORKING GROUPS

S2/WG1—S2 Advisory Planning Committee (A.T. Herfat, Chair; C.F. Gaumond, Vice Chair);

S2/WG2—Terminology and Nomenclature in the Field of Mechanical Vibration and Shock and Condition Monitoring and Diagnostics of Machines (D.J. Evans);

S2/WG3—Signal Processing Methods (T.S. Edwards);

S2/WG4—Characterization of the Dynamic Mechanical Properties of Viscoelastic Polymers (W.M. Madigosky, Chair; J. Niemiec, Vice Chair);

S2/WG5—Use and Calibration of Vibration and Shock Measuring Instruments (D.J. Evans, Chair; B.E. Douglas, Vice Chair);

S2/WG6—Vibration and Shock Actuators (G. Booth);

S2/WG7—Acquisition of Mechanical Vibration and Shock Measurement Data (B.E. Douglas);

S2/WG8—Analysis Methods of Structural Dynamics (M. Mezache);

S2/WG9—Training and Accreditation (R.L. Eshleman, Chair; D. Corelli, Vice Chair);

S2/WG10—Measurement and Evaluation of Machinery for Acceptance and Condition (R.L. Eshleman, Chair; H.C. Pusey, Vice Chair);

S2/WG10/Panel 1—Balancing (R.L. Eshleman);

S2/WG10/Panel 2—Operational Monitoring and Condition Evaluation (R. Bankert);

S2/WG10/Panel 3—Machinery Testing (R.L. Eshleman);

S2/WG10/Panel 4—Prognosis (A.J. Hess);

S2/WG10/Panel 5—Data Processing, Communication, and Presentation (K. Bever);

S2/WG11—Measurement and Evaluation of Mechanical Vibration of Vehicles (Vacant);

S2/WG12—Measurement and Evaluation of Structures and Structural Systems for Assessment and Condition Monitoring (M. Mezache);

S2/WG13—Shock Test Requirements for Shelf-Mounted and Other Commercial Electronics Systems (B. Lang);

S2/WG39 (S3)—Human Exposure to Mechanical Vibration and Shock (D.D. Reynolds, Chair; R. Dong, Vice Chair).

S2 STANDARDS ON MECHANICAL VIBRATION AND SHOCK

ANSI S2.1-2009/ISO 2041:2009 American National Standard Vibration and Shock—Vocabulary (Nationally Adopted International Standard).

ANSI S2.2-1959 (R 2006) American National Standard Methods for the Calibration of Shock and Vibration Pickups.

ANSI S2.4-1976 (R 2004) American National Standard Method for Specifying the Characteristics of Auxiliary Analog Equipment for Shock and Vibration Measurements.

ANSI S2.8-2007 American National Standard Technical Information Used for Resilient Mounting Applications.

ANSI/ASA S2.9-2008 American National Standard Parameters for Specifying Damping Properties of Materials and System Damping.

ANSI S2.16-1997 (R 2006) American National Standard Vibratory Noise Measurements and Acceptance Criteria of Shipboard Equipment.

ANSI S2.19-1999 (R 2004) American National Standard Mechanical Vibration—Balance Quality Requirements of Rigid Rotors, Part 1: Determination of Permissible Residual Unbalance, Including Marine Applications.

ANSI S2.20-1983 (R 2006) American National Standard Estimating Air Blast Characteristics for Single Point Explosions in Air, with a Guide to Evaluation of Atmospheric Propagation and Effects.

ANSI S2.21-1998 (R 2007) American National Standard Method for Preparation of a Standard Material for Dynamic Mechanical Measurements.

ANSI S2.22-1998 (R 2007) American National Standard Resonance Method for Measuring the Dynamic Mechanical Properties of Viscoelastic Materials.

ANSI S2.23-1998 (R 2007) American National Standard Single Cantilever Beam Method for Measuring the Dynamic Mechanical Properties of Viscoelastic Materials.

ANSI S2.24-2001 (R 2006) American National Standard Graphical Presentation of the Complex Modulus of Viscoelastic Materials.

ANSI/ASA S2.25-2004 (R 2009) American National Standard Guide for the Measurement, Reporting, and Evaluation of Hull and Superstructure Vibration in Ships.

ANSI S2.26-2001 (R 2006) American National Standard Vibration Testing Requirements and Acceptance Criteria for Shipboard Equipment.

ANSI S2.27-2002 (R 2007) American National Standard Guidelines for the Measurement and Evaluation of Vibration of Ship Propulsion Machinery.

ANSI/ASA S2.28-2009 American National Standard Guide for the Measurement and Evaluation of Broadband Vibration of Surface Ship Auxiliary Rotating Machinery.

ANSI/ASA S2.29-2003 (R 2008) American National Standard Guide for the Measurement and Evaluation of Vibration of Machine Shafts on Shipboard Machinery.

ANSI S2.31-1979 (R 2004) American National Standard Methods for the Experimental Determination of Mechanical Mobility, Part 1: Basic Definitions and Transducers.

ANSI S2.32-1982 (R 2004) American National Standard Methods for the Experimental Determination of Mechanical Mobility, Part 2: Measurements Using Single-Point Translational Excitation.

ANSI S2.34-1984 (R 2005) American National Standard Guide to the Experimental Determination of Rotational Mobility Properties and the Complete Mobility Matrix.

ANSI S2.42-1982 (R 2004) American National Standard Procedures for Balancing Flexible Rotors.

ANSI S2.43-1984 (R 2005) American National Standard Criteria for Evaluating Flexible Rotor Balance.

ANSI S2.46-1989 (R 2005) American National Standard Characteristics to be Specified for Seismic Transducers.

ANSI S2.48-1993 (R 2006) American National Standard Servo-Hydraulic Test Equipment for Generating Vibration—Methods of Describing Characteristics.

ANSI S2.60-1987 (R 2005) American National Standard Balancing Machines—Enclosures and Other Safety Measures.

ANSI S2.61-1989 (R 2005) American National Standard Guide to the Mechanical Mounting of Accelerometers.

ANSI/ASA S2.62-2009 American National Standard Shock Test Requirements for Equipment in a Rugged Shock Environment.

ANSI S2.70-2006 *(Revision of ANSI S3.34-1986)* American National Standard Guide for the Measurement and Evaluation of Human Exposure to Vibration Transmitted to the Hand.

ANSI S2.71-1983 (R 2006) (*Reaffirmation and redesignation of ANSI S3.29-1983*) American National Standard Guide to the Evaluation of Human Exposure to Vibration in Buildings.

ANSI S2.72/Part 1-2002 (R 2007)/ISO 2631-1:1997 *(Reaffirmation and redesignation of ANSI S3.18/Part 1-2002/ISO 2631-1:1997)* American National Standard Mechanical vibration and shock—Evaluation of human exposure to whole-body vibration—Part 1: General requirements (Nationally Adopted International Standard).

ANSI S2.72/Part 4-2003 (R 2007)/ISO 2631-4:2001 *(Reaffirmation and Redesignation of ANSI S3.18/Part 4—2003/ISO 2631-4:2001)* American National Standard Mechanical vibration and shock—Evaluation of human exposure to whole-body vibration—Part 4: Guidelines for the evaluation of the effects of vibration and rotational motion on passenger and crew comfort in fixed-guideway transport systems (Nationally Adopted International Standard).

ANSI S2.73-2002 (R 2007)/ISO 10819:1996 *(Redesignation of ANSI S3.40-2002/ISO 10819:1996)* American National Standard Mechanical vibration and shock—Hand-arm vibration—Method for the measurement and evaluation of the vibration transmissibility of gloves at the palm of the hand (Nationally Adopted International Standard).

Accredited Standards Committee on Bioacoustics, S3

(C.A. Champlin, Chair; D.A. Preves, Vice Chair)

Scope: Standards, specifications, methods of measurement and test, and terminology in the fields of psychological and physiological acoustics, including aspects of general acoustics, which pertain to biological safety, tolerance and comfort.

S3 WORKING GROUPS

S3/Advisory—Advisory Planning Committee to S3 (C.A. Champlin, Chair; D.A. Preves, Vice Chair);

S3/WG35—Audiometric Equipment (R.L. Grason);

S3/WG36—Speech Intelligibility (R.S. Schlauch);

S3/WG37—Coupler Calibration of Earphones (C.J. Struck);

S3/WG39—Human Exposure to Mechanical Vibration and Shock—Parallel to ISO/TC 108/SC 4 (D.D. Reynolds, Chair; R. Dong, Vice Chair);

S3/WG43—Method for Calibration of Bone Conduction Vibrators (J.D. Durrant);

S3/WG48—Hearing Aids (D.A. Preves);

S3/WG56—Criteria for Background Noise for Audiometric Testing (J.R. Franks);

S3/WG59—Measurement of Speech Levels (M.C.Killion and L.A. Wilber, Co-Chairs);

S3/WG62—Impulse Noise with Respect to Hearing Hazard (G.R. Price);

S3/WG67—Mankinis (M.D. Burkhard);

S3/WG72—Measurement of Auditory Evoked Potentials (R.F. Burkard);

S3/WG79—Methods for Calculation of the Speech Intelligibility Index (C.V. Pavlovic);

S3/WG80—Probe-tube Measurements of Hearing Aid Performance (W.A. Cole);

S3/WG81—Hearing Assistance Technologies (L. Thibodeau and L.A. Wilber, Co-Chairs);

S3/WG82—Basic Vestibular Function Test Battery (C. Wall);

S3/WG83—Sound Field Audiometry (T.R. Letowski);

S3/WG84—Otoacoustic Emissions (Vacant);

S3/WG88—Standard Audible Emergency Evacuation and Other Signals (R. Boyer);

S3/WG89—Spatial Audiometry in Real and Virtual Environments (J. Besing);

S3/WG91—Text-to-Speech Synthesis Systems (C. Bickley and A.K. Syrdal, Co-Chairs).

S3 Liaison Group

S3/L-1 U. S. TAG Liaison to IEC/TC 87 Ultrasonics (W.L. Nyborg).

S3 STANDARDS ON BIOACOUSTICS

ANSI/ASA S3.1-1999 (R 2008) American National Standard Maximum Permissible Ambient Noise Levels for Audiometric Test Rooms.

ANSI/ASA S3.2-2009 American National Standard Method for Measuring the Intelligibility of Speech over Communication Systems.

ANSI S3.4-2007 American National Standard Procedure for the Computation of Loudness of Steady Sounds.

ANSI S3.5-1997 (R 2007) American National Standard Methods for Calculation of the Speech Intelligibility Index.

ANSI S3.6-2004 American National Standard Specification for Audiometers.

ANSI/ASA S3.7-1995 (R 2008) American National Standard Method for Coupler Calibration of Earphones.

ANSI S3.13-1987 (R 2007) American National Standard Mechanical Coupler for Measurement of Bone Vibrators.

ANSI/ASA S3.20-1995 (R 2008) American National Standard Bioacoustical Terminology.

ANSI/ASA S3.21-2004 (R 2009) American National Standard Methods for Manual Pure-Tone Threshold Audiometry.

ANSI S3.22-2003 American National Standard Specification of Hearing Aid Characteristics.

ANSI S3.25-1989 (R 2003) American National Standard for an Occluded Ear Simulator.

ANSI S3.35-2004 American National Standard Method of Measurement of Performance Characteristics of Hearing Aids under Simulated Real-Ear Working Conditions.

ANSI S3.36-1985 (R 2006) American National Standard Specification for a Manikin for Simulated *in situ* Airborne Acoustic Measurements.

ANSI S3.37-1987 (R 2007) American National Standard Preferred Earhook Nozzle Thread for Postauricular Hearing Aids.

ANSI S3.39-1987 (R 2007) American National Standard Specifications for Instruments to Measure Aural Acoustic Impedance and Admittance (Aural Acoustic Immittance).

ANSI S3.41-1990 (R 2008) American National Standard Audible Emergency Evacuation Signal.

ANSI S3.42-1992 (R 2007) American National Standard Testing Hearing Aids with a Broad-Band Noise Signal.

ANSI S3.44-1996 (R 2006) American National Standard Determination of Occupational Noise Exposure and Estimation of Noise-Induced Hearing Impairment.

ANSI/ASA S3.45-2009 American National Standard Procedures for Testing Basic Vestibular Function.

ANSI S3.46-1997 (R 2007) American National Standard Methods of Measurement of Real-Ear Performance Characteristics of Hearing Aids.

Accredited Standards Subcommittee on Animal Bioacoustics, S3/SC1

(D.K. Delaney, Chair; M.C. Hastings, Vice Chair)

Scope: Standards, specifications, methods of measurement and test, instrumentation and terminology in the field of psychological and physiological acoustics, including aspects of general acoustics, which pertain to biological safety, tolerance and comfort of non-human animals, including both risk to individual animals and to the long-term viability of populations. Animals to be covered may potentially include commercially grown food animals; animals harvested for food in the wild; pets; laboratory animals; exotic species in zoos, oceanaria or aquariums; or free-ranging wild animals.

S3/SC1 WORKING GROUPS

S3/SC 1/WG1—Animal Bioacoustics Terminology (A.E. Bowles);

S3/SC 1/WG2—Effects of Sound on Fish and Turtles (R.R. Fay and A.N. Popper, Co-Chairs);

S3/SC 1/WG3—Underwater Passive Acoustic Monitoring for Bioacoustic Applications (A.M. Thode);

S3/SC 1/WG4—Description and Measurement of the Ambient Sound in Parks, Wilderness Areas, and Other Quiet and/or Pristine Areas (K. Fristrup and G.R. Stanley, Co-Chairs).

Accredited Standards Committee on Noise, S12

(W.J. Murphy, Chair; R.D. Hellweg, Vice Chair)

Scope: Standards, specifications, and terminology in the field of acoustical noise pertaining to methods of measurement, evaluation, and control, including biological safety, tolerance and comfort and physical acoustics as related to environmental and occupational noise.

S12 WORKING GROUPS

S12/Advisory—Advisory Planning Committee to S12 (W.J. Murphy and R.D. Hellweg);

S12/WG3—Measurement of Noise from Information Technology and Telecommunications Equipment (K.X.C. Man);

S12/WG11—Hearing Protector Attenuation and Performance (E.H. Berger);

S12/WG14—Measurement of the Noise Attenuation of Active and/or Passive Level Dependent Hearing Protective Devices (W.J. Murphy);

S12/WG15—Measurement and Evaluation of Outdoor Community Noise (P.D. Schomer);

S12/WG23—Determination of Sound Power (B.M. Brooks and J. Schmitt, Co-chairs);

S12/WG32—Revision of ANSI S12.7-1986 Methods for Measurement of Impulse Noise (W.A. Ahroon);

S12/WG38—Noise Labeling in Products (R.D. Hellweg);

S12/WG40—Measurement of the Noise Aboard Ships (S.P. Antonides, Chair; S.A. Fisher, Vice Chair);

S12/WG41—Model Community Noise Ordinances (L.S. Finegold, Chair; B.M. Brooks, Vice Chair);

S12/WG44—Speech Privacy (G.C. Tocci, Chair; D.M. Sykes, Vice Chair);

S12/WG45—Measurement of Occupational Noise Exposure from Telephone Equipment (K.A. Woo, Chair; L.A. Wilber, Vice Chair);

S12/WG46—Acoustical Performance Criteria for Relocatable Classrooms (T. Hardiman and P.D. Schomer, Co-Chairs);

S12/WG47—Underwater Noise Measurements of Ships (M.A. Bahtiarian, Chair; D.J. Vendittis, Vice Chair);

S12/WG48—Railroad Horn Sound Emission Testing (J. Erdreich, Chair; J.J. Earshen, Vice Chair);

S12/WG49—Noise from Hand-operated Power Tools, Excluding Pneumatic Tools (C.S. Hayden, Chair; B.M. Brooks, Vice Chair);

S12/WG50—Information Technology (IT) Equipment in Classrooms (R.D. Hellweg);

S12/WG51—Procedure for Measuring the Ambient Noise Level in a Room (J.G. Lilly);

S12/WG52—Revision of ANSI S12.60-2002 (S. Lind and P.D. Schomer, Co-Chairs).

S12 LIAISON GROUPS

S12/L-1 IEEE 85 Committee for TAG Liaison—Noise Emitted by Rotating Electrical Machines (Parallel to ISO/TC 43/SC 1/WG 13) (R.G. Bartheld);

S12/L-2 Measurement of Noise from Pneumatic Compressors Tools and Machines (Parallel to ISO/TC 43/SC 1/WG 9) (Vacant);

S12/L-3 SAE Committee for TAG Liaison on Measurement and Evaluation of Motor Vehicle Noise (parallel to ISO/TC 43/SC 1/WG 8) (R.F. Schumacher);

S12/L-4 SAE Committee A-21 for TAG Liaison on Measurement and Evaluation of Aircraft Noise (J.D. Brooks);

S12/L-5 ASTM E-33 on Environmental Acoustics (to include activities of ASTM E33.06 on Building Acoustics, parallel to ISO/TC 43/SC 2 and ASTM E33.09 on Community Noise) (K.P. Roy);

S12/L-6 SAE Construction-Agricultural Sound Level Committee (I. Douell);

S12/L-7 SAE Specialized Vehicle and Equipment Sound Level Committee (T.M. Disch);

S12/L-8 ASTM PTC 36 Measurement of Industrial Sound (R.A. Putnam, Chair; B.M. Brooks, Vice Chair).

S12 STANDARDS ON NOISE

ANSI S12.1-1983 (R 2006) American National Standard Guidelines for the Preparation of Standard Procedures to Determine the Noise Emission from Sources.

ANSI/ASA S12.2-2008 American National Standard Criteria for Evaluating Room Noise.

ANSI S12.3-1985 (R 2006) American National Standard Statistical Methods for Determining and Verifying Stated Noise Emission Values of Machinery and Equipment.

ANSI S12.5-2006/ISO 6926:1999 American National Standard Acoustics—Requirements for the Performance and Calibration of Reference Sound Sources Used for the Determination of Sound Power Levels (Nationally Adopted International Standard).

ANSI/ASA S12.6-2008 American National Standard Methods for Measuring the Real-Ear Attenuation of Hearing Protectors.

ANSI S12.7-1986 (R 2006) American National Standard Methods for Measurements of Impulse Noise.

ANSI/ASA S12.8-1998 (R 2008) American National Standard Methods for Determining the Insertion Loss of Outdoor Noise Barriers.

ANSI S12.9/Part 1-1988 (R 2003) American National Standard Quantities and Procedures for Description and Measurement of Environmental Sound, Part 1.

ANSI/ASA S12.9/Part 2-1992 (R 2008) American National Standard Quantities and Procedures for Description and Measurement of Environmental Sound, Part 2: Measurement of Long-Term, Wide-Area Sound.

ANSI/ASA S12.9/Part 3-1993 (R 2008) American National Standard Quantities and Procedures for Description and Measurement of Environmental Sound, Part 3: Short-Term Measurements with an Observer Present.

ANSI S12.9/Part 4-2005 American National Standard Quantities and Procedures for Description and Measurement of Environmental Sound, Part 4: Noise Assessment and Prediction of Long-Term Community Response.

ANSI/ASA S12.9/Part 5-2007 American National Standard Quantities and Procedures for Description and Measurement of Environmental Sound—Part 5: Sound Level Descriptors for Determination of Compatible Land Use.

ANSI/ASA S12.9/Part 6-2008 American National Standard Quantities and Procedures for Description and Measurement of Environmental Sound—Part 6: Methods for Estimation of Awakenings Associated with Outdoor Noise Events Heard in Homes.

ANSI/ASA S12.10-2002 (R 2007)/ISO 7779:1999 American National Standard Acoustics—Measurement of airborne noise emitted by information technology and telecommunications equipment (Nationally Adopted International Standard).

ANSI/ASA S12.11/Part 1-2003 (R 2008)/ISO 10302:1996 (MOD) American National Standard Acoustics—Measurement of noise and vibration of small air-moving devices—Part 1: Airborne noise emission (Modified Nationally Adopted International Standard).

ANSI/ASA S12.11/Part 2-2003 (R 2008) American National Standard Acoustics—Measurement of Noise and Vibration of Small Air-Moving Devices—Part 2: Structure-Borne Vibration.

ANSI/ASA S12.12-1992 (R 2007) American National Standard Engineering Method for the Determination of Sound Power Levels of Noise Sources Using Sound Intensity.

ANSI S12.13 TR-2002 ANSI Technical Report Evaluating the Effectiveness of Hearing Conservation Programs through Audiometric Data Base Analysis.

ANSI/ASA S12.14-1992 (R 2007) American National Standard Methods for the Field Measurement of the Sound Output of Audible Public Warning Devices Installed at Fixed Locations Outdoors.

ANSI/ASA S12.15-1992 (R 2007) American National Standard For Acoustics—Portable Electric Power Tools, Stationary and Fixed Electric Power Tools, and Gardening Appliances—Measurement of Sound Emitted.

ANSI/ASA S12.16-1992 (R 2007) American National Standard Guidelines for the Specification of Noise of New Machinery.

ANSI S12.17-1996 (R 2006) American National Standard Impulse Sound Propagation for Environmental Noise Assessment.

ANSI/ASA S12.18-1994 (R 2009) American National Standard Procedures for Outdoor Measurement of Sound Pressure Level.

ANSI S12.19-1996 (R 2006) American National Standard Measurement of Occupational Noise Exposure.

ANSI S12.23-1989 (R 2006) American National Standard Method for the Designation of Sound Power Emitted by Machinery and Equipment.

ANSI S12.42-1995 (R 2004) American National Standard Microphone-in-Real-Ear and Acoustic Test Fixture Methods for the Measurement of Insertion Loss of Circumaural Hearing Protection Devices.

ANSI/ASA S12.43-1997 (R 2007) American National Standard Methods for Measurement of Sound Emitted by Machinery and Equipment at Workstations and Other Specified Positions.

ANSI/ASA S12.44-1997 (R 2007) American National Standard Methods for Calculation of Sound Emitted by Machinery and Equipment at Workstations and Other Specified Positions from Sound Power Level.

ANSI/ASA S12.50-2002 (R 2007)/ISO 3740:2000 American National Standard Acoustics—Determination of sound power levels of noise sources—Guidelines for the use of basic standards (Nationally Adopted International Standard).

ANSI/ASA S12.51-2002 (R 2007)/ISO 3741:1999 American National Standard Acoustics—Determination of sound power levels of noise sources using sound pressure—Precision method for reverberation rooms. This Standard includes Technical Corrigendum 1-2001. (Nationally Adopted International Standard). *This standard replaces ANSI S12.31-1990 and ANSI S12.32-1990.*

ANSI S12.53/Part 1-1999 (R 2004)/ISO 3743-1:1994 American National Standard Acoustics—Determination of sound power levels of noise sources—Engineering methods for small, movable sources in reverberant fields—Part 1: Comparison method for hard-walled test rooms (Nationally Adopted International Standard). *This standard, along with ANSI S12.53/Part 2-1999 replaces ANSI S12.33-1990.*

ANSI S12.53/Part 2-1999 (R 2004)/ISO 3743-2:1994 American National Standard Acoustics—Determination of sound power levels of noise sources using sound pressure—Engineering methods for small, movable sources in reverberant fields—Part 2: Methods for special reverberation test rooms (Nationally Adopted International Standard). *This standard, along with ANSI S12.53/Part 1-1999 replaces ANSI S12.33-1990.*

ANSI S12.54-1999 (R 2004)/ISO 3744:1994 American National Standard Acoustics—Determination of sound power levels of noise sources using sound pressure—Engineering method in an essentially free field over a reflecting plane. (Nationally Adopted International Standard). *This standard replaces ANSI S12.34-1988.*

ANSI S12.55-2006/ISO 3745:2003 American National Standard Acoustics—Determination of sound power levels of noise sources using sound pressure—Precision methods for anechoic and hemi-anechoic rooms (Nationally Adopted International Standard). *This standard replaces ANSI S12.35-1990.*

ANSI S12.56-1999 (R 2004)/ISO 3746:1995 American National Standard Acoustics—Determination of sound power levels of noise sources using sound pressure—Survey method using an enveloping measurement surface over a reflecting plane (Nationally Adopted International Standard). *This standard replaces ANSI S12.36-1990.*

ANSI/ASA S12.57-2002 (R 2007)/ISO 3747:2000 American National Standard Acoustics—Determination of sound power levels of noise sources using sound pressure—Comparison method *in situ* (Nationally Adopted International Standard).

ANSI/ASA S12.60-2002 (R 2009) American National Standard Acoustical Performance Criteria, Design Requirements, and Guidelines for Schools.

ANSI/ASA S12.60-2009/Part 2 American National Standard Acoustical Performance Criteria, Design Requirements, and Guidelines for Schools, Part 2: Relocatable Classroom Factors.

ANSI/ASA S12.64-2009/Part 1 American National Standard Quantities and Procedures for Description and Measurement of Underwater Sound from Ships — Part 1: General Requirements.

ANSI S12.65-2006 *(Revision of ANSI S3.14-1977)* American National Standard for Rating Noise with Respect to Speech Interference.

ANSI/ASA S12.67-2008 American National Standard Pre-Installation Airborne Sound Measurements and Acceptance Criteria of Shipboard Equipment.

ANSI/ASA S12.68-2007 American National Standard Methods of Estimating Effective A-Weighted Sound Pressure Levels When Hearing Protectors are Worn.

ASA Committee on Standards (ASACOS)

ASACOS (P.D. Schomer, Chair; R.D. Hellweg, Vice Chair)

U.S. Technical Advisory Groups (TAGS) for International Standards Committees

ISO/TC 43 Acoustics, **ISO/TC 43 /SC 1** Noise (P.D. Schomer, U.S. TAG Chair)

ISO/TC 108 Mechanical vibration, shock, and condition monitoring (D.J. Evans, U.S. TAG Chair)

ISO/TC 108/SC 2 Measurement and evaluation of mechanical vibration and shock as applied to machines, vehicles and structures (W.C. Foiles and R.F. Taddeo, U.S. TAG Co-Chairs)

ISO/TC 108/SC 3 Use and calibration of vibration and shock measuring instruments (D.J. Evans, U.S. TAG Chair)

ISO/TC 108/SC 4 Human exposure to mechanical vibration and shock (D.D. Reynolds, U.S. TAG Chair)

ISO/TC 108/SC 5 Condition monitoring and diagnostics of machines (D.J. Vendittis, U.S. TAG Chair; R.F. Taddeo, U.S. TAG Vice Chair)

ISO/TC 108/SC 6 Vibration and shock generating systems (C. Peterson, U.S. TAG Chair)

IEC/TC 29 Electroacoustics (V. Nedzelnitsky, U.S. Technical Advisor)

STANDARDS NEWS FROM THE UNITED STATES

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

American National Standards Call for Comment on Proposals Listed

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

ASHRAE (American Society of Heating, Refrigerating and Air-Conditioning Engineers, Inc.)

Reaffirmations

BSR/ASHRAE 87.3-2001 (R200x), Method of Testing Propeller Fan Vibration-Diagnostic Test Methods (reaffirmation of ANSI/ASHRAE Standard 87.3P-2001 (R2006))

This standard establishes laboratory and on-site diagnostic methods for indentifying causes of vibration problems involving direct-driven propeller fans for condenser cooling in air-conditioning units, heat pumps, and chillers. (October 26, 2009)

ASME (American Society of Mechanical Engineers)

Revisions

BSR/ASME BPVC Section V-200x, Nondestructive Examination (5/14/09 Meeting) (revision of ANSI/ASME BPVC 2007 Edition)

Contains requirements and methods for nondestructive examination (NDE) which are referenced and required by other Sections of the Code. These NDE methods are intended to detect surface and internal imperfections in materials, welds, fabricated parts and components. The following NDE methods are addressed: radiography; ultrasonic; liquid penetrant; magnetic particle; eddy current; visual; leak testing; and acoustic emission. (November 9, 2009)

ASA (ASC S3) (Acoustical Society of America)

Revisions

BSR/ASA S3.22-200x, Specification of Hearing Aid Characteristics (revision of ANSI/ASA S3.22-2009)

Describes air-conduction hearing-aid measurement methods that are particularly suitable for specification and tolerance purposes. Test methods described are output sound pressure level (SPL) with 90-dB input SPL, full-on gain, frequency response, harmonic distortion, equivalent input noise, current drain & induction-coil sensitivity. Specific configurations are given for measuring input SPL to hearing aid. Allowable tolerances in relation to values specified by the manufacturer are given for certain parameters. (November 8, 2009)

American National Standards Final Action

The following American National Standards have received final approval from the ANSI Board of Standards Review.

ASA (ASC S2) (Acoustical Society of America)

New National Adoptions

ANSI ASA S2.1-2009/ISO 2041-2009, Mechanical vibration, shock and condition monitoring—Vocabulary (identical national adoption and revision of ANSI S2.1-2000, ANSI/ISO 2041-1990)

ASA (ASC S3) (Acoustical Society of America)

Revisions

ANSI/ASA S3.25-2009, Occluded Ear Simulator (revision and redesignation of ANSI S3.25-1989 (R2003))

ASA (ASC S12) (Acoustical Society of America)

New Standards

ANSI/ASA S12.60-2009, Part 2, Acoustical Performance Criteria, Design Requirements, and Guidelines for Schools—Part 2: Relocatable Classroom Factors (new standard)

ANSI/ASA S12.64-2009, Quantities and Procedures for Description and Measurement of Underwater Sound from Ships—Part 1: General Requirements (new standard)

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

New Standards

ANSI C63.2-2009, Electromagnetic Noise and Field Strength Instrumentation, 10 Hz–40 GHz Specifications (new standard)

Project Initiation Notification System (PINS)

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

AMCA (Air Movement and Control Association)

BSR/AMCA 300/BSR/ASHRAE 68-201x, Acoustics—Determination of Sound Power Radiated into a Duct by Fans and Other Air-Moving Devices—In-Duct Method (identical national adoption of ISO 5136:2003)

Describes a procedure for the measurement of sound-pressure levels in the inlet or outlet ducts of a fan and a method to use these sound pressure levels to calculate the sound power levels radiated by the fan to the duct system. Project Need: To replace the current ANSI/AMCA 330-ANSI/ASHRAE 68. Stakeholders: Fan manufacturers, acoustic consultants, building owners, acoustic researchers.

ASA (ASC S12) (Acoustical Society of America)

BSR/ASA S12.60-20xx, Part 1, Acoustical Performance Criteria, Design Requirements, and Guidelines for Schools—Part 1: Permanent, Fixed Facilities (revision and redesignation of ANSI S12.60-2002 (R2009))

Provides acoustical performance criteria, design requirements and design guidelines for new or renovated permanent, fixed school classrooms and other learning spaces (excludes modular classrooms). These criteria, requirements, and guidelines are keyed to the acoustical qualities needed to achieve a high degree of speech intelligibility in learning spaces. Test procedures are provided in an annex when conformance to this standard is to be verified. Project Need: To update the existing classroom acoustic performance design guidelines, ANSI S12.60-2002, in order to reflect the existence of Parts 2 and 3, which are nearing completion. The measurement procedures also will be revised to be consistent with better procedures given in Part 2. Stakeholders: School administrators, purchasing agents, school architects and designers, teachers and parents.

ITSDF (Industrial Truck Standards Development Foundation, Inc.)

BSR/ITSDF B56.11.5-2005 (R201x), Measurement of Sound Emitted by Low Lift, High Lift, and Rough Terrain Powered Industrial Trucks (reaffirmation of ANSI/ITSDF B56.11.5-2005)

Establishes the conditions, test procedures, environment, and instrumentation for the determination and reporting of the A-weighted sound pressure level of electric-battery-powered and internal-combustion-engine-powered, low-lift, high-lift, and rough-terrain industrial trucks. This standard provides test procedures for measuring both operator and bystander exposure to sound produced by the trucks. Project Need: To comply with the five-year reaffirmation process. Stakeholders: Users and manufacturers of industrial trucks.

Projects Withdrawn from Consideration

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

ABMA (ASC B3) (American Bearing Manufacturers Association)

BSR/ABMA/ISO 15242-1-200x, Rolling bearings—Measuring methods for vibration—Part 1: Fundamentals (identical national adoption of ISO 15242-1)

BSR/ABMA/ISO 15242-2-200x, Rolling bearings—Measuring methods for vibration—Part 2: Radial ball bearings with cylindrical bore and outside surface (identical national adoption of ISO 15242-2)

BSR/ABMA/ISO 15242-3-200x, Rolling bearings—Measuring methods for vibration—Part 3: Radial spherical and tapered roller bearings with cylindrical bore and outside surface (identical national adoption of ISO 15242-3)

BSR/ABMA/ISO 15242-4-200x, Rolling bearings—Measuring methods for vibration—Part 4: Radial cylindrical roller bearings with cylindrical bore and outside surface (identical national adoption of ISO 15242-4)

STANDARDS NEWS FROM ABROAD

(Partially derived from *ANSI Reporter* and *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

MECHANICAL VIBRATION AND SHOCK (TC 108)

ISO 20806:2009, Mechanical vibration—Criteria and safeguards for the in-situ balancing of medium and large rotors

ISO 7919-2:2009, Mechanical vibration—Evaluation of machine vibration by measurements on rotating shafts—Part 2: Land-based steam turbines and generators in excess of 50 MW with normal operating speeds of 1 500 r/min, 1 800 r/min, 3 000 r/min and 3 600 r/min

ISO 7919-4:2009, Mechanical vibration—Evaluation of machine vibration by measurements on rotating shafts—Part 4: Gas turbine sets with fluid-film bearings

ISO 10816-1/Amd1:2009, Mechanical vibration—Evaluation of machine vibration by measurements on non-rotating parts—Part 1: General guidelines—Amendment 1

ISO 10816-2:2009, Mechanical vibration—Evaluation of machine vibration by measurements on non-rotating parts—Part 2: Land-based steam turbines and generators in excess of 50 MW with normal operating speeds of 1 500 r/min, 1 800 r/min, 3 000 r/min and 3 600 r/min

ISO 10816-4:2009, Mechanical vibration—Evaluation of machine vibration by measurements on non-rotating parts—Part 4: Gas turbine sets with fluid-film bearings

IEC Standards

ELECTROACOUSTICS (TC 29)

IEC 60318-1 Ed. 2.0 b:2009, Electroacoustics—Simulators of human head and ear—Part 1: Ear simulator for the measurement of supra-aural and circumaural earphones

PERFORMANCE OF HOUSEHOLD ELECTRICAL APPLIANCES (TC 59)

IEC 60704-2-3 Ed. 2.0 b:2001, Household and similar electrical appliances—Test code for the determination of airborne acoustical noise—Part 2–3: Particular requirements for dishwashers

SAFETY OF HAND-HELD MOTOR-OPERATED ELECTRIC TOOLS (TC 116)

IEC 60745-2-12 Ed. 2.0 b:2003, Hand-held motor-operated electric tools—Safety—Part 2–12: Particular requirements for concrete vibrators

ISO Draft Standards

MECHANICAL VIBRATION AND SHOCK (TC 108)

ISO 14839-1/DAmd1, Mechanical vibration—Vibration of rotating machinery equipped with active magnetic bearings—Part 1: Vocabulary—Draft Amendment 1—12/26/2009

COMPRESSORS, PNEUMATIC TOOLS AND PNEUMATIC MACHINES (TC 118)

ISO/DIS 28927-4, Hand-held portable power tools—Test methods for evaluation of vibration emission—Part 4: Straight grinders—12/19/2009

ISO/DIS 28927-10, Hand-held portable power tools—Test methods for evaluation of vibration emission—Part 10: Percussive drills, hammers and breakers—12/19/2009

ISO/DIS 28927-11, Hand-held portable power tools—Test methods for evaluation of vibration emission—Part 11: Stone hammers—12/19/2009

TRACTORS AND MACHINERY FOR AGRICULTURE AND FORESTRY (TC 23)

ISO/DIS 22867, Forestry and garden machinery—Vibration test code for portable hand-held machines with internal combustion engine—Vibration at the handles—1/14/2010

IEC Draft Standards

51/970/FDIS, IEC 62333-3 Ed.1: Noise Suppression Sheet for Digital Devices and Equipment—Part 3: Characterization of parameters of noise suppression sheet, 12/11/2009

100/1620/FDIS, IEC 60728-2: Cable networks for television signals, sound signals and interactive services—Part 2: Electromagnetic compatibility for equipment, 11/27/2009

100/1621/FDIS, IEC 60268-7: Sound system equipment—Part 7: Headphones and earphones, 11/27/2009

100/1624/FDIS, IEC 62458: Sound system equipment—Electroacoustical transducers—Measurement of large signal parameters, 12/04/2009

REVIEWS OF ACOUSTICAL PATENTS

Sean A. Fulop
Dept. of Linguistics, PB92
California State University Fresno
5245 N. Backer Ave., Fresno, California 93740

Lloyd Rice
11222 Flatiron Drive, Lafayette, Colorado 80026

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

Reviewers for this issue:

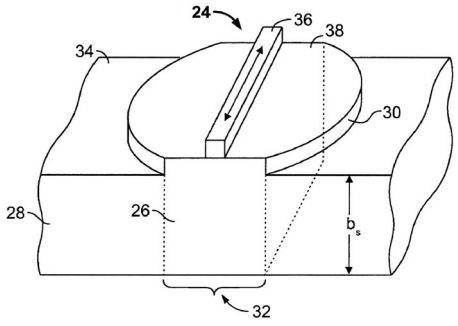
GEORGE L. AUGSPURGER, *Perception, Incorporated, Box 39536, Los Angeles, California 90039*
JEROME A. HELFFRICH, *Southwest Research Institute, San Antonio, Texas 78228*
DAVID PREVES, *Starkey Laboratories, 6600 Washington Ave. S., Eden Prairie, Minnesota 55344*
NEIL A. SHAW, *Menlo Scientific Acoustics, Inc., Post Office Box 1610, Topanga, California 90290*
ERIC E. UNGAR, *Acentech, Incorporated, 33 Moulton Street, Cambridge, Massachusetts 02138*
ROBERT C. WAAG, *Department of Electrical and Computer Engineering, University of Rochester, Rochester, New York 14627*

7,564,167

43.38.Ar SYSTEM AND METHOD OF ASSEMBLING A TRAPPED ACOUSTIC WAVE SYSTEM

Brian J. Truesdale, assignor to Illinois Tool Works Incorporated
21 July 2009 (Class 310/320); filed 7 May 2008

The author discloses the principles behind construction of a touch-sensitive switch that uses only a two-terminal piezoelectric transducer as the sensor element. The details are not very well spelled out—not even the frequency range of intended operation—though one gets the impression that it would be in the Lamb wave range where one wavelength is near the thickness of the substrate pictured. The principle of operation is that one has defined a "resonant mesa" for surface (Lamb) waves in the substrate using a change in thickness (what the author calls a moat), and the trapped energy can be dissipated effectively by placing a finger on top of the mesa. The



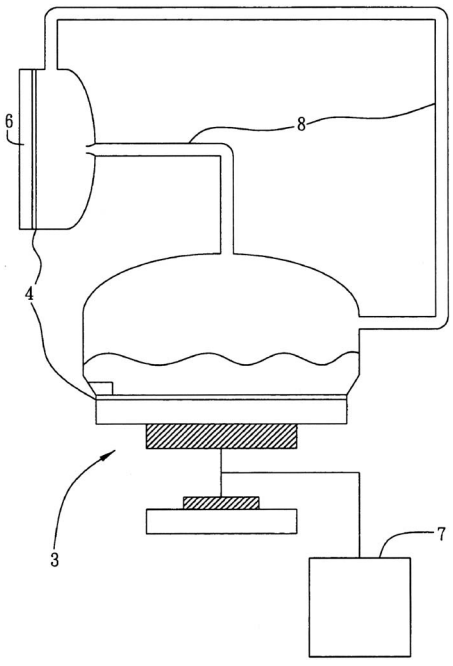
change in two-terminal impedance of the transducer is then detected as a touch. Most of the discussion in this patent is on the construction of and means of bonding to the mesa, the details of operation presumably disclosed in the earlier U.S. Patent No. 7106310. This is certainly a simple and robust switch concept, though probably not widely used in comparison to capacitive switches.—JAH

7,567,015

43.38.Ar COMPOSITE MODE TRANSDUCER AND COOLING DEVICE HAVING THE COMPOSITE MODE TRANSDUCER

Syh-Yuh Cheng *et al.*, assignors to Industrial Technology Research Institute
28 July 2009 (Class 310/321); filed in Taiwan 3 March 2006

This patent discloses the use of vibrating piezoelectric diaphragm-type devices to pump or atomize fluid from a cooler area to a warmer one to facilitate heat exchange. The emphasis seems to be on open-cycle systems where the fluid is atomized from a reservoir and impinges on the hot part, but is not reclaimed. However, in the last configuration shown, the authors allow for the recycling of the fluid via a return path. The system looks very



complicated for actual installation in a computer case, as the authors propose, not to mention the problem of having your computer emit screeching sounds continually.—JAH

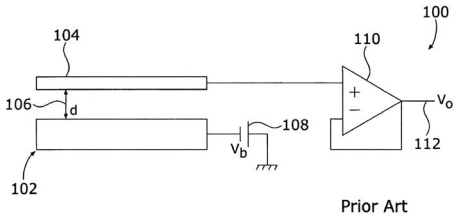
 0001-4966/2010/127(1)/603/9/$25.00

7,545,945

43.38.Bs COMB SENSE MICROPHONE

Ronald Miles, assignor to The Research Foundation of the State University of New York
9 June 2009 (Class 381/174); filed 5 August 2005

A micro-machined capacitor microphone is described where stiff fingers 202, part of diagram 204, are inter-digitated with stiff fingers 206 of fixed backplate 208, which is said to remove the problem of the attraction of diagram 204 to backplate 208, and therefore decouples the sense voltage



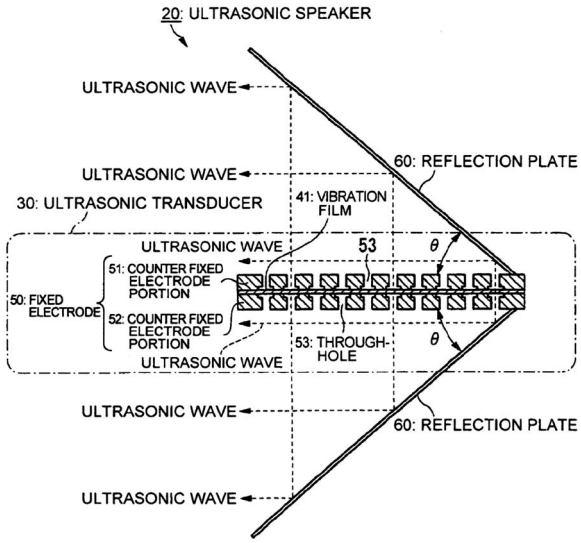
from the bias voltage. Other enhancements regarding the micro-machined suspension are also described in a clear manner.—NAS

7,564,986

43.38.Bs ULTRASONIC SPEAKER AND PROJECTOR

Hirokazu Sekino *et al.*, assignors to Seiko Epson Corporation
21 July 2009 (Class 381/191); filed in Japan 21 June 2004

A conventional push-pull electrostatic loudspeaker 53 is used as an ultrasound generator. A pair of reflectors 60 redirect sound waves from both sides of the diaphragm to form a right-angle beam, as indicated by the dotted arrows. This simple arrangement is essentially what has been patented and is described at greater length in the single independent patent claim. Since the two sound sources operate in phase opposition there can be



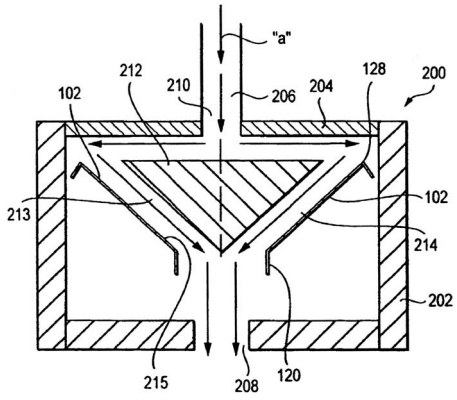
no acoustic output directly on axis, which would seem to be a severe disadvantage. However, a variant includes provisions for spacing one reflector a half-wavelength farther from the diaphragm, thus producing a coherent wavefront. The device is intended to be used as a parametric loudspeaker, producing a concentrated beam of audible sound in thin air.—GLA

7,539,324

43.38.Dv LOUDSPEAKER DIAPHRAGM SYSTEMS

Robert Polfreman and Michael Harris, assignors to Harman International Industries, Incorporated
26 May 2009 (Class 381/423); filed in United Kingdom 8 July 2002

A method of creating an anodized diaphragm/voice coil former in one piece is described, wherein the thickness and tapering of the conical and cylindrical portions of the assembly can be varied in a controlled manner.—NAS

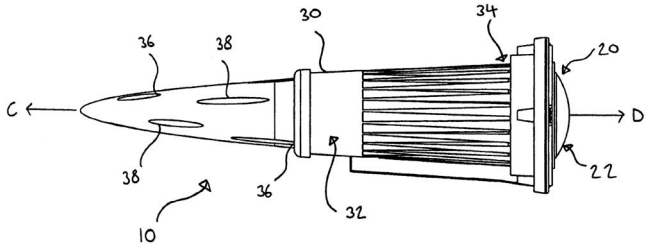


7,536,024

43.38.Ja LOUDSPEAKER

Daniel Bailey and Graeme Foy, assignors to Mordaunt-Short Limited
19 May 2009 (Class 381/338); filed in United Kingdom 17 May 2004

The invention appears to describe a high frequency audio reproducing assembly 10 that can be used alone or as a tweeter in a multi-way loudspeaker system. The design seems to allow the device 10 to also form a visual design element that can be separated from the rest of the multi-way loudspeaker system and so is form as well as function driven. To control the resonant peak that may occur when open-backed drive units are coupled to a short tubular enclosure, the invention incorporates a number of tubes 32 of



different lengths, each with an opening 34 to "encourage sound to be radiated." The tubes are of different lengths and can be arranged "both axially and radially with respect to the central axis."—NAS

7,542,583

43.38.Ja LOUDSPEAKER

Osamu Funahashi, assignor to Panasonic Corporation
2 June 2009 (Class 381/398); filed in Japan 8 March 2004

"A loudspeaker has a configuration in which edge diameter in the cross section of second edge coupled to suspension holder is set to be larger than edge diameter in the cross section of first edge coupled to diaphragm. With such a configuration, a loudspeaker with reduced harmonic distortion is provided." Or, if the radius of surround 14 is greater by about 50% than that of surround 15, the harmonic distortion of the diaphragm and suspension assembly is reduced.—NAS

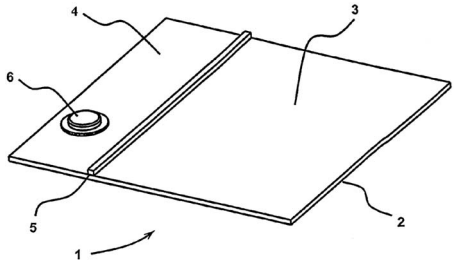
7,570,771

43.38.Ja LOUDSPEAKERS

Timothy Christopher Whitwell *et al.*, assignors to New Transducers Limited

4 August 2009 (Class 381/152); filed in United Kingdom 11 March 2004

The illustration shows a rear view of this bending wave loudspeaker. An impedance divider 5 separates the diaphragm into two areas 3 and 4. Area 3 is transparent and is intended to cover a visual display screen, with a separation of only 1 or 2 mm between the two. It functions mainly as a high



frequency reproducer. Area 4 has a larger back cavity so that it can reproduce lower frequencies. Thus, the transducer operates as a full-range loudspeaker, yet without optically distorting the video display.—GLA

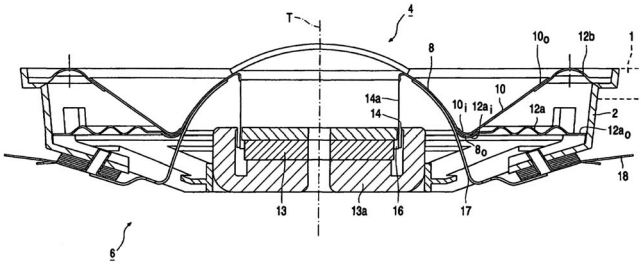
7,570,780

43.38.Ja LOUDSPEAKER HAVING A COMPOSITE DIAPHRAGM STRUCTURE

Jozef Arnold Frans Baeten, assignor to PSS Belgium N.V.

4 August 2009 (Class 381/423); filed in the European Patent Office 22 August 2003

The illustration is misleading; a loudspeaker constructed as shown fails to exploit this patent's most distinctive feature. However, it is easy to see that voice coil tube 14a could be much shorter, allowing magnetic assembly 13 to move farther up into the dome-shaped cavity behind dia-



phragm 8. The result would indeed be a very shallow moving-coil loudspeaker.—GLA

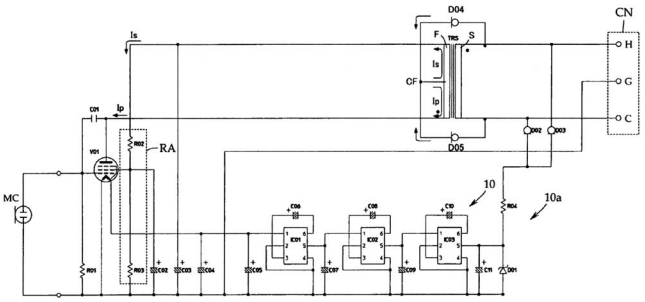
7,542,578

43.38.Kb CONDENSER MICROPHONE

Hiroshi Akino, assignor to Kabushiki Kaisha Audio-Technica

2 June 2009 (Class 381/113); filed in Japan 14 May 2004

Vacuum tubes used as impedance converters for "classic" type condenser microphones require both an "A" voltage for the heater and a "B" voltage for the plate. By using a specially built vacuum tube V01 and a step-down converter circuit 10a the A voltage can be supplied by a 48 V



phantom power supply sent over terminals H and C with return via terminal G, as can the B voltage. However, the patent asserts that current designs can suffer from dc saturation of transformer TRS from the B voltage since this voltage appears across the transformer primary winding, which prior art solves by using a bigger and heavier transformer. This design uses a center tap on the primary side of the transformer and two constant-current diodes D04 and D05 so that two currents Is (screen) and Ip (plate) are developed that "prevent" dc saturation of the transformer core. If these currents are within 5% of one another, the patent states that a smaller in both volume (by about 1/7) and weight (by about 1/8), transformer can then be used.—NAS

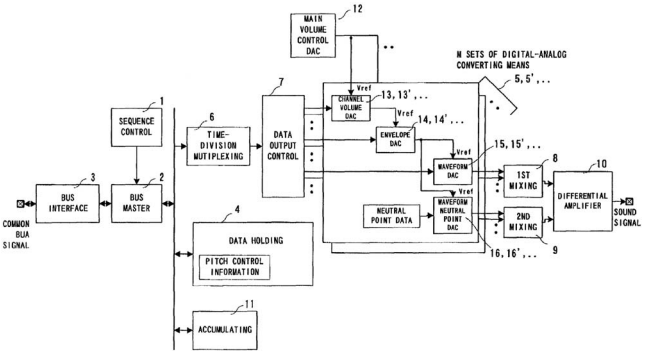
7,561,931

43.38.Lc SOUND PROCESSOR

Shuhei Kato and Koichi Sano, assignors to SSD Company Limited

14 July 2009 (Class 700/94); filed 10 August 2000

Computer games generate and modify a variety of sound effects in response to actions of the player. Moreover, the effects may be generated in two-channel stereo or surround sound. According to this patent, the interface and data storage systems used in typical sound modules can place severe demands on the host computer. "Therefore, it is a primary object of the



present invention to provide a sound processor and sound processing apparatus which is free from functional restriction due to the capacity of a local memory with less processing loading on the host-side units, such as a CPU, but capable of achieving simultaneous reproduction for a multiplicity of channels at low cost." A generalized single-chip circuit and embedded program are disclosed.—GLA

7,564,968

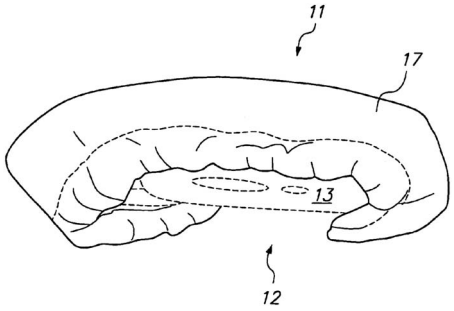
43.38.Si VOICE CONTAINMENT DEVICE FOR A TELEPHONE

Ronald Lee Ayers and Ada Marie Bull, both of Sunnyvale, California

21 July 2009 (Class 379/440); filed 20 January 2005

This invention is a kind of acoustic muff that slips over a cellular phone. Its purpose is to confine the voice of the user to the microphone, thus providing privacy and improving voice pickup. Does it work? Can it work?

The patent gives no indication that the device has actually been tested.—GLA

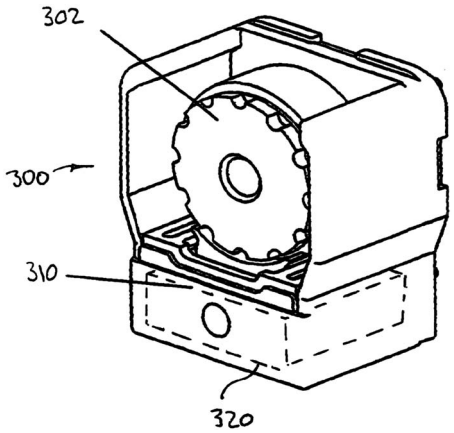


7,565,070

43.38.Si TELEPHONE VIBRATOR

Roman C. Gutierrez, assignor to Siimpel Corporation
21 July 2009 (Class 396/75); filed 28 March 2006

A multi-purpose cellular phone may include a camera. The camera includes a miniature motor to operate its zoom lens. If the lens is moved



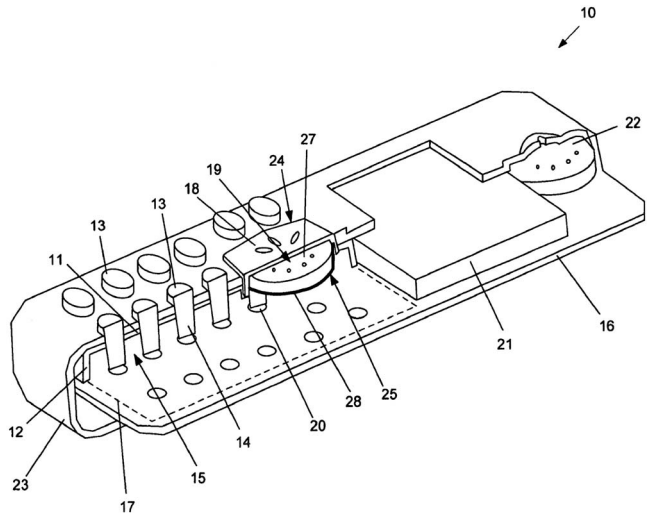
rapidly back and forth it will provide tactile vibration, eliminating the need for a separate ring-alert vibrator.—GLA

7,565,178

43.38.Si PORTABLE HANDSET WITH INTEGRATED SPEAKER

Sidney Sitachitt *et al.*, assignors to Kyocera Wireless Corporation
21 July 2009 (Class 455/569.1); filed 16 August 2005

A wireless handset 10 has a numeric keypad that includes a large navigation key 24. The navigation key is perforated 18 to allow the passage of sound from speaker 19. The geometry described in the patent creates a



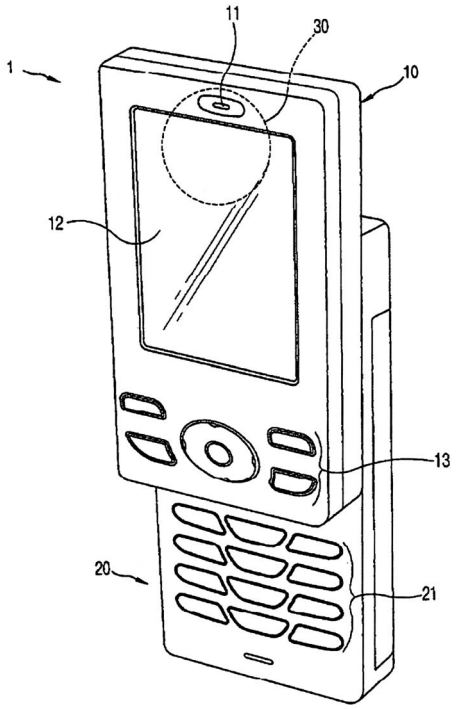
small sealed rear cavity that is said to improve the quality of sound projected from the front of the speaker.—GLA

7,565,184

43.38.Si MOBILE COMMUNICATION TERMINAL

Seong-Cheol Lee and Yong-Hum Lee, assignors to LG Electronics Incorporated
21 July 2009 (Class 455/575.1); filed in Republic of Korea 24 May 2005

This patent describes yet another attempt to house a relatively large loudspeaker in a very small cellular telephone. In this case, high-frequency sound from the front of ceramic disk speaker 30 is confined to a small chamber and emanates from opening 11. "Low or middle pitched sound is



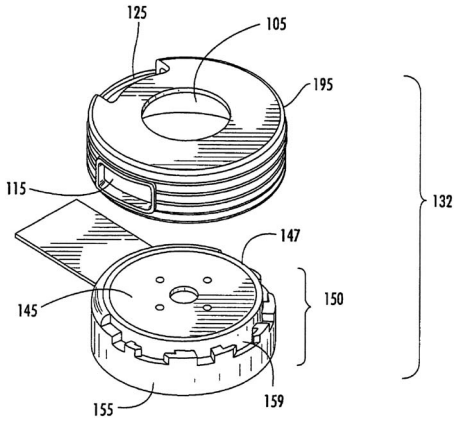
emitted from the rear surface of the speaker 30 and resonated in the first body 10." None of which makes much sense acoustically, but the geometry is clearly described in the patent.—GLA

7,567,680

43.38.Si DUAL-DIAPHRAGM SPEAKER ASSEMBLIES WITH ACOUSTIC PASSAGEWAYS AND MOBILE TERMINALS INCLUDING THE SAME

Matthew Justin Murray *et al.*, assignors to Sony Ericsson Mobile Communications, AB
28 July 2009 (Class 381/181); filed 29 October 2004

This patent is difficult to follow. Components are numbered differently in different diagrams, and the numbers do not always fit the descriptions of individual components. The device shown in the exploded view contains two microspeakers. The upper speaker serves as a handset receiver, projecting sound from the front of its diaphragm through opening 105. The lower



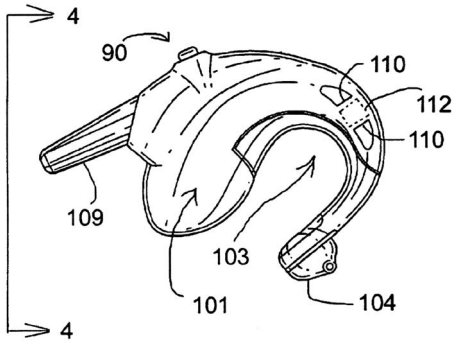
speaker is intended for hands-free communication or music reproduction. Its output is intended to emanate from side opening 115. An important feature of the invention is "grommet" 195, which fits down over housing 155 and maintains at least a partial seal between the two sound passages so that high acoustic levels are not ported too close to a user's ear.—GLA

7,570,777

43.38.Si EARSET ASSEMBLY

Jon C. Taenzer *et al.*, assignors to Step Labs, Incorporated
4 August 2009 (Class 381/381); filed 13 January 2004

This earpiece is intended to "…provide great comfort in the use of an earset assembly." To that end a rotatable ear bud 104 allows the assembly to fit almost any person's right or left ear, the connecting wire is brought down behind the ear through a flexible tube 109, and the portion of the ear bud that fits into the ear canal is made non-occluding.—GLA

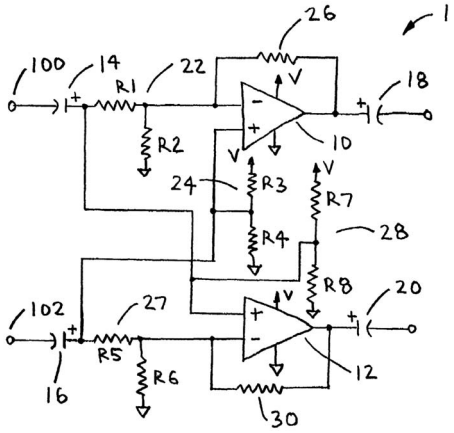


7,564,982

43.38.Vk TWO CHANNEL AUDIO SURROUND SOUND CIRCUIT

Paul R. Ambourn, assignor to Phantom Technologies, Incorporated
21 July 2009 (Class 381/120); filed 25 June 2002

This is a strange patent. The circuit shown is intended to create a surround-sound effect in a two-channel stereo system by enhancing the difference signal component. The scheme is common prior art found in many small stereo systems. However, this patent points out that an earlier circuit (Bruney, 1985) used 1 μF input and output capacitors, "…which act as low pass filters and provide poor frequency response in lower frequencies in the symmetrical circuit." The solution is to replace the 1 μF capacitors with much larger polarized electrolytic capacitors. Is this modification conceivably novel, non-obvious, and deserving of 30 patent claims? The mind boggles.—GLA



7,561,976

43.40.Le METHOD AND SYSTEM FOR MONITORING THE PERFORMANCE OF A PIPE CONTAINING A PRESSURIZED FLUID

Olivier Bernard and Bruno Gerard, assignors to Oxand
14 July 2009 (Class 702/42); filed in France 11 March 2003

Monitoring of aging pipes is desirable for determining their expected useful lives and scheduling of repairs. Monitoring of pipes, particularly of pipes made of pre-stressed or reinforced concrete, according to this patent is accomplished by a combination of static and dynamic measurements. Static measurements are used to determine a pipe's circumferential expansion under pressure, providing information on longitudinal cracks and the pipe's effective stiffness. Measurement of the pipe's modes and natural frequencies adds further information on the pipe's stiffness and on its residual cross-section. The resulting data are combined in a predictive model to predict the pipe's remaining service life.—EEU

7,565,957

43.40.Tm DAMPER AND VIBRATION DAMPING STRUCTURE USING THE SAME

Masami Mochimaru *et al.*, assignors to Oiles Corporation
28 July 2009 (Class 188/297); filed in Japan 21 February 2002

The dampers described in this patent are intended for the reduction of the vibrations of buildings due to earthquakes. The dampers in essence consist of nests of tubes in which the space between the tubes is filled with a viscous fluid or with a viscoeleastic material. The tubes are attached to end-pieces so that axial relative motion of the end pieces causes adjacent

tubes to move relative to each other, inducing shear in the damping materials. The damping assemblies are intended to be mounted diagonally in frames of the building or vertically in parallel with the building columns.—EEU

7,562,904

43.40.Vn VIBRATION CONTROL DEVICE PROVIDING A DESIRED VIBRATION CHARACTER IN A STEERING WHEEL OF A VEHICLE

Mats Gustavsson, assignor to A2 Acoustic AB
21 July 2009 (Class 280/779); filed in Sweden 1 April 2004

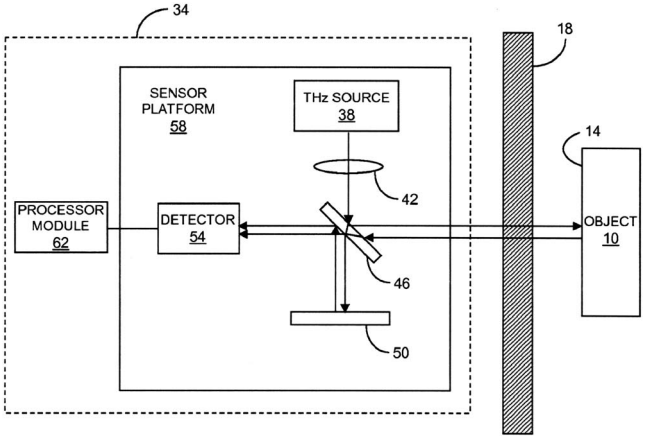
Active control of the vibrations of steering wheels is designed so that undesirable vibration components (such as those produced by the engine) are attenuated while others (such as those that provide "road feel") are not reduced or even amplified. The outputs of sensors mounted on the vehicle frame relate to the road conditions and are used to determine which components should not be attenuated by the active system that controls the vibrations to which the steering wheel is subject.—EEU

7,564,567

43.40.Yq SENSOR FOR MEASURING A VIBRATING SURFACE OBSCURED FROM VIEW

Jerry C. Chen and Sumanth Kaushik, assignors to Massachusetts Institute of Technology
21 July 2009 (Class 356/502); filed 25 January 2007

The vibration measurement system described in this patent is basically similar to a laser vibrometer. However, instead of light, it uses terahertz electromagnetic radiation, which can penetrate optically opaque objects. In one embodiment including a Michelson interferometer arrangement that employs a narrow-band source 38 of a terahertz beam, a beamforming optical module (such as a polyethylene lens) 42, beamsplitter 46, mirror 50, and



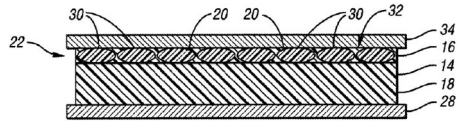
detector 54, all components are mounted rigidly on a stiff platform 58 that has small coefficient of thermal expansion. The processor 62 extracts the vibration signal from the detector output and also may be employed to move the reference mirror 50 to maintain maximum linearity. The source may radiate at frequencies between 300 GHz and 3 THz (with a wavelength between 1 mm and 100 μm), but needs to have a bandwidth of tens of megahertz or less.—EEU

7,566,475

43.55.Ti ACOUSTIC INSULATOR WITH CONTROLLED AIRFLOW RESISTANCE AND METHOD OF MAKING SAME

Terence Connelly *et al.*, assignors to International Automotive Components Group North America, Incorporated
28 July 2009 (Class 427/195); filed 9 November 2004

The authors disclose the fabrication of an acoustic insulator wall material that includes a porous first surface 26 that is manufactured in a novel way. The surface covering (unspecified, but probably a fabric) is bonded to the substrate with thermoplastic material that diffuses into the covering un-



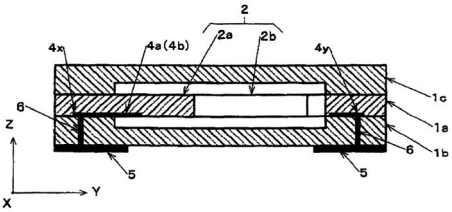
der heat and pressure as a processing step. The resulting sandwich has desirable semi-porous properties, along with an areal mass density of 0.1–1.0 kg/m^2. The idea seems to be sound if the manufacturing is not too costly.—JAH

7,564,177

43.58.Hp CRYSTAL UNIT HAVING STACKED STRUCTURE

Masahiro Yoshimatsu *et al.*, assignors to Nihon Dempa Kogyo Company, Limited
21 July 2009 (Class 310/365); filed in Japan 26 December 2006

This patent is interesting for the glimpse it gives to the future of quartz clock crystal fabrication. The authors disclose a new technique for bonding quartz tuning fork type resonators into a quartz case, using adhesion-promoting chemical treatment with siloxane to promote a direct quartz to



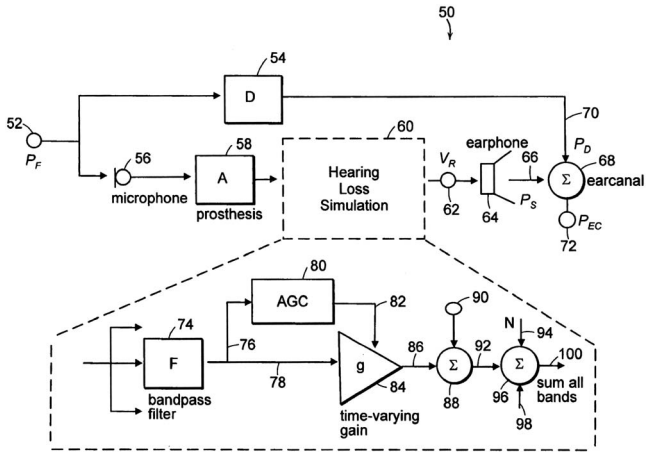
quartz bond. In this way, the package can be made thinner and the problems of dissimilar thermal expansion coefficients can be alleviated. There would also be simplification of the fabrication process since it does not have to involve as many dissimilar materials (glass and low-temperature solder) on the assembly line.—JAH

7,564,980

43.66.Ba SYSTEM AND METHOD FOR IMMERSIVE SIMULATION OF HEARING LOSS AND AUDITORY PROSTHESES

Patrick M. Zurek and Joseph G. Desloge, assignors to Sensimetrics Corporation
21 July 2009 (Class 381/60); filed 21 April 2005

To demonstrate the effects of hearing loss as well as improvements provided by hearing aids, a hearing loss simulator, which may be head-worn and binaural, provides automatic gain control (AGC) on acoustic input signals with variable and fixed gains for inputs exceeding first and second thresholds, respectively, and then adds a controlled amount of noise to the processed signals so as to make input signals inaudible when they are below another input threshold level. AGC processing and noise addition may be performed in multiple frequency bands.—DAP

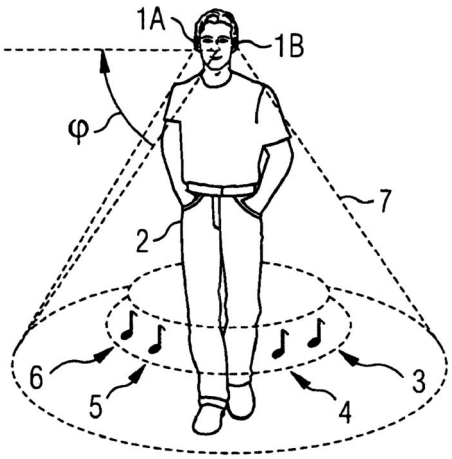


7,561,707

43.66.Ts HEARING AID SYSTEM

Ulrich Kornagel, assignor to Siemens Audiologische Technik GmbH
14 July 2009 (Class 381/310); filed in Germany 20 July 2004

Status signals having varying virtual spatial locations within a rotationally symmetric cone surrounding the wearer of binaural hearing aids are created by adjusting the relative amplitude and relative phase of the audio signals presented to the two ears, factoring in a head related transfer function. The different locations of the virtual signal sources are said to make more easily accessible information for the wearer about system states or hearing aid settings, such as battery life remaining or active program number.—DAP

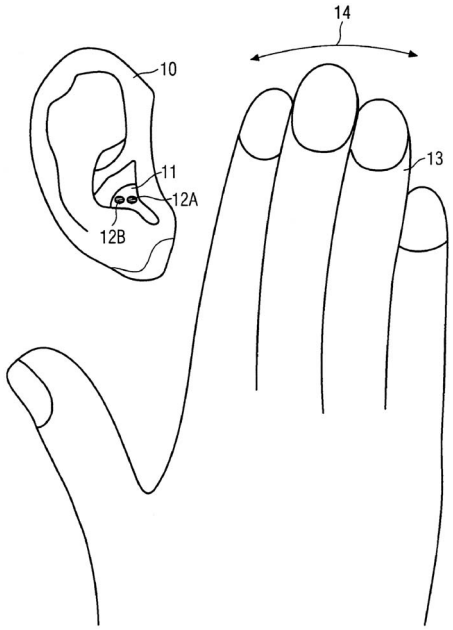


7,561,708

43.66.Ts HEARING AID

Gerhard Röhrlein, assignor to Siemens Audiologische Technik GmbH
14 July 2009 (Class 381/315); filed in Germany 21 April 2004

To eliminate the need for a remote control device, a magnetic activator, or other external devices, a proximity sensor on the hearing aid, which could be infrared or ultrasonic, detects when the wearer's hand is moved into or out of an area near the ear and triggers an operating function of the hearing aid accordingly. Usefulness and detection accuracy are said to be improved by causing several different operating functions, such as a volume adjustment or changing which signal processing program is active, to be triggered by different chronological sequences of multiple entries or exits of the hand into and out of the detection area, respectively.—DAP

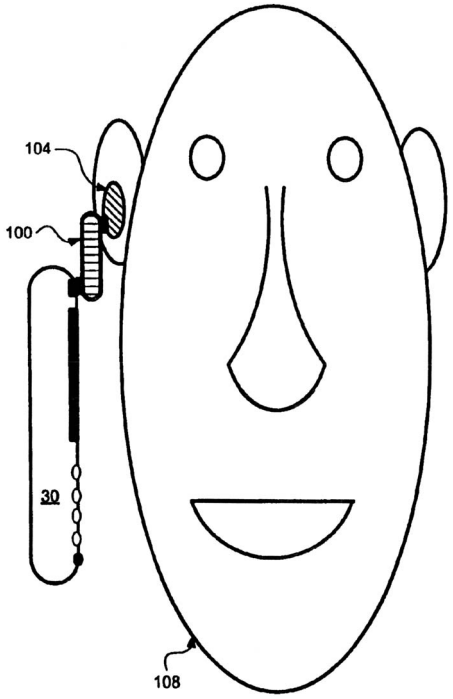


7,561,711

43.66.Ts AUDIO PERIPHERAL FOR AN ELECTRONIC DEVICE

Larry Hawker *et al.*, assignors to Research In Motion Limited
14 July 2009 (Class 381/331); filed 20 September 2005

A peripheral device containing input and output telecoils spaced by 2–6 cm and an optional amplifier, frequency response shaping, and noise-filtering circuitry in between is placed between a wireless device and a



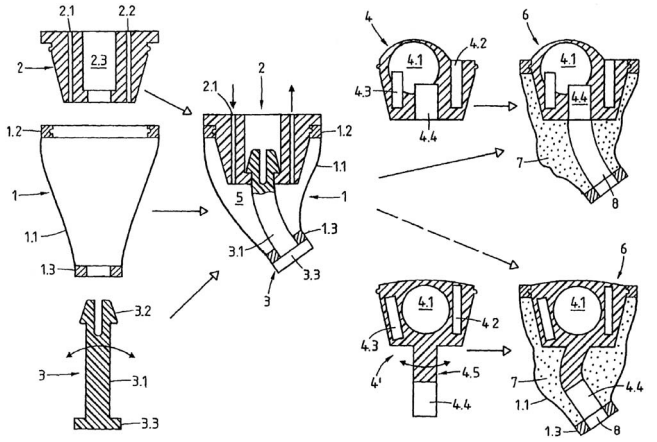
hearing aid to help reduce electromagnetic interference in hearing aids produced by devices such as cellular telephones. The peripheral device may be removably attached to the wireless device.—DAP

7,564,987

43.66.Ts SET AND METHOD FOR PRODUCING A HEARING AID AND HEARING AID PRODUCED ACCORDING TO THIS METHOD

Marcel Aeschlimann *et al.*, assignors to Bernafon AG
21 July 2009 (Class 381/322); filed in Switzerland 13 November 2002

A casting cavity for an instant-fit custom hearing aid is filled while it is in place in the ear canal of the wearer. The casting cavity is created from a central casting space element comprising outer and inner rings with a tubular expandable membrane extending between them, an outer casting space element with a casting opening matched to the outer ring, and a partially



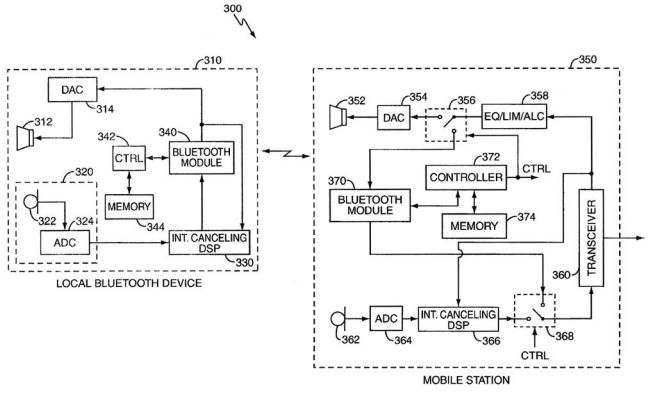
flexible canal element connected to the inner ring and outer casting element. During casting, the electronics module and loudspeaker, which replace the outer casting element, are allowed to float to adapt to the shape of the ear canal.—DAP

7,570,974

43.66.Ts BLUETOOTH^{R} ENABLED HEARING AID

Matt Andrew Kaltenbach *et al.*, assignors to Sony Ericsson Mobile Communications AB
4 August 2009 (Class 455/569.1); filed 29 May 2007

To help reduce electromagnetic interference in hearing aids produced by close-proximity mobile stations such as cellular telephones, a short-range wireless communication link, such as bluetooth, communicates between the device providing hearing aid functionality, such as a wireless headset, and



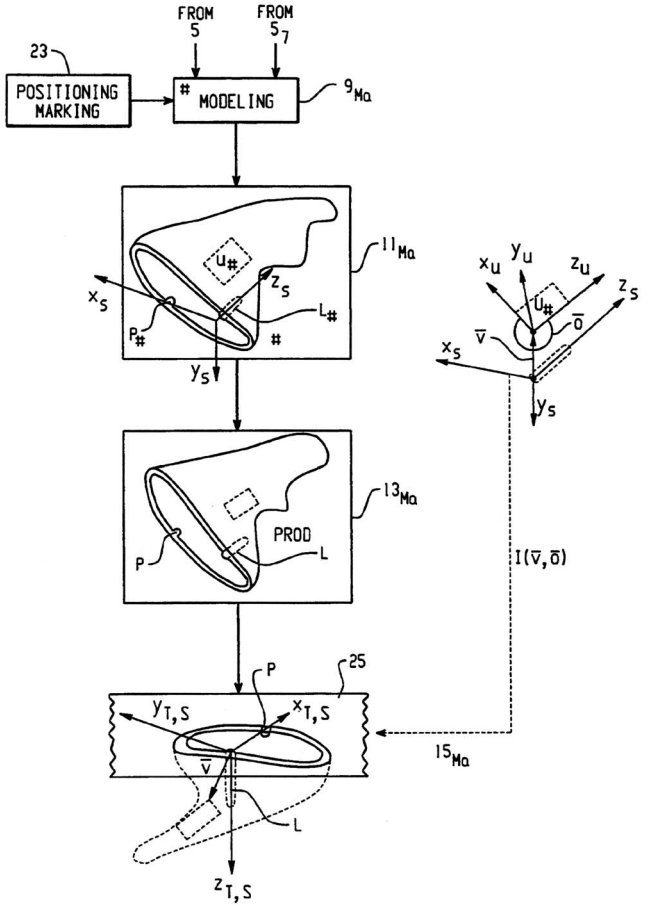
the mobile station. Interference suppression circuitry, whose properties are determined in concert with the mobile station, is implemented in the device providing hearing aid functionality between a microphone or microphones and the short-range transceiver. Interference suppression may include noise suppression and acoustic echo suppression.—DAP

7,571,018

43.66.Ts PRESERVING LOCALIZATION INFORMATION FROM MODELING TO ASSEMBLING

Samuel Hans Martin Roth *et al.*, assignors to Phonak AG
4 August 2009 (Class 700/98); filed 14 June 2006

The three-dimensional orientation system, assigned during formulation of the computer model of the impression/shell for a custom hearing aid, is preserved by adding embossments or projection markings on the shell during manufacturing to help define shell modifications needed for good component fit.—DAP

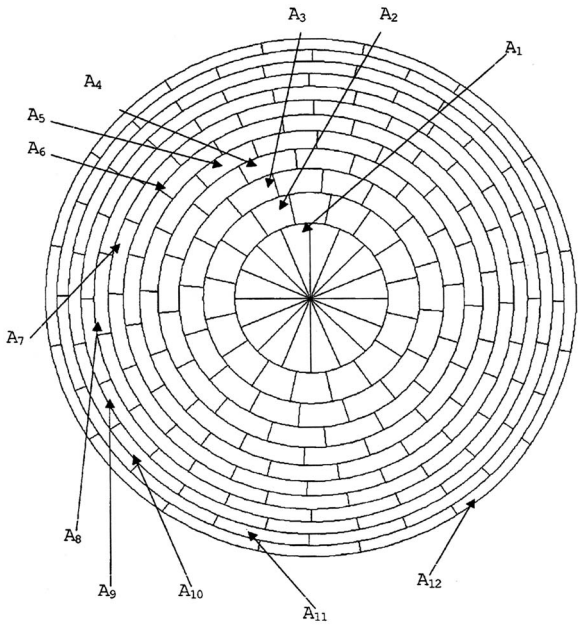


7,559,897

43.80.Vj ELECTRONIC ARRAY PROBE FOR ULTRASONIC IMAGING

Marino Cerofolini, assignor to Esaote, S.p.a.
14 July 2009 (Class 600/459); filed in the European Patent Office 1 July 2003

This probe contains an annular array of transducer elements. The array surface is shaped like the cap of a sphere to minimize the range of electronic focusing delays needed to position the focus at different axial locations. The transducer elements are staggered from annulus to annulus, and the width of each annulus decreases as the radius of the annulus increases.—RCW



A_i with i ∈ {2, 4, 6, 8, 10, 12, .., n even}
Annular bands with position of even order number

A_i with i ∈ {1, 3, 5, 7, 11, .., n odd}
Annular bands with position of odd order number

7,563,228

43.80.Vj STEREOSCOPIC THREE OR FOUR DIMENSIONAL ULTRASOUND IMAGING

Qinglin Ma *et al.*, assignors to Siemens Medical Solutions USA, Incorporated
21 July 2009 (Class 600/437); filed 24 January 2005

A video display system is integrated in an ultrasound imaging instrument. The imaging system includes a three-dimensional image processor that renders images obtained from different viewing angles. The renderings are displayed either simultaneously or sequentially at a high rate.—RCW

7,569,015

43.80.Vj INTEGRATED PHYSIOLOGY AND IMAGING WORKSTATION

Brenda Donaldson *et al.*, assignors to General Electric Company
4 August 2009 (Class 600/437); filed 15 July 2005

Physiologic signals derived from a subject and ultrasound data obtained from a region of interest in the subject are sent to a workstation. An ultrasound processing unit receives and processes the ultrasound data to produce images. The physiologic signals are combined with ultrasound images and displayed simultaneously with the images in real time during a patient examination.—RCW