Effect of source spectrum on sound localization in an everyday reverberant room

Antje Ihlefeld and Barbara G. Shinn-Cunningham[a)]
Hearing Research Center, Boston University, Boston, Massachusetts 02215

(Received 28 February 2010; revised 10 May 2010; accepted 12 May 2011)

Two experiments explored how frequency content impacts sound localization for sounds containing reverberant energy. Virtual sound sources from thirteen lateral angles and four distances were simulated in the frontal horizontal plane using binaural room impulse responses measured in an everyday office. Experiment 1 compared localization judgments for one-octave-wide noise centered at either 750 Hz (low) or 6000 Hz (high). For both band-limited noises, perceived lateral angle varied monotonically with source angle. For frontal sources, perceived locations were similar for low- and high-frequency noise; however, for lateral sources, localization was less accurate for low-frequency noise than for high-frequency noise. With increasing source distance, judgments of both noises became more biased toward the median plane, an effect that was greater for low-frequency noise than for high-frequency noise. In Experiment 2, simultaneous presentation of low- and high-frequency noises yielded performance that was less accurate than that for high-frequency noise, but equal to or better than for low-frequency noise. Results suggest that listeners perceptually weight low-frequency information heavily, even in reverberant conditions where high-frequency stimuli are localized more accurately. These findings show that listeners do not always optimally adjust how localization cues are integrated over frequency in reverberant settings. © *2011 Acoustical Society of America*. [DOI: 10.1121/1.3596476]

PACS number(s): 43.66.Qp, 43.66.Pn, 43.66.Yw [MAA] Pages: 324–333

I. INTRODUCTION

Most models of sound localization are based on human localization data gathered in anechoic space. In anechoic settings, interaural time differences (ITDs) and interaural level differences (ILDs) of broadband sources vary reliably with source lateral angle, and, for nearby sources, with source distance (Brungart *et al.*, 1999). Thus, interaural acoustic cues alone should enable listeners to determine source location to within a cone or torus of confusion (e.g., see Shinn-Cunningham *et al.*, 2000). When source bandwidth and/or spectral density decrease, source azimuth judgments become less accurate and less precise, showing that listeners normally integrate binaural cues across frequency when estimating lateral source angle (Jeffress, 1972; Hartmann, 1983; Trahiotis and Stern, 1989). Although both ITDs and ILDs provide reliable, detectable cues for source azimuth, in anechoic settings listeners tend to weight information from low frequencies (where ITDs are the dominant cue) more than high-frequency ILD cues (Strutt, 1907; McFadden and Pasanen, 1976; Wightman and Kistler, 1992; MacPherson and Middlebrooks, 2002).

Spatial acoustic cues, including those provided by ITDs and ILDs, are altered by reverberation in everyday rooms, where sound travels directly to the ears from the sound source as well as indirectly via reflections from surfaces in the room. While there are some studies exploring localization in reverberant settings (Hartmann, 1983; Rakerd and Hartmann 1985, 1986, 2004, 2010; Giguère and Abel, 1993; Devore *et al.*, 2009; Devore and Delgutte, 2010), we still have an incomplete understanding of how listeners localize when reverberant energy transforms the spatial cues that they hear. The aim of the current study was to examine the relative perceptual weight that listeners give to high- and low-frequency spatial cues when localizing noise in the presence of reverberant energy.

For noise presented in reverberant settings, spatial cues are distorted by the reflected energy reaching the ears, which causes random fluctuations in the short-term values of spatial cues from one instant to the next (Hartmann, 1983; Rakerd and Hartmann 1985, 1986, 2004, 2010; Giguère and Abel, 1993; Devore *et al.*, 2009). These effects increase as the ratio of direct to reverberant energy (D/R) decreases (Shinn-Cunningham and Kawakyu, 2003; Shinn-Cunningham *et al.*, 2005a, 2005b). For ITDs, these moment-to-moment variations generally do not alter the mean ITD, but they increase ITD variability, decreasing interaural coherence and decreasing the reliability of ITD spatial information (Rakerd and Hartmann, 2010; Shinn-Cunningham and Kawakyu, 2003). Indeed, consistent with these acoustic effects, ITD cues may provide no useful azimuth information to a listener when the D/R is too small (Rakerd and Hartmann, 1985, 2004).

On average, reverberant energy adds approximately equal amounts of energy to the total sound reaching each ear. Because, on a decibel scale, addition of the same amount of reverberant energy has a relatively larger effect on the total signal level of the ear receiving less direct-sound energy than the ear receiving more direct-sound energy, reverberant energy tends to reduce the magnitude of ILDs. Thus, as D/R decreases, ILDs become both less reliable and more biased toward zero (Shinn-Cunningham *et al.*, 2005a).

[a)] Also at: Department of Cognitive and Neural Systems, Boston University, 677 Beacon St., Room 311, Boston, MA 02215. Author to whom correspondence should be addressed. Electronic mail: shinn@cns.bu.edu

0001-4966/2011/130(1)/324/10/$30.00

Perceptually, reverberant energy can degrade the accuracy with which listeners can identify the direction of a source (Hartmann, 1983; Giguère and Abel, 1993; Rakerd and Hartmann, 1985, 2004; Devore *et al.*, 2009). In reverberant space, listeners consistently misjudge the direction of lateral, distant sources, perceiving them as closer to midline than their true location (Shinn-Cunningham *et al.*, 2005b; Devore *et al.*, 2009). These perceptual results are consistent with recent physiological recordings from the inferior colliculus of cats showing that reverberant sound degrades the neural representation of ITD (Devore *et al.*, 2009).

Because reverberant energy affects ITDs and ILDs differently, the perceptual weight that listeners give to high- and low-frequency spatial cues may depend on the environment. For instance, in conditions with low D/Rs, where interaural coherence may be so low to render ITDs unusable, listeners may give greater perceptual weight to high-frequency ILDs than in anechoic space. Of course, ILDs are also degraded by reverberant energy (Rakerd and Hartmann, 2010); however, given that ILDs become increasingly biased toward zero as reverberant energy increases, listeners may continue to rely on ITD cues in reverberant space since their values are not systematically biased by reverberant energy.

One way to explore whether listeners adjust how they compute location in different environments is to compare localization in conditions in which the level of reflected energy is manipulated, comparing (1) narrowband, low-frequency stimuli, (2) narrowband, high-frequency stimuli, and (3) stimuli containing both low and high frequencies. Results for low- and high-frequency bands will reveal the ability of listeners to extract useful localization information from the cues in the different frequencies, and how this ability changes as the D/R decreases and both ITDs and ILDs become less reliable. Moreover, since the dominant lateralization cues differ in low frequencies (where ITDs are most important) and high frequencies (where ILDs are more important; see Henning, 1974; Wightman and Kistler, 1992; Macpherson and Middlebrooks, 2002), differences in performance for low-frequency and high-frequency stimuli can give insight into differences in how reverberant energy alters the perception of ITD and ILD information. Comparison of localization judgments for narrowband stimuli and stimuli containing both low-frequency and high-frequency content can reveal how listeners combine localization cues in different settings: if listeners optimally adjust localization computations based on the spatial information available in low and high frequencies, then localization accuracy of stimuli containing both low- and high-frequency energy should be equal to or better than localization of the constituent bands presented alone.

Here, two experiments were conducted to determine how spectral content influences the ability to localize a sound source over a range of D/Rs (controlled by varying the simulated source distance). The first experiment examined how listeners localized narrowband bursts of pink noise centered at either 750 Hz (where ITDs are the dominant perceptual cue) or 6 kHz (where ILDs dominate). The second experiment examined performance when listeners localized sounds comprised of the low- and high-frequency centered bursts presented together, asking whether listeners can combine localization information from low-frequency and high-frequency components optimally to achieve more accurate localization.

II. METHODS

A. Listeners

Seven college students were paid for their participation in the experiments. All listeners had hearing thresholds within 20 dB of normal at octave frequencies between 250 and 8000 Hz, verified by an audiologist. All listeners gave written informed consent to participate in the study, as overseen by the Boston University Charles River Campus Institutional Review Board.

B. Spatial cues

Using a blocked meatus technique, non-individualized, binaural room impulse responses (BRIRs) were measured at the entrances to the ear canals of a Knowles Electronic Manikin for Acoustic Research (KEMAR) (for details about measurement procedures see Shinn-Cunningham *et al.*, 2005a). Briefly, KEMAR was placed on a wooden chair in a small rectangular office [room dimensions 3.3 m (width) × 5.8 m (length) × 2.6 m (height)] with its ears approximately 1.5 m above the ground and its back 0.5 m from the nearest long wall. The room had a thin carpet on the floor and a suspended ceiling composed of acoustically treated tiles. Several pieces of office furniture were in the room, including a whiteboard on the wall nearest to KEMAR, causing some modest asymmetries in the reverberant energy for sources in the left and right hemifield. Reverberation times (T60), estimated from the left ear BRIR recordings for a source at 1 m distance and 0° azimuth, were 490, 418, 487, 578, and 557 ms at 0.5, 1, 2, 4, and 8 kHz, respectively (Schroeder, 1965). Broadband D/Rs, defined as the ratio of the energy of the first 10 ms of each BRIR to the energy in the remainder,[1] were 24, 18, 12, and 6 dB at distances of 15 cm, 40 cm, 1 m, and 1.7 m, respectively.

BRIR measurements were taken with a Bose mini-cube speaker as the source, which was positioned in the horizontal plane containing KEMAR's ears for all combinations of thirteen lateral angles (every 15° from $-90°$ in the left to 90° in the right hemifield) and four distances (0.15, 0.40, 1.00, and 1.70 m from the center of KEMAR's head to the center of the speaker driver). Raw impulse responses were recorded using 65534-point Maximum Length Sequences at a sampling rate of 25 kHz. The raw BRIRs were band-pass filtered between 200 Hz and 12 kHz (to limit the BRIRs to frequencies at which the measurements were reliable; see Shinn-Cunningham *et al.*, 2005a) using zero-phase digital filtering by processing the raw BRIRs with a fifth-order Butterworth band-pass filter, time-reversing the output, processing the time-reversed signal by the same filter, and then reversing this output in time (MATLAB 6.5, The Mathworks, Inc., Natick, MA). To limit the influence of the measurement noise in the long tails of the BRIRs, the BRIRs were time-windowed with an exponentially decaying function whose time constant was set to match that estimated from the room responses. This time-windowing started at the time point at which the amplitude of the recorded signal reached the noise floor (see footnote 3 in Shinn-Cunningham *et al*., 2005a).

In order to provide more quantitative insights into the cues available in the reverberant BRIRs used in the study, we analyzed four spatial acoustic attributes of the signals presented in the experiments. This analysis, which describes how spatial cues in the stimuli vary with simulated source distance and direction, is provided in the Appendix to allow more direct comparison of the current results with those of other experiments investigating the role of reverberant energy on sound localization.

C. Stimuli

Sound localization performance was measured using three types of band-limited pink noise, generated digitally with a sampling frequency of 25 kHz. Stimuli were 250 ms in duration, including 2-ms-long squared cosine ramps at the onset and offset. During training, stimuli consisted of tokens of broadband pink noise, band-pass filtered with finite impulse response filters with 100 dB/octave roll-off with 3-dB down points at 200 Hz and 12 kHz. To create one binaural stimulus, a noise token was processed with a pseudo-anechoic BRIR (producing separate left and right ear signals). The set of pseudo-anechoic BRIRs were generated by time-windowing "Center" BRIRs measured in a previous study, taken with KEMAR in the center of a classroom at a distance of 15 cm from the listener (Shinn-Cunningham *et al*., 2005a).

Experiment 1 separately presented either low-frequency noise (termed "Lo," centered at 750 Hz) or high-frequency noise (termed "Hi," centered at 6 kHz), which were created by band-pass filtering tokens of broadband pink noise. The filters had 3-dB down points at 500 Hz and 1 kHz for the Lo noise and 4 and 8 kHz for the Hi noise (24 dB/octave frequency roll-off). Experiment 2 presented the Lo and Hi noises simultaneously, with the Lo and Hi components gated on and off together. Based on informal listening, the combined Lo+Hi stimulus consistently produced one fused image.

During the training sessions and in both experiments, stimuli were digitally convolved with the appropriate set of BRIRs, then converted by a 16-bit D/A and amplified using Tucker Davis Technology hardware (TDT PD1). Signals were presented via Sennheiser HD 580 headphones at an average level of approximately 65 dB sound pressure level to a listener seated in a sound-treated chamber (IAC). To reduce the usefulness of direct sound level as a cue, noise tokens were first normalized to have the same root mean square (RMS) value, processed with BRIRs, and roved in overall level from trial to trial (+/− 5dB). With this processing, when averaged across all stimuli, the direct-sound energy reaching a listener's ears was somewhat smaller for more distant source than for closer sources. However, these overall level cues varied less systematically with distance than they would have if the source level had been held fixed.

D. Procedures

On each trial, one newly generated token of noise was presented. Listeners were explicitly told that the simulated source distances ranged from 0.15 to 1.7 m and that the source angles ranged from −90° to +90°. However, they were asked to report only the perceived lateral angle of the simulated sources. To indicate perceived lateral angle, listeners used a computer mouse to position a marker on a graphical user interface (GUI). The GUI showed a top-down, cartoon view of the listener with a semicircle marking the range of directions of lateral source angles relative to the listener (extending from −90°, or left, to +90°, or right) and a line extending from the center of the listener's head along the 0° radius on the semicircle. The same GUI was used for all experimental conditions.

Each listener performed a total of 11 one-hour-long sessions: two initial training sessions followed by nine test sessions. In training sessions, listeners were presented with broadband stimuli at a fixed distance of 15 cm using pseudo-anechoic stimuli. Each training session consisted of 12 blocks, each of which contained three repetitions of the thirteen source azimuth angles, in random order. Listeners were instructed that the noises could originate anywhere in the frontal plane. We wished to measure *where* listeners perceived the simulated sources, rather than testing whether they could learn to respond with the "correct," simulated locations. Therefore, to encourage listeners to respond based on the perceived locations, no response feedback was provided during training. At the end of these two training sessions, experimenters verified that each listener could perform this localization task self-consistently. Specifically, using only data from the second training session, for each source angle, the standard deviation across responses was calculated. A listener was judged as providing self-consistent responses if the mean of these standard deviations, averaged across all source angles, was not greater than 20°. All listeners passed this screening.

Data for Experiment 1 were gathered in the six initial test sessions, each of which consisted of 24 experimental blocks. In each block, frequency content (two values), stimulus distance (four values), and stimulus angles (13 values) varied randomly from trial to trial. Each of the 24 blocks contained 26 trials that were randomly ordered, with the constraint that over the course of the six sessions, each subject responded to each combination of 13 azimuths, four distances, and two types of stimuli differing in frequency content (104 conditions) exactly 36 times. As in the training sessions, no feedback was given during any of the test sessions.

Experiment 2 was conducted in the final three test sessions, each of which consisted of 24 blocks of Lo+Hi stimuli, with distances and angles varying randomly from trial to trial. Each of the 24 blocks contained 26 randomly ordered trials, with the constraint that over the course of the three sessions, each subject responded to each combination of 13 azimuths and four distances (52 conditions) exactly 36 times. Again, no feedback was provided during Experiment 2.

E. Statistical analysis

Data were analyzed with repeated measures analysis of variance (ANOVA) using the open-source package CLEAVE (T.J. Herron). For all significant effects and interactions, we computed partial omega square, ω_p^2, the estimated proportion of variability for which that factor accounted.

III. RESULTS

A. Experiment 1

Figure 1 shows the mean lateral response angle as a function of lateral stimulus angle for Lo and Hi noises from the first six experimental test sessions (Experiment 1; the error bars in the figure show one standard deviation of the means across listeners). Panels B–G show results for a pair of sources, one in the left hemifield and one in the corresponding location in the right hemifield (panel A shows results for a source from directly in front). Response patterns were similar across listeners, so only across-listener averages are shown.

Overall, listeners' responses were less accurate for more lateral sources than for sources closer to the median plane (responses tend to fall farther from the gray lines, denoting the simulated source angles, in panels F and G than in panels A–E). For sources at 0° azimuth, responses were generally close to 0° [Fig. 1(A)]. Responses for sources at 15°, 30°, and 45° azimuth were laterally biased [in Figs. 1(B)–1(D), responses fall outside the gray lines]. This lateral bias tended to decrease as distance increased [in Figs. 1(B)–1(D), responses fall closer to the gray line as radial distance increases]. In contrast, responses for the more lateral sources (from 60°, 75°, and 90°) were medially biased [in Figs. 1(E)–1(G), responses fall between the gray lines]; this radial bias increased as distance increased [responses fell farther from the gray line as radial distance increases in Figs. 1(E)–1(G)]. Overall, the medial bias of responses for sources with lateral angles of 60° and beyond was greater in magnitude than the lateral bias of responses for the other sources.[2]

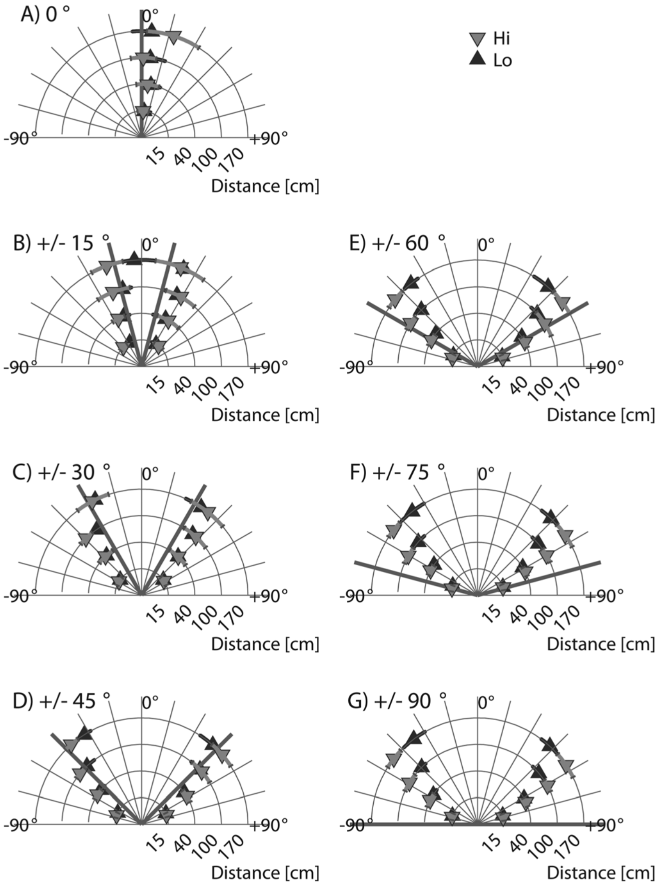

FIG. 1. Experiment 1: Mean localization performance for Lo and Hi noise (black upward and gray downward triangles, respectively). Panels B–G each show one symmetric pair of lateral source angles. Error bars show one standard deviation across listeners.

Performance differed slightly (but consistently) for Lo and Hi noises (compare black and gray triangles within each panel of Fig. 1), an effect that was larger for greater source distances and for greater lateral angles [see Fig. 1(G), largest radial distance]. At 15 cm, localization results were similar for Lo and Hi noises (for the closest simulated distance in each panel, black and gray triangles largely overlap). In contrast, at 1.7 m, Lo noises tended to be perceived as closer to the median plane than Hi noises, especially for sources simulated from more lateral angles [at the greatest radial distance in Figs. 1(D)–1(G), black triangles are closer to midline than gray triangles]. This discrepancy, whereby responses to Lo stimuli were closer to the median plane than responses to Hi stimuli, was as large as 15° for source distances of 40 cm and beyond [compare gray and black triangles in Figs. 1(E)–1(G) at the larger radial distances].

Signed response biases were calculated by subtracting the perceived angle from the source angle. Biases were mirror symmetric for left and right hemifields, so data were collapsed and plotted as a function of the angle from the median plane (Fig. 2); more positive values correspond to responses that are closer to the median plane and more negative values correspond to more lateral judgments.

In general, listeners tended to over-estimate source azimuth for sources near the midline (biases tend to be positive in the left sides of the panels in Fig. 2). This tendency decreased as source distance increased; for instance, there was no bias in the responses for sources near midline for the most distant sources [see left side of Fig. 2(D)]. For sources that were simulated from more lateral locations, listeners tended to underestimate the source lateral angle; this effect increased with increasing source distance [in the right sides

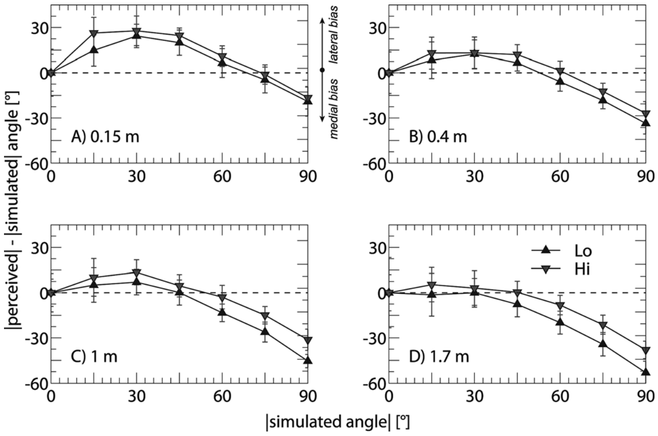

FIG. 2. Experiment 1: Mean signed response biases for Lo and Hi noise (black upward and gray downward triangles, respectively). Panels A–D each show one source distance. Error bars show one standard deviation across listeners.

of the panels in Fig. 2, bias is negative; this negative bias increased systematically from Fig. 2(A) to Fig. 2(D)].

Perceived lateral angle judgments were subjected to a three-factor repeated-measures ANOVA with factors of frequency content, stimulus distance, and lateral angle. This statistical analysis confirmed that all of the trends discussed above were statistically significant. Specifically, we found that the main effects of stimulus distance [$F(3,18) = 3.9$, $p = 0.027$, $\omega_p^2 = 0.99$] and lateral stimulus angle [$F(12,72) = 546.1$, $p < 0.0001$, $\omega_p^2 = 0.99$] were significant. Although the main effect of frequency content was not significant [$F(1,6) = 1.6$, $p = 0.258$], all interactions between the frequency content and the other main variables were [frequency content × distance: $F(3,18) = 1.1$, $\omega_p^2 = 0.01$; frequency content × angle: $F(12,72) = 7.8$, $\omega_p^2 = 0.53$; frequency content × distance × angle: $F(36,216) = 4.9$, $\omega_p^2 = 0.39$; $p < 0.0001$ in all cases]. These significant interactions reflect the fact that lateral position judgments for sources at the sides were less accurate for Lo than Hi noise, and that the size of this difference grew with source distance. *Post hoc* tests with repeated-measures ANOVAs of the difference between perceived lateral angles of Hi and Lo noises, calculated for the two extreme lateral angles (+/−90°), confirmed that distance affected perceived direction significantly and differently for Lo and Hi noises [$F(3,18) = 8.4$, $p = 0.022$ with Greenhouse–Geisser and Bonferroni corrections, $\omega_p^2 = 0.55$, and $F(3,18) = 9.0$, $p = 0.016$ with Greenhouse–Geisser and Bonferroni corrections, $\omega_p^2 = 0.57$, for +90° and −90°, respectively].

The results of Experiment 1 show that in a reverberant room, judgments of source direction change systematically with simulated source distance; moreover, these effects depend on the frequency content of the stimuli. Specifically, we found that perceived lateral source angle was more biased toward the median plane as source distance increased (i.e., with increasing levels of reverberant energy), an effect also reported in past studies (e.g., Shinn-Cunningham *et al.*, 2005b; Devore *et al.*, 2009). For sources within 45° of the median plane, listeners' judgments of source direction were similar for Lo and Hi noises. However, for stimulus angles of 60° and greater, judgments were more biased toward the median plane for low-frequency than for high-frequency sounds, an effect that increased with increasing source distance. These results show that when the D/R was low (for more distant sources), listeners were more accurate at judging the lateral angle of high-frequency sounds than of low-frequency sounds.

B. Experiment 2

Figure 3 displays mean localization judgments for Lo+Hi noise using the same format as Fig. 1. Many of the trends seen in the results from Experiment 1 were also present for the Lo+Hi noise. In particular, judgments for sources from azimuths of 60° or greater often were biased toward the median plane, and this bias tended to increase with increasing source distance and with increasing source lateral angle. Thus, bias was generally largest for sources in panel G (+/−90°) at the greatest radial distance from the origin.

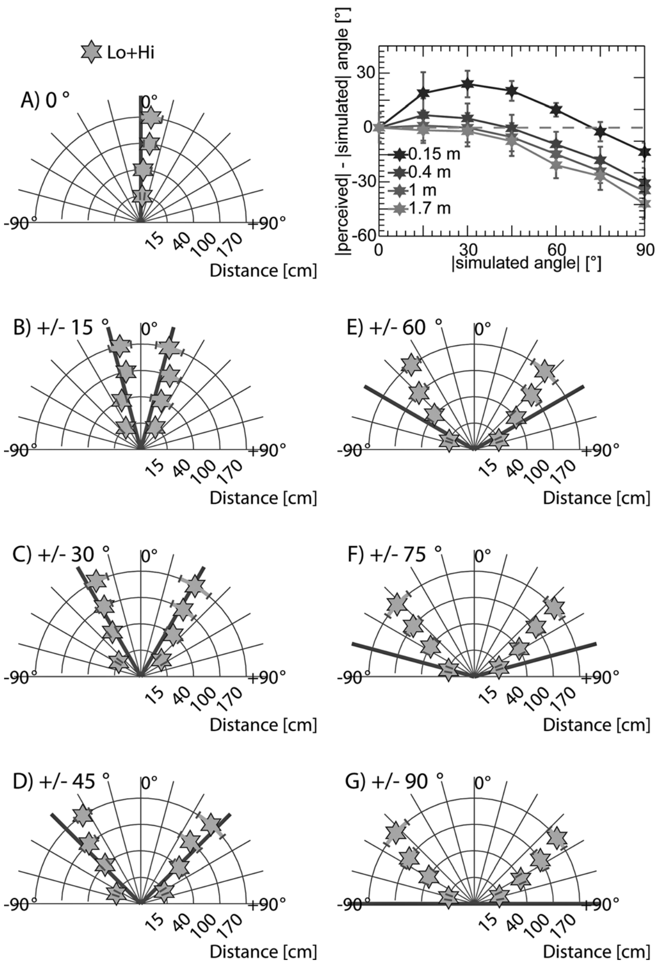

FIG. 3. Experiment 2: Mean localization performance for Lo+Hi noise (light gray hexagrams). Panels B–G each show one symmetric pair of lateral source angles. Error bars show one standard deviation across listeners.

Perceived angles were subjected to a two-factor (distance, lateral angle) repeated-measures ANOVA. Both distance and angle had a significant effect on lateralization [$F(3,18) = 9.7$, $p = 0.001$, $\omega_p^2 = 0.59$, and $F(12,72) = 533.1$, $p < 0.0001$, $\omega_p^2 = 0.99$, for distance and angle, respectively]. The interaction between distance and angle was also significant [$F(36,216) = 36.4$, $p < 0.0001$, $\omega_p^2 = 0.86$], consistent with the fact that with increasing distance, responses for lateral sources were more biased toward midline than for frontal sources, while direction judgments for frontal sources were relatively accurate.

Results for the Lo+Hi noise lend further support to the idea that reverberant energy causes sources from the sides of a listener to be heard closer to midline than their true location. Consequently, as distance increases and D/R decreases, localization accuracy is reduced.

C. Comparison of results from Experiment 1 and Experiment 2

In Experiment 1, judgments for Lo and Hi noises were similar in the least reverberant condition (the 15 cm distance); however, at larger distances, judgments of Hi noises were less biased than judgments of Lo noises. This result implies that in the reverberant space tested, high-frequency localization cues provide more reliable localization information than low-frequency cues. The Lo+Hi stimuli used in Experiment 2 provided listeners with both low- and high-frequency localization cues. To the extent that listeners made optimal use of the information available at low and high frequencies, judgments of the Lo+Hi noise in Experiment 2 should be at least as accurate as both the judgments for Lo noise alone and the judgments for Hi noise alone.

Figure 4 compares localization accuracy in Experiment 1, when only the composite narrow bands were presented (black upward and gray downward triangles for Lo and Hi, respectively), to Lo+Hi noise performance in Experiment 2 (light gray hexagrams). For each listener and simulated distance, the RMS difference between source and perceived lateral angles was computed. These individual values were then averaged across all listeners and all lateral angles to produce the mean RMS error in perceived lateral angle as a function of source distance. Computed in this way, lower RMS values correspond to better performance. If listeners integrated information optimally across frequency, then RMS errors for a Lo+Hi stimulus should be lower than or equal to the smaller of the two RMS errors for the corresponding Lo and Hi single-band stimuli. Therefore, for an optimal listener, RMS errors for Lo+Hi should be equal to or smaller than those of Lo at 15 cm, and equal to or smaller than those of Hi at 40 cm, 1 m, and 1.7 m.

At the highest D/Rs (nearest distances), RMS errors for Lo+Hi stimuli were never larger than those for the corresponding Lo and Hi noises (at distances of 15 and 40 cm, the hexagrams fall on top of the lower, single-band-response triangles). However, at lower D/Rs (far distances), the RMS errors were larger for Lo+Hi noises than for Hi noises (at 1 and 1.7 m, downward triangles fall below light gray hexagrams). Moreover, RMS performance for Lo+Hi stimuli fell between the levels for Lo and Hi stimuli. These results suggest that when localizing Lo+Hi at low D/Rs, listeners made use of high-frequency information when judging azimuth angle, because they performed better than with Lo noise alone. However, at these low D/Rs listeners did not weight information in the different frequency bands optimally, as performance was worse than for the high frequencies presented alone. Listeners would have localized the Lo+Hi noise more accurately if they had ignored the low-frequency portion of stimuli; thus, listeners are weighting low-frequency cues more heavily than is optimal in the low-D/R conditions.

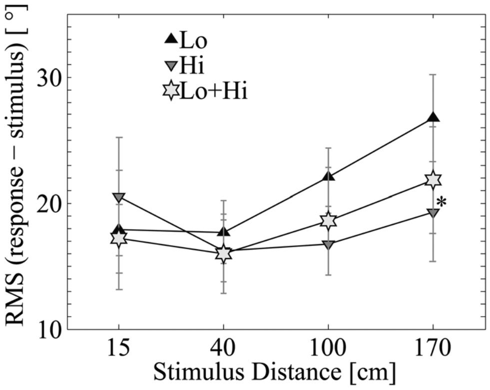

FIG. 4. Mean localization RMS errors for location judgment with Lo, Hi, and Lo+Hi (black upward, gray downward triangles, and light gray hexagrams, respectively). Error bars show one standard deviation across listeners.

A repeated measures ANOVA of RMS error with main factors of distance and frequency content found that both main effects were significant [$F(2,12) = 10.7$, $p = 0.002$, $\omega_p{}^2 = 0.62$, and $F(3,18) = 9.6$, $p = 0.001$, $\omega_p{}^2 = 0.59$, for distance and frequency content, respectively], and found a significant interaction between the two factors [$F(6,36) = 6.6$, $p < 0.0001$, $\omega_p{}^2 = 0.48$]. Pairwise *post hoc* testing between Lo+Hi and Lo or Hi at each stimulus distance (two-tailed t-tests at the 5% significance level) found a statistically significant difference at 1.7m between Lo+Hi stimuli and Hi stimuli ($p = 0.0204$, with Bonferroni correction).

V. SUMMARY AND DISCUSSION

A. Effects of acoustic spatial cues

As outlined in the Introduction, reverberant energy degrades the acoustic information in ITDs and ILDs differently. In everyday reverberant rooms, the average and mode of the distribution of short-term ITDs are unaffected by reverberant energy (Shinn-Cunningham and Kawakyu, 2003; Shinn-Cunningham *et al.*, 2005a; Rakerd and Hartmann, 2010). In contrast, across-time average ILDs decrease systematically as reverberation increases (Shinn-Cunningham *et al.*, 2005a). Although reverberation does not change the mean of short-time ITDs, it does add variability (Shinn-Cunningham and Kawakyu, 2003; Rakerd and Hartmann, 2010). Reverberant energy also increases variability in short-term ILDs systematically as a source is displaced away from the median plane (Ihlefeld and Shinn-Cunningham, 2004).

The Appendix reports detailed acoustic analyses of the spatial cues available in the stimuli used in the current study. These analyses confirm that for these stimuli, mean ITDs changed very little with source distance, whereas both binaural coherence and mean ILDs decreased with increasing source distance. This simple acoustic analysis suggests that if listeners could compute the across-time average of the ITD and ILD cues available in the signals reaching the ears, performance should have been more accurate for the Lo stimuli than for the Hi stimuli. Instead, behaviorally, we found that listeners performed similarly for Lo and Hi stimuli at a distance of 15 cm, but responded more accurately to Hi stimuli than Lo stimuli for more distant sources. This finding suggests that short-term fluctuations in ITDs interfere with localization, even when the across-time average ITD is reliable. In short, listeners were able to interpret noisy, but biased ILD cues more accurately than they could interpret noisy, but unbiased ITD cues.

For sources more than 45° from the median plane, responses were biased toward the median plane, an effect that was stronger the greater the distance of the source from the listener. In contrast, for sources within 45° of the median plane, listeners' responses tended to be laterally biased; however, even these sources were perceived as coming from more medial directions as distance increased. Indeed, for sources within 45° of the median plane but at a distance of

1.7 m, sources were perceived at the correct angle (the lateral bias disappeared).

B. Potential complications of the current study

1. Novelty of sources very close to the head

In everyday life, listeners rarely hear sources as close to the head as 15 cm. Such nearby sources contain exceptionally large ILDs, even at low frequencies where ILDs are near zero for more distant sources (Brungart 2001; Shinn-Cunningham *et al.,* 2005a). The lateral bias of perceived source azimuth for nearby sources in the current results may reflect a misinterpretation of the large ILDs sources contain when they are very close to the listener, as if listeners either do not take into account source distance when interpreting ILDs, or overestimate the source distance of nearby sources. Consistent with the latter explanation, some previous studies find that listeners tend to overestimate source distance for sources near the listener (e.g., see Brungart 2001; Zahorik 2002). Either way, the current results suggest that listeners mistake the large ILDs present in nearby sources as signaling that the source is more lateral than its true direction, increasing lateral response bias.

2. Limited response range

Some of the medial bias in responses could be due to edge effects, as the GUI limited listener responses to a range of +/−90°. However, previous work measuring perceived lateral angle with ILD pointers instead of a GUI (a method without a hard-limited response range) showed a similar pattern (Devore *et al.,* 2009). In that study, BRIRs were simulated in different reverberant conditions, without including the acoustic effects of head shadow and ears. Listeners matched the perceived direction of a low-frequency, band-limited noise by adjusting the ILD of a high-frequency band-limited noise pointer.

Results showed that the range of ILDs used to match different perceived source directions decreased as the amount of reverberant energy increased, consistent with listeners perceiving more distant sources as coming from closer to the median plane. This previous result supports the idea that the medial bias in the current results is not due solely to limitations in the response range.

In the current study we also found that the size of the medial bias for sources simulated from large lateral angles depended on the frequency content of the stimuli, even when the same response method was used: even if there were an “edge effect” caused by the limited allowable response range, it cannot account for these differences in medial bias across stimulus conditions. Together, these results support the notion that reverberant energy biases listeners to perceive sources as closer to the median plane (cf., Devore *et al.,* 2009).

C. Effects of T60 and D/R

There are undoubtedly many factors contributing to differences in lateralization of Lo and Hi noises in the current study. One important factor is simply how much reverberant energy is present in different frequencies. T60s were, if anything, longer in the frequency range of the Hi stimuli than in the frequency range of the Lo stimuli (see Sec. II), which might lead one to expect greater acoustic degradation of spatial cues (and a bigger effect of reverberant energy) for Hi stimuli than for Lo stimuli. However, a detailed analysis of D/R in the Lo and Hi stimuli revealed that for the same source azimuth and distance, there was always less reverberant energy in the Hi than in the Lo condition [see the Appendix, Figs. 5(G) and 5(H)]. Thus, the reverberant energy is likely to have had a smaller effect, acoustically, on spatial cues in Hi stimuli than in Lo stimuli. Still, even though lateralization tended to be less affected by reverberant energy for high-frequency stimuli than for low-frequency stimuli in our simulated room, further studies are needed to examine whether this result generalizes to other rooms.

Although the acoustic degradations of high and low frequencies in our reverberant simulations may explain why Hi stimuli are more accurately localized than Lo stimuli, one of the main points of this study is to consider how listeners combine information across frequencies. Here, our results are unequivocal. At low D/Rs, listeners were slightly less accurate when localizing Lo+Hi stimuli than they were at localizing Hi stimuli. In anechoic conditions, laterality judgments are generally dominated by low-frequency cues (Wightman and Kistler, 1992). Given that at low D/Rs, localization performance appeared worse for Lo than for Hi stimuli, listeners could have reduced their reliance on low-frequency cues when localizing relatively distant sources containing both low and high frequencies in reverberant space to improve accuracy in their localization judgments. However, Experiment 2 indicates that at low D/Rs, listeners were less accurate at localizing Lo+Hi noise than they were at localizing Hi noise presented alone. This pattern of responses is consistent with listeners weighting low-frequency spatial cues too strongly in reverberant space: they would have been more accurate if they ignored the Lo frequency information entirely. However, while the RMS localization error was greater for Lo+Hi noise than for Hi noise for relatively distant sources (with relatively low D/Rs), the RMS localization error for Lo+Hi noise was still smaller than it was for Lo noise. This latter result shows that listeners utilize high-frequency spatial cues when judging the azimuths of these sources: they do not ignore the high-frequency information altogether.

D. Effects of perceived distance

Lateral response angles varied with simulated source distance. In our experiments, listeners were not asked to report source distance, so it is impossible to know whether perceived distance varied with simulated source distance; the current results do not address whether or not perceived source distance affects source laterality judgments. However, using an experimental approach similar to that in the current study, with different room impulse responses and different listeners, previous work from our own lab shows that source laterality judgments are not affected by the requirement that listeners report source distance as well as source laterality (Shinn-Cunningham *et al.,* 2005b). This result at least shows that requiring listeners to make dual judgments of both direction and distance yields lateralization judgments that are indistinguishable from when they are required only to report lateral angle, as in the current study. Future work is necessary to tease apart whether there are perceptual interactions between perceived direction and perceived distance, especially in reverberant conditions; however, this is beyond the scope of the current study.

E. Onset localization cues

The current study was designed to explore how realistic reverberant energy affects localization judgments of stimuli with different frequency content, without focusing on how spatial information is integrated over time. In contrast, many past studies of the effects of reverberation on localization have concentrated on exploring the fact that information at the onsets of sounds (where the D/R is much larger than it is during on- going sound) is both more reliable, acoustically (e.g., see Lochner and Burger, 1964; Faller and Merimaa, 2004) and more-heavily weighted, perceptually (e.g., see Wallach *et al.,* 1949; Litovsky *et al.,* 1999; Rakerd and Hartmann, 2004) than later-arriving spatial information. Indeed, if a listener focused only on onset cues and ignored the degraded, ongoing portion of a reverberant sound source containing both direct and reflected sound energy, they should localize sounds nearly equally well in reverberant and anechoic conditions (e.g., see Devore *et al.,* 2009). However, past studies show that while information in onsets is perceptually weighted more heavily than later-arriving spatial cues, later information still influences localization perception (Hartmann 1983; Stecker and Hafter, 2002; Devore *et al.,* 2009).

Consistent with past work, in the current study we found that even though our stimuli had strong onsets that contained reliable, unambiguous information about the source direction, reverberant energy biased perceived source locations toward the median plane for all three types of noise tested. A similar effect has been reported when listeners localize broadband noise (Shinn-Cunningham *et al.,* 2005b). Moreover, recent work linking neural responses in cat to human localization of broadband noise in simulated reverberation found similar biases in localization with reverberation for stimuli containing only ITDs (Devore *et al.,* 2009). Thus, the current results, taken together with past work, confirm that the degradation of acoustic spatial cues caused by reverberant energy in the ongoing portions of stimuli systematically biases sound localization judgments.

F. Perceptual realism

The BRIRs used to simulate source location in experiments 1 and 2 were recorded on an acoustic mannequin, rather than being measured for each individual subject. Head-related transfer functions (HRTFs) can differ across individuals due to anatomical differences (such as differences in pinnae sizes and shapes), especially at high frequencies (Pralong and Carlile, 1996). Previous work shows that use of nonindividualized HRTFs interferes with elevation judgments and increases the likelihood of cone-of-confusion errors; however, even without individualized HRTFs, listeners are generally good at judging the correct angle relative to the median plane (Wenzel *et al.,* 1993).

The focus of the current study was on how reverberant energy influenced the accuracy of judgments of the cone-of-confusion angle, and how this depended on the frequency content of the stimuli. Thus, even though listeners may have had trouble accurately judging the direction of BRIR-simulated sources in the up/down and front/back dimensions, it is not surprising that they were able to judge the perceived angle of the sources relative to the median plane. Indeed, even without any response feedback, listeners were able to respond consistently in the anechoic training conditions. Furthermore, all of our conclusions are supported by comparisons of lateralization judgments across stimuli with different frequency content, or comparisons across conditions where source angle and distance are parametrically changed, all using the same BRIR simulation method. Moreover, we found consistent effects with the stimulus manipulations used in our study. Even if the use of nonindividualized BRIRs reduced the simulation realism, all of the tested conditions were likely to be affected similarly. Therefore, we expect our conclusions to generalize to studies using individualized simulations as well as to free-field experiments.

VI. CONCLUSIONS

1. In virtual reverberant simulations, perceived lateral angles of sources more than 45° from the median plane are biased toward the median plane, an effect that grows with increasing distance (as D/R decreases). More medial sources tend to be biased laterally; this lateral bias decreases with increasing distance.
2. The localization bias caused by reverberation is greater for low-frequency sounds than for high-frequency sounds.
3. Listeners do not always weight low- and high-frequency localization cues optimally in a reverberant space. In the most reverberant stimulus condition (1.7 m simulated source distance), localization accuracy, as measured by RMS error, is poorer for noises containing both low-frequency and high-frequency components than for a narrowband, high-frequency noise alone, and better than for a narrowband, low- frequency noise alone.

ACKNOWLEDGMENTS

This work was supported by NIH Grant No. R01 DC009477 and the NSF CELEST Science of Learning Center Grant No. SBE-0354378. The authors would like to thank Dr. Sasha Devore and Dr. Eric Thompson for insightful comments. Timothy Streeter and Justin Kiggins assisted with data collection. Two anonymous reviewers provided helpful comments on an earlier version of this manuscript. The BRIRs used in these simulations are available for use in other studies; simply write to shinn@cns.bu.edu.

APPENDIX: ACOUSTIC CUES IN REVERBERANT SETTINGS

BRIRs were convolved with a token of pink noise and processed with band-pass filters identical to those used for generating Lo and Hi noise. For each type of stimulus (frequency content), distance, and lateral source angle, ITDs were estimated by cross-correlating left- and right-ear noise tokens for interaural delays ranging from -800 to 800 μs. These cross correlation functions were then normalized by the square root of the product of the squared left and right ear signals to yield the normalized interaural correlation function. The interaural delay of the peak in the normalized cross correlation function was used to estimate the ITD in the stimulus, while the peak height estimated the interaural coherence.

Figures 5(A) and 5(B) show the ITD for stimuli centered on low-frequency (750 Hz) and high-frequency (6000 Hz) portions of the BRIR-processed noise tokens, respectively. Different lines denote different source distances. ITDs generally increased monotonically with stimulus angle for all source distances. For sources at the greatest distance, the greatest peak in the cross correlation was sometimes in the hemifield opposite the true source angle. However, these "reversals" generally corresponded to ITDs that were one cycle away from the expected ITD in the correct hemi-field (e.g., for the low-frequency band centered at 750 Hz, the reversal occurs 1/750 Hz or 1.3 ms away from the expected ITD). Such reversals are more likely to occur for ITD computations encompassing a narrow bandwidth, where the cross correlation function averages across fewer frequencies, than for spectrally wider sounds. Consistent with this, such reversals in the peak ITD were observed in the Lo but not in the Hi band [cf. Figs. 5(A) and 5(B)]. For narrowband sounds, multiple peaks in the ITD function can lead to bimodal distributions of sound localization judgments or ambiguity in spatial percepts (e.g., see Zhang and Hartmann, 2006). However, in the Lo conditions of Experiment 1, there was no evidence of such confusion: no reversals were observed for Lo noise (see main results). It may be that other directional cues (such as ILDs or ITDs in other off-frequency channels) disambiguate which of two nearly equal peaks in the cross-correlation of the "on frequency" band represent the true source laterality (Trahiotis and Stern, 1989).

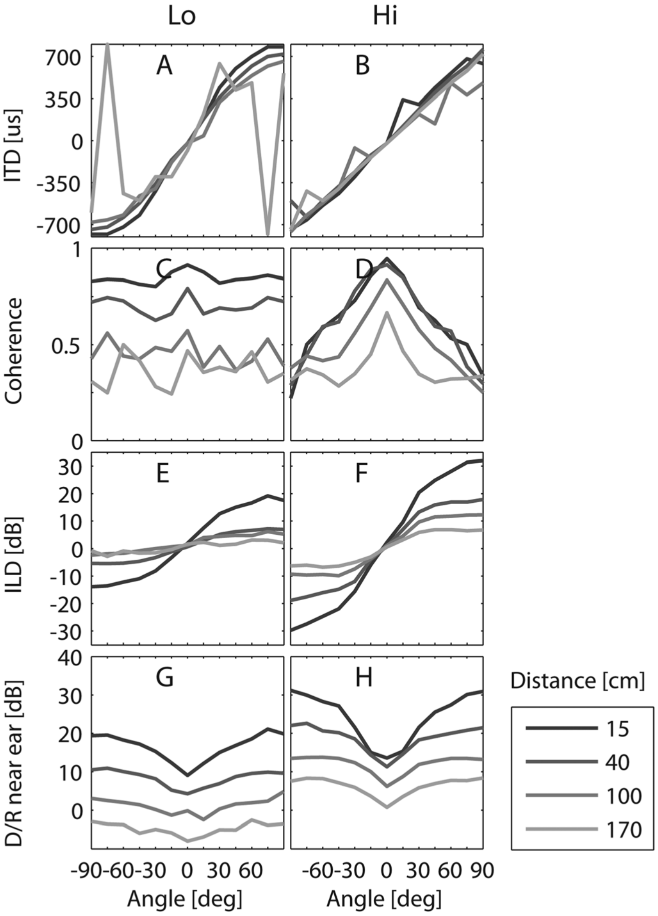

FIG. 5. Binaural acoustic cues (see text for details). (A,B) ITDs, (C,D) coherences, (E,F) ILDs, (G,H) D/R. Left and right column show Lo and Hi conditions, respectively. Different shades of gray denote different distances (dark to light for 15 cm to 1.7 m).

In general, the magnitude of the estimated ITDs decreased with increasing distance. At the 1.7m distance, ITDs were not exactly zero for sources at 0° azimuth, a fact that helps explain the slight rightward bias in localization judgments of these sources [see Fig. 1(A) at 1.7 m distance].

Binaural coherence is shown in Figs. 5(C) and 5(D) for Lo and Hi noises, respectively. In general, binaural coherence decreased with increasing source distance. This result shows that the reliability of ITDs decreases as source distance increases.

ILDs were calculated as the dB ratio between the left and right ear signal levels [Figs. 5(E) and 5(F)]. For a fixed lateral angle, ILDs decreased with increasing distance; at a fixed distance, ILDs increased monotonically with lateral source angle. ILDs were greater for high-frequency sounds than for low-frequency sounds.

Finally, D/R was estimated as the dB difference in energy between the first 10 ms of each BRIR and the remainder of the recording [Fig. 5(G) and 5(E)]. D/Rs decreased with increasing distance and were generally larger for high-frequency than for low-frequency sounds.

---

[1] The time window 10 ms was determined by visual inspection of all BRIR recordings in this study. In all conditions here, the first 10 ms of each BRIR includes all of the direct sound, but excludes the majority of the reflected energy (although some of the first reflection from the floor may be included; this reflection, however, has the same lateral angle as the direct sound; see, for example, Shinn-Cunningham *et al.*, 2005a).

[2] At the greatest (1.7 m) distance, listeners' responses were biased slightly toward the right, consistent with a modest asymmetry in the ITDs of our measured HRTFs [cf., Appendix, Fig. 5(A)].

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

Brungart, D. S. (**2001**). "Preliminary model of auditory distance perception for nearby sources," in *Computational Models of Auditory Function*, edited by S. Greenberg and M. Slaney (IOS Press, Amsterdam), pp. 83–95.

Devore, S., Ihlefeld, A., Hancock, K., Shinn-Cunningham, B. G., and Delgutte, B. (**2009**). "Accurate sound localization in reverberant environments is mediated by robust encoding of spatial cues in the auditory midbrain," Neuron **62**, 123–134.

Devore, S., and Delgutte, B. (**2010**). "Effects of reverberation on the directional sensitivity of auditory neurons across the tonotopic axis: influences of interaural time and level differences," J. Neurosci. **30**, 7826–7837.

Faller, C., and Merimaa, J. (**2004**). "Source localization in complex listening situations: Selection of binaural cues based on interaural coherence," J. Acoust. Soc. Am. **116**, 3075–3089.

Giguère, C., and Abel, S. M. (**1993**). "Sound localization: Effects of reverberation time, speaker array, stimulus frequency, and stimulus rise/decay," J. Acoust. Soc. Am. **94**, 769–776.

Hartmann, W. M. (**1983**). "Localization of sound in rooms," J. Acoust. Soc. Am. **74**, 1380–1391.

Henning, G. B. (**1974**). "Detectability of interaural delay in high-frequency complex waveforms," J. Acoust. Soc. Am. **55**, 84–90.

Ihlefeld, A., and Shinn-Cunningham, B. G. (**2004**). "Effect of source locations and listener location on ILD cues in a reverberant room," J. Acoust. Soc. Am. **115**, 2598.

Jeffress, L. A. (**1972**). "Binaural signal detection: Vector theory," in *Foundations of Modern Auditory Theory*, edited by J. V. Tobias (Academic Press, New York), Vol. II, pp. 351–368.

Litovsky, R. Y., Colburn, H. S., Yost, W. A., and Guzman, S. J. (**1999**). "The precedence effect," J. Acoust. Soc. Am. **106**, 1633–1654.

Lochner, J. P. A., and Burger, J. F. (**1964**). "The influence of reflections on auditorium acoustics," J. Sound Vibr. **1**, 426–454.

Macpherson, E. A., and Middlebrooks, J. C. (**2002**). "Listener weighting of cues for lateral angle: The duplex theory of sound localization revisited," J. Acoust. Soc. Am. **111**, 2219–2236.

McFadden, D., and Pasanen, E. G. (**1976**). "Lateralization at high frequencies based on interaural time differences," J. Acoust. Soc. Am. **59**, 634–639.

Pralong, D., and Carlile, S. (**1996**). "Generation and validation of auditory space," in *Virtual Auditory Space: Generation and Applications*, edited by S. Carlile (Landes, Austin), pp. 142–147.

Rakerd, B., and Hartmann, W. M. (**1985**). "Localization of sound in rooms II: The effect of a single reflecting surface," J. Acoust. Soc. Am. **78**, 524–533.

Rakerd, B., and Hartmann, W. M. (**1986**). "Localization of sound in rooms III: Onset and duration effects," J. Acoust. Soc. Am. **80**, 1695–1706.

Rakerd, B., and Hartmann, W. M. (**2004**). "Localization of noise in a reverberant environment," in *Auditory Signal Processing: Physiology, Psychoacoustics, and Models*, edited by D. Pressnitzer, A. de Cheveigné, S. McAdams, and L. Collet (Springer Verlag, Berlin), pp. 414–422.

Rakerd, B., and Hartmann, W. M. (**2010**). "Localization of sound in rooms, V: Binaural coherence and human sensitivity to interaural time differences in noise," J. Acoust. Soc. Am. **128**, 3052–3063.

Schroeder, M. R. (**1965**). "New method of measuring reverberation time," J. Acoust. Soc. Am. **37**, 409–412.

Shinn-Cunningham, B. G., Santarelli, S. G., and Kopĉo, N. (**2000**). "Tori of confusion: Binaural cues for sources within reach of a listener," J. Acoust. Soc. Am. **107**, 1627–1636.

Shinn-Cunningham, B. G., and Kawakyu, K. (**2003**). "Neural representation of source direction in reverberant space," in *Proceedings of the 2003 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, New Pfaltz, New York, pp. 79–82.

Shinn-Cunningham, B. G. Kopĉo, N., and Martin, T. (**2005a**). "Localizing nearby sound sources in a classroom: Binaural room impulse responses," J. Acoust. Soc. Am. **117**, 3100–3115.

Shinn-Cunningham, B. G., Lin, I. F., and Streeter, T. (**2005b**). "Trading directional accuracy for realism," in *Proceedings of the Human–Computer Interaction International 2005/1st International Conference on Virtual Reality*, pp. 22–27.

Stecker, G. C., and Hafter, E. R. (**2002**). "Temporal weighting in sound localization," J. Acoust. Soc. Am. **112**, 1046–1057.

Strutt, J. W. (**1907**). "On our perception of sound direction," Philos. Mag. **13**, 214–232.

Trahiotis, C., and Stern, R. M. (**1989**). "Lateralization of bands of noise: effects of bandwidth and differences of interaural time and intensity," J. Acoust. Soc. Am. **86**, 1285–1293 (L).

Wallach, H., Newman, E. B., and Rosenzweig, M. R. (**1949**). "The precedence effect in sound localization," Am. J. Psychol. **52**, 315–336.

Wenzel, E. M., Arruda, M., Kistler, D. J., and Wightman, F. L. (**1993**). "Localization using nonindividualized head-related transfer functions," J. Acoust. Soc. Am. **94**, 111–123.

Wightman, F. L., and Kistler, D. J. (**1992**). "The dominant role of low-frequency interaural time differences in sound localization," J. Acoust. Soc. Am. **91**, 1648–1661.

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

Zhang, P. X., and Hartmann, W. M. (**2006**). "Lateralization of sine tones-interaural time vs phase (L)," J. Acoust. Soc. Am. **120**, 3471–3474.