Proceedings of the ACM SIGGRAPH and EUROGRAPHICS Campfire: Acoustic Rendering for Virtual Environments, Snowbird, Utah, 26-29 May 2001, 17-22.

LOCALIZING SOUND IN ROOMS

Barbara Shinn-Cunningham

Department of Cognitive and Neural Systems and Biomedical Engineering
Boston University, 677 Beacon St., Boston, MA 02215
Email: shinn@cns.bu.edu
Ph: 617-353-5764
FAX: 617-353-7755

INTRODUCTION

Relatively little psychoacoustic work has examined how realistic echoes and reverberation affect spatial auditory perception. Within psychoacoustics, echoes and reverberation are generally thought to 1) cause little degradation in directional perception (as suggested by studies of the "precedence effect"; e.g., see Litovsky, Colburn, Yost & Guzman, 1999) and 2) improve distance perception (by some essentially unknown mechanism; e.g., see Mershon & King, 1975).

Head-related transfer functions (HRTFs) show how the signals that reach the two ears are related to the original source signal from a specific location in space (Wightman & Kistler, 1989a; Wightman & Kistler, 1989b; Wenzel, 1992; Carlile, 1996). HRTFs have been examined in detail in anechoic space as a function of source direction and, more recently, as a function of source distance (Brungart & Rabinowitz, 1999b). Typically, such HRTFs are relatively smooth (as a function of frequency) at low frequencies, with notches and peaks above about 6 kHz. The frequency locations of these notches and peaks depend on source elevation and are used by listeners to determine source elevation (e.g., see Wenzel, Arruda, Kistler & Wightman, 1993; Middlebrooks, 1997). Changes in the laterality of the source (relative to the median plane) cause changes in the interaural time difference (ITD) between the signals reaching the left and right ear, a cue known to mediate perception of source laterality (for a review, see Middlebrooks & Green, 1991). Changes in both source laterality and source distance cause changes in the interaural level difference (ILD, difference in the magnitude spectra of the left and right HRTFs; e.g., see Shinn-Cunningham, Santarelli & Kopčo, 2000b). Recent studies of anechoic localization show that ILDs convey some distance information to listeners when sources are near the head (Brungart & Durlach, 1999a).

Recent work in my laboratory addresses how echoes and reverberation influence localization in two ways: by 1) taking empirical measures of the sounds that reach a listener's ears in a room (and studying how these signals vary with source location and listener position) and 2) measuring human localization performance (in three dimensions) when listeners are presented with realistic reverberant signals. Results suggest that spatial perception is affected by room acoustics more than the literature might suggest; and that high-level factors, such as knowledge and experience, have a notable impact on how subjects interpret spatial cues in a reverberant space.

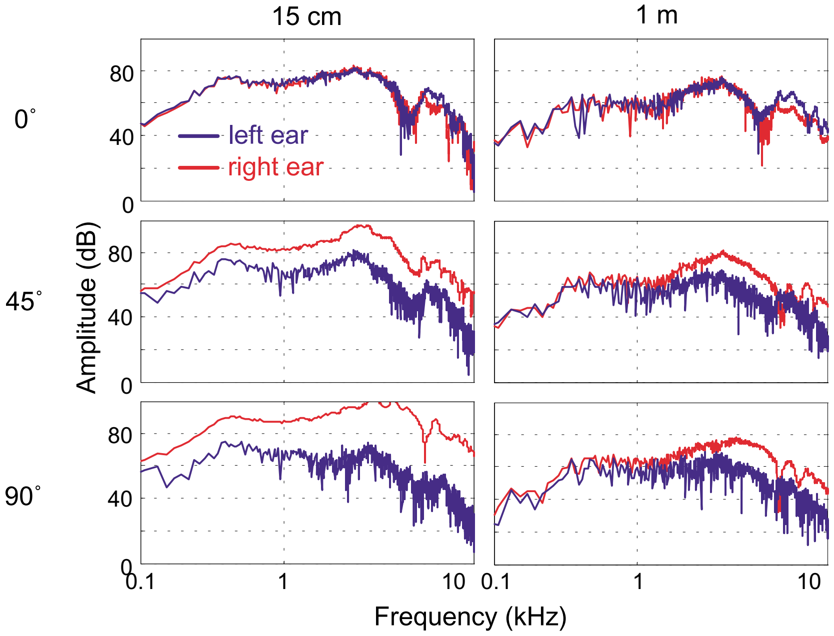

Figure 1: Magnitude spectrum (dB) of HRTFs in the center of a reverberant room as a function of source position relative to listener. Left and right columns show near (15 cm) and far (1 m) sources, respectively. Top, middle, and bottom rows show the lateral angle of the source relative to median plane (0ß, 45ß, and 90ß to the right, respectively).

ACOUSTIC MEASURES

In order to understand how human perceivers perceive auditory source position in rooms, it is important to examine how echoes and reverberation affect the cues thought to underlie spatial perception. HRTFs were measured for a source and listener in a reverberant room (broadband T_{60} ~

450 ms) using a maximum-length-sequence technique. Measurements were made for individual human listeners as well as a KEMAR manikin for sources at different positions (relative to the head) as well as different listener positions within the room (Brown, 2000).

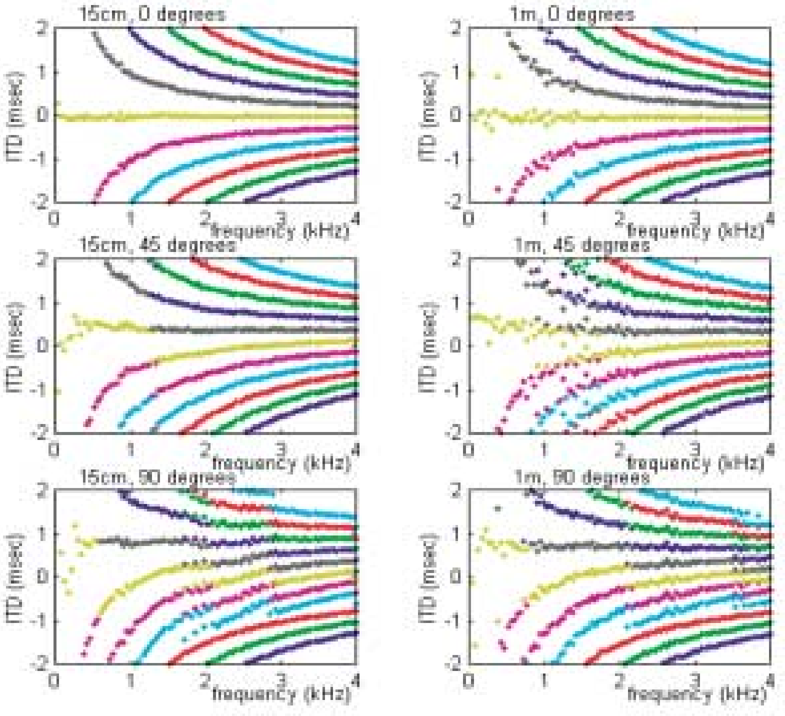

Figure 2: Interaural phase difference versus frequency for the same listener and source positions as in Figure 1.

Figure 1 shows the HRTF magnitude spectra for a source at various positions relative to KEMAR, who was positioned in the center of the room. In addition to randomly distorting the signal spectra reaching the ears, reverberation reduces the depth of any spectral notches. Because the notch depths are reduced by reverberation, one might expect that elevation perception is less robust in a real room than it is in anechoic space (see also Begault, 1992b). The effects are greatest at the ear farther from the source because it receives less direct energy (making the reverberant energy relatively stronger). For the source positions shown (to the right of the listener), the left ear signal (blue) is affected more than the right ear signal (red). The effect of reverberation increases with distance for both left and right ears because the direct sound level decreases; for the cases shown, the effect of reverberation is greater in the right column (source at 1 m) than the left column (source at 15 cm). Finally, source laterality affects the influence of reverberation as well; the effects increase at the left ear and decrease at the right ear as the source moves from 0ß (top row) to 90ß right (bottom row).

Echoes and reverberation also distort interaural differences, and the amount of distortion grows with source distance and laterality. Figure 2 shows ITD as a function of frequency for the same source positions and listener position shown in Figure 1. At any single frequency, there is an essential ambiguity in the interaural time difference that corresponds to a phase difference (at that frequency) of 2π rad. The true interaural time delay is that value which yields approximately the same ITD at all frequencies. In anechoic space, similar calculations lead to an essentially flat line as a function of frequency (e.g., see blue symbols in Figure 4). However, as seen in Figure 2, the effect of reverberation is to introduce noise into the ITD as a function of frequency. Thus, one might expect judgments of source laterality to be affected by echoes and reverberation, although these effects may be small due to the precedence effect (e.g., see Litovsky et al., 1999). Similar results obtain when one examines interaural level differences (ILDs), although there is a tendency for echoes and reverberation to reduce the ILD magnitude in addition to generating frequency-to-frequency distortions.

Results show that the effects of echoes and reverberation depend on the location of the source relative to the listener. Of course, results also depend on the listener location in the room. For a listener located near a wall or other reflective surface, the influence of the resulting early-arriving, intense echo can cause large distortions in the magnitude spectra at the ears, the interaural phase differences, and the interaural level differences. These distortions are much more dramatic than those that occur when the listener is in the center of the room. In fact, early-arriving reflections cause comb-filtering effects characterized by deep notches and rapid phase shifts with frequency, both of which can lead to large distortions of spatial cues.

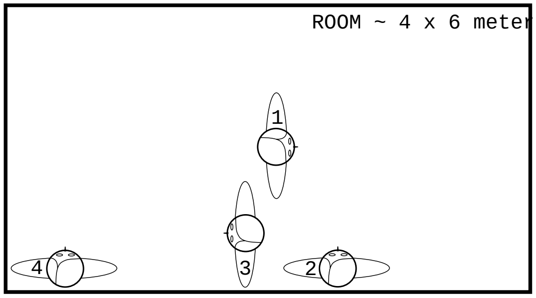

Figure 3: Four listener configurations for which HRTFs were measured in a reverberant room (not to scale).

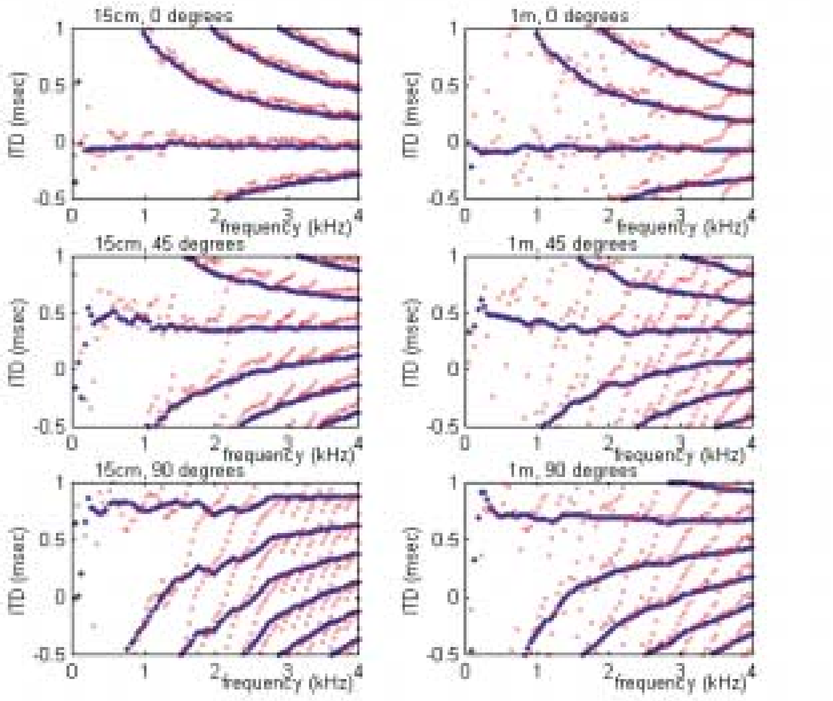

Figure 4: ITD versus frequency for the same source positions as in Figure 2. Blue symbols show anechoic and red symbols reverberant results for a listener located in the corner of the room.

In order to systematically evaluate HRTFs as a function of listener location and orientation, HRTFs were measured for four different configurations of the listener (Kopčo & Shinn-Cunningham, 2001). Figure 3 diagrams the listener positions/orientations for which HRTFs were measured (for the same six relative source positions shown in Figures 1 and 2). Results show that the cleanest results are obtained when a listener is in the center of the room (configuration 1 in Figure 3). Spatial cues becoming increasingly degraded as the listener approaches a wall (configuration 3), are even worse when the subject has his back to the wall (configuration 2), and are most distorted when the listener is located in the corner of the room (configuration 4).

Figure 4 demonstrates how much worse the acoustic distortion can be by showing ITD as a function of frequency for the same relative source positions as in Figure 2, but for listener configuration 4 (corner of the room; note that Figs 2 and 4 use different ITD scales). The blue symbols show the ITD that would arise for anechoic HRTFs; the red symbols show the corresponding ITD for the reverberant HRTFs. For a source near to and directly in front of the listener, echoes and reverberation only marginally affect ITD; however, for all other conditions, the ITD is dramatically distorted.

Taken as whole, acoustic measures suggest that directional localization performance should be degraded in a room compared to in anechoic space, and that this degradation should depend on where the listener is located in the room. Directional performance should be worst when a subject is located in the corner of the room and best when a listener is in the center of the room. In contrast, reverberation should provide source distance information. The degree to which distance perception varies with source and listener position may help in teasing out what aspects of reverberation provide distance information to the listener.

BEHAVIORAL MEASURES

Human localization performance was measured in the same room in which acoustic measures were made (Santarelli, Kopčo & Shinn-Cunningham, 1999; Santarelli, 2000; Santarelli, Kopčo & Shinn-Cunningham, 2000; Kopčo et al., 2001) using an experimental procedure essentially identical to that employed in a previous anechoic localization study (Brungart et al., 1999a). In the experiments, a human experimenter positioned a small speaker at a random location near the listener, whose eyes were closed, and a broadband signal was presented. The actual position of the speaker was measured using an electromagnetic tracker (Polhemus) mounted on the speaker, and the speaker was moved to a neutral position. The listener then opened his eyes and used a pointer to indicate the heard position of the source (in three-dimensional space). A second electromagnetic tracker, affixed to the end of the pointer, measured the response. At the beginning of the experiments, subjects were given an hour of practice on the task, just as in the previous anechoic study (Brungart et al., 1999a).

An initial experiment (Santarelli et al., 1999) confirmed that directional perception was degraded in the room compared to anechoic space, but that distance perception was vastly improved. However, in this initial experiment, two conditions were run. In both conditions, the listener was located in the center of the room. In the first condition, there were no objects near the listener. In the second condition, a 6 x 4 plywood board, covered in acrylic paint, was positioned just to the left of the listener. We anticipated that subjects localization accuracy would be much worse in the second condition compared to the first, due to the presence of the board (and the concomitant early, intense reflections). Instead, we found that the listeners, all of whom performed the two conditions in the same order (first without the board, then with the board in place), were more accurate in localizing sources in the second condition in every spatial dimension. Further examination of the data showed that listener s accuracy improved over hours of practice in the first condition but was essentially unchanged during the second condition. No similar change were seen in the previous anechoic data (reanalyzed for these trends). These results imply that subjects "adapt" to a room over time, and that whatever the subjects learn transfers from one configuration (without a board) to another (with the board in place) that is very different, acoustically.

A follow-up study was recently conducted to explore how robust these effects are (Kopčo et al., 2001). We hypothesized that with practice in a room, subjects adapt and localization improves, and that this learning transfers from one listener configuration to another; i.e., that there is some "room specific" characteristics of reverberation common across all listener positions and orientations in the room. To examine these hypotheses, two groups of listeners performed a localization task similar to that in the initial experiment. Each listener performed four sessions of localization, each from one of the four configurations shown in Figure 3. The first group performed the sessions in the order indicated in Figure 3, starting in the center of the room (configuration 1) and ending in the corner of the room (configuration 4). The second group performed the sessions in the opposite order.

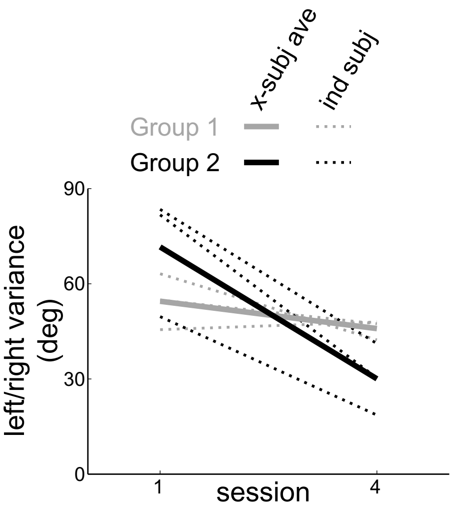

Figure 5: Response variability versus session. Solid lines show across-subject means. Dashed lines show individual subjects.

To the extent that room position affected localization accuracy, we hypothesized that performance would be best when listeners were in the center and worst when listeners were in the corner of the room. To the extent that practice in the room improved localization, performance should be better in the last session of the experiment and worst in the initial session, independent of room configuration order. If both factors influence localization accuracy, the second subject group should show the largest improvement from session one to session four, because both the acoustic and the learning effects would push the results in the same direction. In contrast, for the first subject group, who begin the experiment in the easiest acoustic setting (but without any prior experience in the room), the two effects would interact. In this case, insight into the relative importance of learning and room acoustics on localization performance could be gleaned by comparing results for the two groups.

Response variability in the left/right dimension is shown in Figure 5 for the two groups for the initial session (left) and the final session (right). Results show that the Group 2 subjects (for whom both learning and acoustic effects should cause performance to be best in session 4) show much larger changes in response variability between session 1 (the most acoustically-challenging, corner configuration) and session 4 (the room center configuration). The Group 1 subjects, who started in the easy room configuration and moved to the hardest room configuration, showed only a modest decrease in variability between sessions 1 and 4.

These results support the hypothesis that both learning and room acoustics influence localization accuracy. In addition, since the learning transfers across room configurations that are acoustically very different (and that lead to very different signals at the ears), the results suggest that with practice on the task, subjects learn some very general characteristic about the room reverberation that is similar for all room positions, independent of the exact structure of the echoes and reverberation interacting with the direct sound.

In another set of experiments (Shinn-Cunningham, Santarelli & Kopčo, 2000a), measured HRTFs were used to simulate anechoic and reverberant listening conditions under headphones. Subjects were asked to indicate the heard distance of the simulated sources for sources that were presented both binaurally and monaurally, for sources to the side (along the interaural axis) and to the front. Simulated source distances ranged from 15 cm to 1 m, the range in which ILD cues vary dramatically with distance for sources along the interaural axis. We expected to find that subjects could judge source distance accurately for binaural presentations of lateral sources because subjects in a real anechoic space have been shown to do relatively well on a similar task. Binaural and monaural presentations of reverberant simulations were used so that we could determine whether the reverberation cue for source distance arose from monaural effects (such as the direct-to-reverberant energy ratio; e.g., see Mershon & King, 1975; Bronkhorst et al., 1999) or binaural effects (such as the interaural decorrelation caused by reverberant energy, which is correlated with the direct-to-reverberant energy ratio).

Results of this study were compelling. In every anechoic condition, subject performance was near chance. In all reverberant conditions, subject performance was well above chance. Further, for lateral sources simulated with the reverberant HRTFs, monaural and binaural distance perception was essentially equal; binaural cues were irrelevant for the task. Interestingly, for medial sources, turning off one ear did affect distance judgments slightly, with subjects consistently overestimating the simulated source distance. However, we believe this bias arises because the simulated sources were heard in the wrong direction (i.e., along the interaural axis), where the pattern of reverberation varies differently with distance than it does for medial sources.

Results suggest that reverberation is an important distance cue. Even when sources are so close to the listener that there exist reliable ILD distance cues, these cues are ignored when listeners expect (are calibrated for) a reverberant listening environment. The cue for distance is probably correlated with the direct-to-reverberant energy ratio, although it is unlikely that the human auditory system can accurately compute such a ratio from the total signal reaching the ear. Further, the distance cue provided by reverberant energy is not a binaural cue, but a monaural cue; however, perceived direction (which is strongly influenced by binaural cues) affects perceived distance.

SUMMARY

Inclusion of realistic echoes and reverberation in virtual auditory environments will have a number of dramatic effects, including increasing the realism of the display (Begault, 1992b; Durlach, Rigapulos, Pang, Woods, Kulkarni, Colburn & Wenzel, 1992; Gilkey, Simpson & Weisenberger, 2001), improving distance perception (Shinn-Cunningham, 2000a), providing information about the room itself (Gilkey et al., 2001), and degrading directional accuracy, albeit slightly (Shinn-Cunningham, 2000b). Relatively little is known about which aspects of reverberation are most critical for each of these perceptual results. Further, it is likely that these different perceptual effects arise from different aspects of the reverberation. For instance, while our results hint that distance perception is driven more by monaural than binaural cues, impressions about room size depend on the amount of interaural decorrelation induced by echoes and reverberation, a binaural cue.

These results have a number of implications for the design of effective, efficient acoustic room simulators, pointing to the need to take into account how various aspects of reverberation influence perception. Further work is necessary to tease apart how reverberation influences various percepts important in virtual environments. More specifically, we must examine how accurately room reflection patterns must be simulated in a virtual environment to achieve accurate distance perception as well as realism (while some work addressed these issues, e.g., Begault, 1992a; Zahorik, Kistler & Wightman, 1994, much more work remains). The fact that, in a real reverberant room, listeners adapt their spatial percepts over time suggests that the human perceiver makes subtle perceptual calibrations in ways that we don t yet understand. In turn, this fact hints that listeners are perceptually sensitive to room acoustics in ways that must be explored and understood in order to develop room simulations that recreate what is important for the human perceiver.

ACKNOWLEDGEMENTS

Norbert Kopčo, Scott Santarelli, Tara Brown, and Scarlet Constant assisted in data collection and analysis. Portions of this work were supported by grants from the Air Force Office of Scientific Research, the Whitaker Foundation, and the Sloan Foundation.

REFERENCES

Begault, D. R. (1992a). Binaural auralization and perceptual veridicalityy. Journal of the Audio Engineering Society, preprint 3421.

Begault, D. R. (1992b). Perceptual effects of synthetic reverberation on three-dimensional audio systems. Journal of the Audio Engineering Society, **40**(11): 895-904.

Bronkhorst, A. W. and T. Houtgast (1999). Auditory distance perception in rooms. Nature, **397**(11 February): 517-520.

Brown, T. J. (2000). Characterization of Acoustic Head-Related Transfer Functions for Nearby Sources. Electrical Engineering and Computer Science. Cambridge, MA, Massachusetts Institute of Technology.

Brungart, D. S. and N. I. Durlach (1999a). Auditory localization of nearby sources II: Localization of a broadband source in the near field. Journal of the Acoustical Society of America, **106**(4): 1956-1968.

Brungart, D. S. and W. M. Rabinowitz (1999b). Auditory localization of nearby sources I: Head-related transfer functions. Journal of the Acoustical Society of America, **106**(3): 1465-1479.

Carlile, S. (1996). Virtual Auditory Space: Generation and Applications. New York, RG Landes.

Durlach, N. I., A. Rigapulos, X. D. Pang, W. S. Woods, A. Kulkarni, H. S. Colburn and E. M. Wenzel (1992). On the externalization of auditory images. Presence , **1**: 251-257.

Gilkey, R., B. D. Simpson and J. M. Weisenberger (2001). Creating auditory presence. Human Computer Interaction, International, New Orleans, LA.

Kopčo, N. and B. G. Shinn-Cunningham (2001). Effect of listener location on localization cues and localization performance in a reverberant room. 24th MeetingAssoc Res Otolaryng, St. Petersburg Beach, FL.

Litovsky, R. Y., H. S. Colburn, W. A. Yost and S. J. Guzman (1999). The precedence effect. Journal of the Acoustical Society of America, **106**(4): 1633-1654.

Mershon, D. H. and L. E. King (1975). Intensity and reverberation as factors in auditory perception of egocentric distance. Perception and Psychophysics, **18**: 409-415.

Middlebrooks, J. C. (1997). Spectral shape cues for sound localization. Binaural and Spatial Hearing in Real and Virtual Environments. R. Gilkey and T. Anderson. New York, Erlbaum**:** 77-98.

Middlebrooks, J. C. and D. M. Green (1991). Sound localization by human listeners. Annual Review of Psychology, **42**: 135-159.

Santarelli, S. (2000). Auditory Localization of Nearby Sources in Anechoic and Reverberant Environments. Cognitive and Neural Systems. Boston, MA, Boston University.

Santarelli, S., N. Kopčo and B. G. Shinn-Cunningham (1999). Localization of near-field sources in a reverberant room. 22nd MeetingAssoc Res Otolaryng, St. Petersburg Beach, FL.

Santarelli, S., N. Kopčo and B. G. Shinn-Cunningham (2000). Distance judgements of nearby sources in a reverberant room: Effects of stimulus envelope. Journal of the Acoustical Society of America, **107**(5).

Shinn-Cunningham, B. G. (2000a). Distance cues for virtual auditory space. Proceedings of the IEEE-PCM 2000, Sydney, Australia.

Shinn-Cunningham, B. G. (2000b). Learning reverberation: Implications for spatial auditory displays. International Conference on Auditory Displays, Atlanta, GA.

Shinn-Cunningham, B. G., S. Santarelli and N. Kopčo (2000a). Distance perception of nearby sources in reverberant and anechoic listening conditions: Binaural vs. monaural cues. 23rd MeetingAssoc Res Otolaryng, St. Petersburg Beach, FL.

Shinn-Cunningham, B. G., S. Santarelli and N. Kopčo (2000b). Tori of confusion: Binaural localization cues for sources within reach of a listener. Journal of the Acoustical Society of America, **107**(3): 1627-1636.

Wenzel, E. M. (1992). Localization in virtual acoustic displays. Presence, **1**(1): 80-107.

Wenzel, E. M., M. Arruda, D. J. Kistler and F. L. Wightman (1993). Localization using nonindividualized head-related transfer functions. Journal of the Acoustical Society of America, **94**: 111-123.

Wightman, F. L. and D. J. Kistler (1989a). Headphone simulation of free-field listening. I. Stimulus synthesis. Journal of the Acoustical Society of America, **85**: 858-867.

Wightman, F. L. and D. J. Kistler (1989b). Headphone simulation of free-field listening. II. Psychophysical validation. Journal of the Acoustical Society of America, **85**: 868-878.

Zahorik, P., D. J. Kistler and F. L. Wightman (1994). Defining and redefining limits on human performance in auditory spatial displays. Second Intl Conf Aud Display, Santa Fe, NM, Santa Fe Institute.