Proceedings of the 2003 International Conference on Auditory Display, Boston, MA, USA, July 6-9, 2003

IDENTIFYING WHERE YOU ARE IN A ROOM: SENSITIVITY TO ROOM ACOUSTICS

Barbara Shinn-Cunningham

Boston University
677 Beacon St.
Boston, MA 02215
shinn@cns.bu.edu

Suraj Ram

Boston University
677 Beacon St.
Boston, MA 02215
sram23@cns.bu.edu

ABSTRACT

In a spatial auditory display, reverberation provides a reliable cue for source distance, increases the subjective realism of the display, and improves the externalization of simulated sound sources. However, relatively little is known about perceptual sensitivity to differences in reverberation patterns or how precisely reverberation must be simulated in a spatial auditory display. This paper presents preliminary results of a study examining sensitivity to changes in listener location in a simulated room. Results suggest that monaural cues in the ear receiving the least direct-sound energy provide the most salient cues for identifying room location. However, many details in the reverberation pattern are not easily perceived. These results indicate that including reverberation from simplified room models may provide the benefits of reverberation without noticeably degrading the realism of the display.

1. INTRODUCTION

Many studies have examined how echoes and reverberation (henceforth referred to jointly as "reverberation") influence perception in both real-world settings and spatial auditory displays. In most ordinary environments, reverberation provides a robust cue for source distance [1-8], but only modestly degrades directional perception [4] and speech intelligibility [9, 10]. Reverberation is also important for improving the subjective realism and externalization achieved in a spatial auditory display [2, 11]. Speech perception and the ability to monitor and understand one source in the presence of competing sources can be degraded by reverberation, but these effects can be modest for the reverberation in many rooms [9, 12]. These results suggest that reverberation is generally useful in an auditory display.

While modest amounts of reverberation are helpful, relatively little is known about how sensitive listeners are to changes in the pattern of reverberation reaching the ears. In a given environment, the exact timing, direction, and intensity of individual reflections reaching the listener depends on the location of the source relative to the listener as well as the location and orientation of the listener in the room [6, 13]. As a result, the reverberation pattern changes whenever the listener or source moves (relative to each other or to the room), making it computationally intensive to calculate and update the pattern of reverberation in real time [2, 14]. Thus, determining sensitivity to changes in reverberation is important for determining how to design a cost-efficient but effective spatial auditory display. This paper reports preliminary results exploring how well listeners can tell where they are in a room from the signals reaching their ears.

2. EXPERIMENT 1

Listeners heard headphone simulations that included realistic reverberation (measured at four listener locations in a classroom). Listeners were asked to identify their room location from the acoustic signals they heard for different source azimuths and distances relative to the head. We hypothesized that the ability to judge room location would improve with source distance because the relative level of the reverberant energy in the signals reaching the listener increases with source distance. We expected source azimuth to have some impact because the levels and pattern of direct and reverberant energy also change with azimuth.

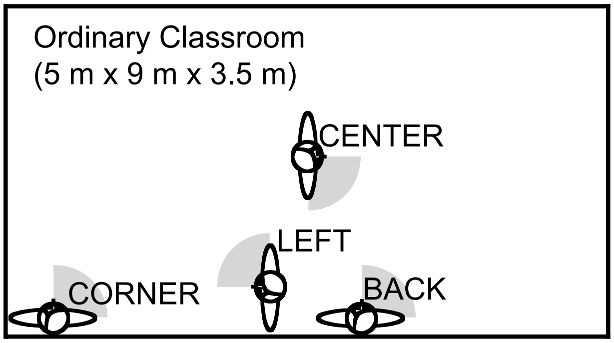

Figure 1. *Sketch of the listener locations in the room (gray area shows source locations re: listener).*

2.1. Methods

Six subjects (students at Boston University) were paid to participate in the study. All subjects had normal hearing as confirmed by an audiometric screening.

KEMAR manikin Head-Related Transfer Functions (HRTFs) were measured in a quiet classroom (broadband T_{60} = 700 ms). Measurements were taken with KEMAR at one of four locations in the room (depicted schematically in Figure 1), denoted *center*, *back*, *left*, and *corner*, which differed in the proximity and number of nearby walls. For each of the room locations, measurements were taken with the sound source in the right front horizontal plane (at ear height) at one of nine source locations (all combinations of azimuths 0, 45, and 90º to the right and distances 0.15, 0.40, and 1 m).

A Maximum-Length-Sequence [15] was used to measure HRTFs with Tucker-Davis Technologies hardware (PD1 D/A converter, HB6 amplifier, Bose loudspeaker). Knowles FG-3329c microphones mounted in earplugs inserted into the entrance of the ear canals measured the acoustic responses. Microphone outputs drove a custom-built microphone amplifier connected to a TDT A/D converter (PD1). Details regarding the measurements are given in [16].

Eight independent samples of pseudo-random white noise bursts (length 743 ms at a 44.1 kHz sampling rate) were convolved in MATLAB with the set of 36 HRTFs (9 source locations x 4 listener locations) to generate binaural stimuli. In order to remove gross intensity differences, these stimuli were normalized so that the right-ear signal in each binaural pair (i.e., the louder of the signals in the binaural pair for the tested source locations) had the same RMS value. Signals were presented from MATLAB through a Creative Labs SoundBlaster soundcard driving Sennheiser HD270 headphones. At the start of each session, listeners adjusted the signals to a comfortable listening level.

On each trial, the subject indicated the perceived listener location by clicking with a computer mouse on one of four graphical buttons labeled with the locations. After each response, the button corresponding to the correct answer was highlighted to provide feedback.

Subjects performed 36 blocks of trials. In each block, all trials simulated the same source location (azimuth and distance); the only differences from trial to trial were due either to changes in the reverberation pattern or random variation across the eight noise samples. Each block consisted of 32 trials (one presentation of each of the 8 noise samples for each of the 4 room locations, in random order). Prior to each block of 32 trials, subjects could listen to presentations from each of the room locations as many times as they wished by clicking on the appropriate button; testing began only when subjects felt ready to proceed.

Subjects performed 12 blocks per experimental session, during which the source distance was held constant (randomly ordered for each subject). The three source azimuths were presented in random order within each session (different orders for each day and subject), so that each day, a subject heard four consecutive blocks from the same source location followed by two sets of four blocks, each simulating a different source azimuth. To reduce any artifacts due to training, the first block in each condition are not analyzed in the results reported here.

2.2. Results

The information transfer ratio T was calculated for each source location and subject. Let $c_{i,j}$ denote the number of times a subject responds that room location j was presented when the actual room location was i. The joint probability of hearing i and responding j is estimated by

$$\hat{p}_{i,j} = \frac{c_{i,j}}{\sum_i \sum_j c_{i,j}} \quad (1)$$

Similarly, the probabilities of hearing stimulus i and of responding j can be estimated as

$$\hat{p}_i = \frac{\sum_j c_{i,j}}{\sum_i \sum_j c_{i,j}} \text{ and } \hat{p}_j = \frac{\sum_i c_{i,j}}{\sum_i \sum_j c_{i,j}}, \quad (2)$$

respectively. The estimated mutual information $M_{I,J}$ between stimuli I and responses J can be computed as

$$M_{I,J} = -\sum_j \hat{p}_j \log\hat{p}_j + -\sum_i \hat{p}_i \log\hat{p}_i - -\sum_j \sum_i \hat{p}_{i,j} \log\hat{p}_{i,j}. \quad (3)$$

The information transfer ratio T can then be computed as

$$T = \frac{M_{I,J}}{-\sum_i \hat{p}_i \log\hat{p}_i}. \quad (2)$$

T ranges between zero and one, and is exactly equal to one if knowing the subject's response perfectly predicts the stimulus presented. Low values of T arise when responses are independent of the stimulus.

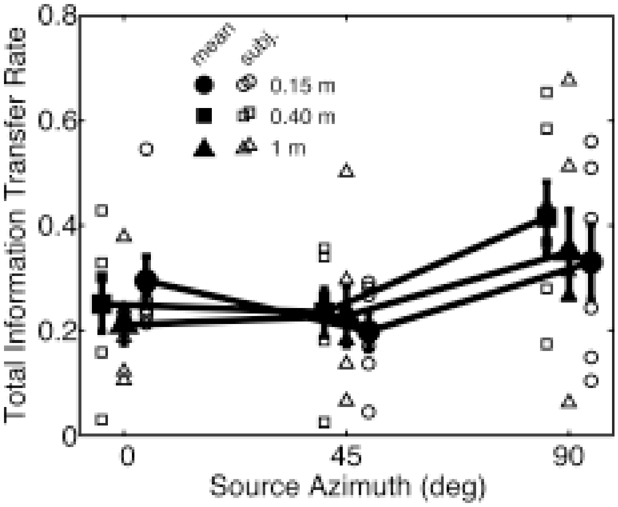

Figure 2. *Information transfer rate T as a function of source azimuth for three source distances (Experiment 1). Across-subject means shown by solid lines (with standard error bars); individual subject results shown by open symbols.*

2.2.1. *Total Information Transfer Ratio*

Overall, performance is relatively poor, with average T values around 0.3 (see Figure 2). There is a modest effect of azimuth: performance is slightly (but significantly) better for sources at 90˚; however, distance caused no significant main effect (multi-way ANOVA analysis, $p > 0.05$). Individual differences are large (e.g., T ranges from 0.1 – 0.7 for 90˚ sources).

2.2.2. *Information Transfer Ratio for Room Pairs*

In order to gain further insight into what the listeners could hear, T was analyzed separately for each pair of room locations (by analyzing the stimulus/response counts $c_{A,j}$

and $c_{B,j}$ for stimulus pairs A≠B). This analysis examines which pairs of room locations were relatively difficult to tell apart and which were relatively easy to discriminate.

For each subject, *T* was calculated for all room pairs for each source location. Analysis showed that the pattern of confusions listeners made varied systematically with source location; however, the effect due to source azimuth was small. For simplicity, values were averaged over source azimuth for each listener and source distance. Figure 3 shows the across-subject mean and standard error of these values for all room pairings and for each of the source distances. These results show that for nearby sources, all room pairings were roughly equally discriminable; however, as distance increased, performance for four room pairs increased while for two pairs it decreased. More specifically, at the 1 m source distance, subjects could not distinguish between the two locations in which the left ear faced a wall (*corner* and *left* positions) or between the two locations in which neither ear faced a wall (*center* and *back)*; however, they rarely made confusions across these categories (e.g., they rarely confused *center* and *corner* locations). These results show that although distance has little impact on overall information transfer (Figure 2), it has a dramatic effect on the pattern of response confusions (Figure 3).

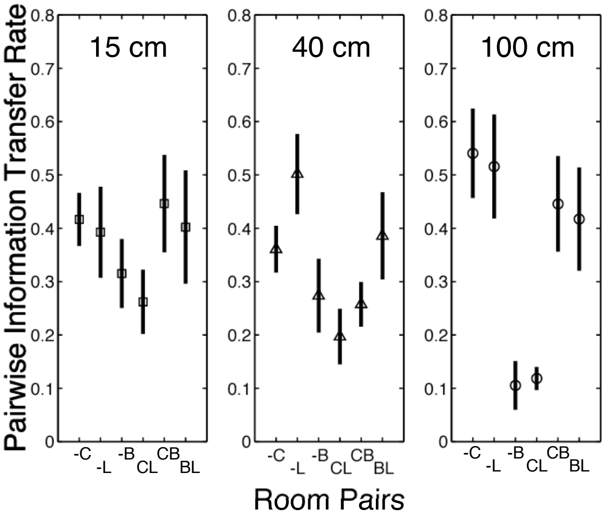

Figure 3. *Pairwise information transfer rate for all combinations of room locations averaged across subjects (with across-subject standard error). Room locations: -: center; C: corner; B: back; L: left.*

2.2.3. *Discussion*

Overall, results show that the ability to discriminate between different room locations depends on the source location relative to the listener. For distant sources, listeners are relatively good at discriminating between room locations in which one ear faces a wall and room locations in which neither ear is facing a wall, but cannot discriminate locations within these categories. However, for nearer sources, listeners have a more modest ability to discriminate across all combinations of room locations. These results suggest that the presence of very early echoes that are strong relative to the direct-sound energy (as occurs when one ear faces the wall and the source is relatively far from the listener) are easy to hear, but that later-arriving energy (which differs for *center* and *back* conditions and for *corner* and *left* positions) is less salient, perceptually.

3. EXPERIMENT 2

Many acoustic cues present in the stimuli in Experiment 1 could be used to perform the task, including binaural decorrelation; monaural pitch, timbre, and spectral-shape effects due to comb-filtering by early, intense echoes; overall loudness differences due to variations in the signal level at the head-shadowed ear [6]; and the pattern of energy over time. Experiment 2 was designed to begin to tease apart which cues are salient by comparing binaural and monaural listening. We expected binaural performance to be best, because all possible cues are available to listeners in this condition. We expected monaural left-ear performance to be next best, because although any dominant spectral cues would be present in the left-ear signal, binaural information would not be available in this condition.

3.1. Methods

Procedures were the same as in Experiment 1, with the following exceptions. Two subjects from Experiment 1 also performed Experiment 2, along with two new subjects. Stimuli were identical except that RMS levels were first equated (by equating the right-ear signal levels in the binaural and right-ear monaural conditions and the left-ear signal level in the left-ear monaural condition) and then randomly roved by 10 dB on each trial to remove any overall intensity cues. Only the farthest distance (1 m) was presented, but three presentation conditions were used (left-ear monaural, right-ear monaural, and binaural). The blocking of trials was otherwise similar (with presentation condition varied from day to day).

3.2. Results

Figure 4 shows *T* for left-ear, right-ear, and binaural conditions as a function of azimuth. Binaural performance is worse than in Experiment 1, probably due to the uncertainty caused by the rove in stimulus intensity. The finding that the level rove appears to degrade binaural performance in Experiment 2 suggests that overall level (changes in the left-ear signal level) provided information in Experiment 1. Results for the right-ear monaural presentation were essentially equal to binaural levels; however, performance for the left-ear monaural presentations is far better than for the other conditions. These results suggest that monaural cues in the left ear signal provide the most information about the reverberation pattern, and that the presence of the right-ear signal during binaural presentation actually makes these cues less easily perceived.

4. DISCUSSION AND CONCLUSIONS

Results of these experiments suggest that listeners reliably hear monaural spectral cues due to very prominent, early echoes in the signals reaching the ears (especially in the ear that receives the least direct sound energy, which is the left ear in the current experiments) and can use them to judge room location. However, listeners are relatively insensitive to other differences in reverberation that change with room and source locations. In addition, the monaural cues in the ear receiving less direct sound energy are less useful when signals are presented binaurally. It is worth noting that all source distances tested in this study are relatively close to the listener, and results may differ when the relative level of reverberation is greater (sources are more distant). Further analysis of how various acoustic cues vary as a function of source and listener location may provide additional insights into how listeners discriminate their location in a room from the signals reaching the ears.

These results support the interpretations of many previous researchers, who conclude that the pattern of late-arriving echoes are not perceptually salient. The current results further suggest that even early echoes need not be simulated with great accuracy. Listeners are relatively insensitive to the exact timing of and arrival direction of echoes: listeners cannot always discriminate between the *center* location (were there are no nearby reflecting surfaces) and *back* location (where the listener has his back nearly touching a wall) or between the *corner* location (where both the listener's back and left side are near a wall) and the *left* location (where only the listener's left side is near a wall).

5. ACKNOWLEDGEMENTS

Portions of this work were supported by the Air Force Office of Scientific Research, the Sloan Foundation, and the National Science Foundation. Nick Cunningham assisted in proofreading this manuscript. Steve Colburn and Bill Hartmann provided valuable comments on this work.

6. REFERENCES

[1] D. H. Mershon, W. L. Ballenger, A. D. Little, P. L. McMurtry, and J. L. Buchanan, "Effects of room reflectance and background noise on perceived auditory distance," *Perception*, vol. 18, pp. 403-416, 1989.

[2] D. R. Begault, E. M. Wenzel, A. S. Lee, and M. R. Anderson, "Direct comparison of the impact of head-tracking, reverberation, and individualized head-related transfer functions on the spatial perception of a virtual speech source," *J Aud Eng Soc*, vol. 49, pp. 904-916, 2001.

[3] P. Zahorik, "Assessing auditory distance perception using virtual acoustics," *J Acoust Soc Am*, vol. 11, pp. 1832-1846, 2002.

[4] S. Santarelli, "Auditory Localization of Nearby Sources in Anechoic and Reverberant Environments," unpublished Ph.D. thesis, Dept. Cognitive and Neural Systems. Boston, MA: Boston University, 2000.

[5] W. L. Martens, "Pseudophonic listening in reverberant environments: Implications for optimizing auditory display for the human user of a telerobotic system," in *Proc Int Conf Info Soc: Emerging Technologies and New Challenges*, Aizu-Wakamatsu, Japan, 2000.

[6] B. G. Shinn-Cunningham, "Distance cues for virtual auditory space," in *Proc IEEE-PCM 2000*, Sydney, Australia, 2000.

[7] B. G. Shinn-Cunningham, "Localizing sound in rooms," in *Proc ACM/SIGGRAPH Eurographics Campfire: Acoustic Rendering for Virtual Environments*, Snowbird, Utah, 2001.

[8] B. G. Shinn-Cunningham, "Speech intelligibility, spatial unmasking, and realism in reverberant spatial auditory displays," in *Proc Int Conf Aud Disp*, Atlanta, GA, 2002.

[9] T. Houtgast, H. J. M. Steeneken, and R. Plomp, "Predicting speech intelligibility in rooms from the modulation transfer function I. General room acoustics," *Acustica*, vol. 46, pp. 60-72, 1980.

[10] K. L. Payton, R. M. Uchanski, and L. D. Braida, "Intelligibility of conversational and clear speech in noise and reverberation for listeners with normal and impaired hearing," *J Acoust Soc Am*, vol. 95, pp. 1581-1592, 1994.

[11] N. I. Durlach, A. Rigapulos, X. D. Pang, W. S. Woods, A. Kulkarni, H. S. Colburn, and E. M. Wenzel, "On the externalization of auditory images," *Presence*, vol. 1, pp. 251-257, 1992.

[12] B. G. Shinn-Cunningham, L. Mraz, and N. Kopco, "Effect of reverberation on spatial unmasking for nearby speech sources," *J Acoust Soc Am*, vol. 109, pp. 2468, 2001.

[13] B. G. Shinn-Cunningham, J. G. Desloge, and N. Kopco, "Empirical and modeled acoustic transfer functions in a simple room: Effects of distance and direction," in *Proc IEEE-WASPAA*, New Pfalz, New York, 2001.

[14] D. R. Begault, B. U. McClain, and M. R. Anderson, "Early reflection thresholds for virtual sound sources," *Proc Int Workshop Spat Media*, Aizu-Wakamatsu, Japan, 2001.

[15] J. Vanderkooy, "Aspects of MLS measuring systems," *J Aud Eng Soc*, vol. 42, pp. 219-231, 1994.

[16] T. J. Brown, "Characterization of Acoustic Head-Related Transfer Functions for Nearby Sources," unpublished M.Eng. thesis, Dept. Elec Eng Comp Sci. Cambridge, MA: Massachusetts Institute of Technology, 2000.

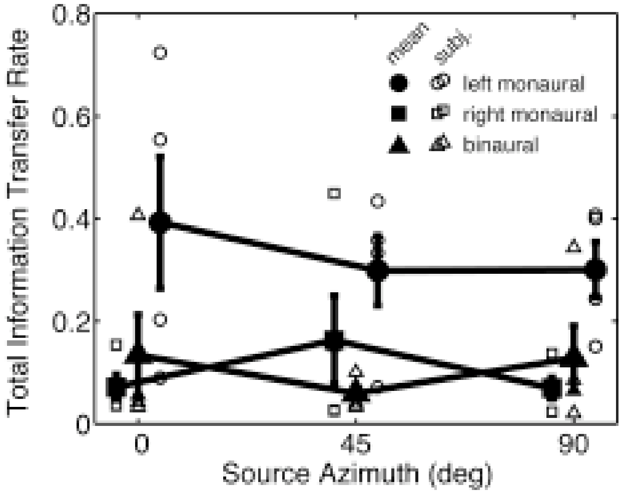

Figure 4. *Information transfer rate T as a function of source azimuth in Experiment 2. Across subject means shown by solid lines (with standard error bars); individual results shown by open symbols.*