The Perceptual Consequences of Creating a Realistic, Reverberant 3-D Audio Display

Barbara G. Shinn-Cunningham

Boston University Hearing Research Center
Boston University, Boston, MA, USA
shinn@cns.bu.edu

Abstract

Relatively little is known about the perceptual sensitivity of listeners to reverberant energy like that present in most everyday environments. This paper briefly summarizes some of the effects reverberant energy can have on the acoustic signals at the ears, listener perception, and models of auditory processing. The consequences of including realistic reverberation in 3-D sound reproduction systems are discussed in light of these findings.

1. Introduction

In an ordinary day, listeners are bombarded with echoes and reverberation, often with energy that exceeds the direct-sound energy. Reverberation affects the spectro-temporal aspects of the signals reaching the ears and performance on many different behavioral tasks. Past studies show that reverberation provides listeners with distance information [1], enables listeners to judge room properties [2], increases audibility of quiet sources [3], and increases the realism of 3-D sound reproduction systems [4]. However, reverberant energy also degrades speech intelligibility [5, 6], localization accuracy [7], and the ability to cope with "cocktail party" listening conditions [8, 9].

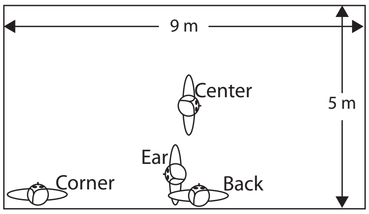

Figure 1: Block diagram of the classroom and listener locations within the classroom.

The computational demands required to render realistic reverberation make understanding how reverberation influences perception a critical issue. What aspects of reverberation must be modeled to gain the benefits of reverberation? Which details can be ignored without loss of realism? What are the consequences of including reverberation on tasks such as understanding speech? This paper reviews some studies of the effects of reverberant energy on acoustic, psychoacoustic, and computational experiments.

2. Classroom Reverberation

Most results reviewed here were gathered with listeners located at one of four locations in an ordinary, moderate-sized classroom (volume roughly 158 m^3; see Figure 1). The classroom had a broadband T_{60} of 700 ms, with a carpeted floor and relatively hard walls.

2.1. Acoustic Effects of Reverberation

The effects of reverberation on the acoustic signals reaching a listener vary with the direct-to-reverberant energy ratio (D/R). In addition to depending on room characteristics, the influence of reverberation depends on the distance and direction of the sound source relative to the listener, which of the two ears one considers, and the listener location in the room [10, 11].

Figure 2 shows how D/R varies at the left (black) and right (gray) ears of a listener in the classroom as a function of the source azimuth relative to the listener (ranging 0° to 90° to the right). Within each panel, the different lines show results for different source distances (from 0.15 m to 1 m). Each individual panel shows results for a different listener location in the room.

For the tested source locations (most of which are near to and to the right of the listener) and listener locations (none of which had a wall near the right ear of the listener), the left ear D/R (black) is substantially smaller than the right ear D/R (gray). The listener location in the room also influences the relative energy of the reverberation reaching a listener: whenever the listener was located with one ear facing a nearby wall, the D/R was much smaller in the ear facing the wall than for other listener locations in the room (compare the left-ear, black results in the bottom two panels to those in the top two panels).

These results demonstrate that the ear near to the source (here, the right ear, shown in gray in Fig. 2) is generally less affected by the distortion caused by reverberation than the ear farther from the source. Consequently, the distortion of interaural time differences (ITDs) and interaural level differences (ILDs) caused by reverberant energy is dominated by the monaural distortion at the ear farther from the source [10]. Analysis of the effects of reverberation on the cues reaching the listener in the classroom show that ITD information can be radically altered by reverberation;

both random and systematic frequency-to-frequency deviations (relative to the corresponding anechoic condition) can arise, depending on the geometry of the source, listener, and room. Despite substantial distortion due to reverberation, the direct-source ITD can be extracted reliably for the levels of reverberation present in the room [10]. However, ILDs in the signals reaching a listener's ears are less robust; not only are the ILDs randomly distorted, they are reduced in magnitude, even for the relatively modest levels of reverberation in the classroom conditions tested in these experiments [10].

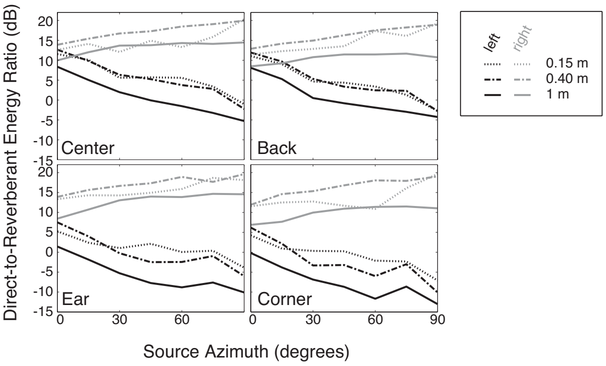

Figure 2: D/R as a function of source azimuth. Each panel shows results for one listener location, three source distances, and both ears.

2.2. Perceptual Effects of Reverberation

Despite the fact that reverberant energy distorts the signals reaching the listener's ears, the perceptual effects of reverberation are often modest. For instance, in the classroom, localization behavior is nearly as good as in anechoic space [12] and does not depend strongly on the listener location in the room [13]. Evidence suggests past experience in an acoustic environment is important for enabling accurate localization performance in a room in at least two ways. Consistent short-term experience allows a listener to calibrate how to interpret spatial cues in a particular environment (e.g., listeners rapidly learn how to judge distance from reverberation cues) [14]. Longer-term experience allows the listener to fine-tune localization judgments so that response variability decreases with experience [12].

Past studies show that in anechoic space, listening with both ears provides substantial benefit over listening monaurally with the ear that has the more advantageous signal-to-noise ratio (i.e., the "better ear") when the target and masking sounds arise from different directions; however, this "binaural advantage" degrades substantially when there is reverberation in the environment [8, 9]. However, in some cases room reverberation does not destroy a listener's ability to understand the content of a spoken sentence in the presence of a masking sound [15]. For instance, in a recent experiment, BRIRs were used to simulate speech and noise sources in different directions and distances relative to the listener. The speech intensity was adaptively varied to find the level at which the speech intelligibility reached threshold. Both center-of-the-room (see Figs. 1 and 2) and anechoic conditions were simulated. Binaural and monaural conditions were tested to determine the binaural advantage.

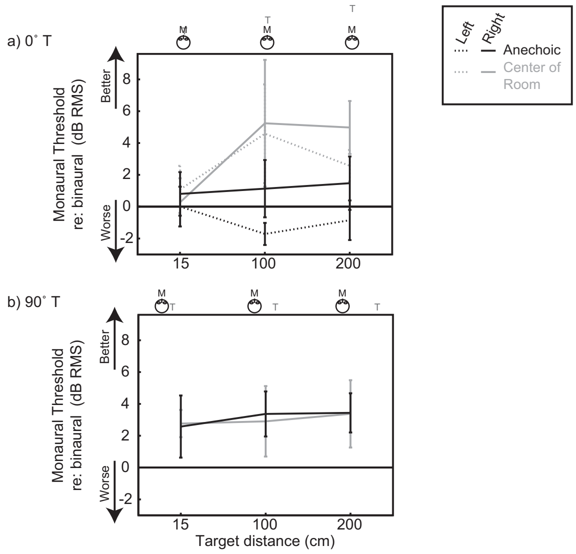

Figure 3: Binaural advantage for a speech-in-noise task. a) Target and masker both at 0°. b) Target at 90° to the right, masker at 0°.

Figure 3 plots the across-subject mean binaural advantage (the improvement when listening with two ears compared to listening only with the acoustically better ear) for the four listeners in the experiment; error bars show the across-subject standard deviation. Within each panel, both reverberant (gray) and anechoic (black) results are shown. Fig. 3a shows results for the conditions in which the target and masker were in the same direction (straight ahead of the listener). Fig. 3b shows results when the target is 90° to the right of the listener and the masker is straight ahead (see cartoons above each panel, which summarize the spatial positions of target and masker relative to the listener).

Results show that there is a binaural advantage in all of the tested conditions when the target and masker are in different directions (Fig. 3b), even when the listener is in the center of a reverberant classroom (gray results). In fact, for these conditions (masker at 0° and speech at 90°), the binaural advantage is of the same magnitude in both anechoic and reverberant conditions (compare gray and black results in Fig. 3b). In other words, reverberation does not always destroy the binaural advantage that arises when target and masker are in different directions; here the binaural advantage is unaffected by reverberant energy.

When target and masker are at 0° (Fig. 3a), there is no "better ear;" both ears have essentially the same target-to-masker ratio. In these anechoic conditions, there is no binaural advantage: performance is the same when listening binaurally and monaurally to either ear (in Fig. 3a, see the solid and dashed black lines representing the difference between binaural threshold and thresholds for the right and left ears, respectively). Perhaps even more intriguingly, when target and masker are both straight ahead in the reverberant condition, a binaural advantage arises when the target is at a different distance than the masker (the gray lines in Fig. 3b are positive for the target at 1 and 2 m).

Binaural advantages usually arise because the target causes interaural decorrelation of the masker. Current results are consistent with this explanation. In all conditions, the masker is so close to the listener that it produces very similar left and right ear signals. In anechoic simulations, the 0˚ target is also highly correlated at the two ears and thus causes no interaural decorrelation when added to the masker (and no binaural advantage; see solid and dashed black lines in Fig. 3a). When the target is at 90˚, it decorrelates both anechoic and reverberant maskers and thus produces a binaural advantage (see all results in Fig. 3b). However, when a reverberant target is in the same direction but a different distance from the listener, the target reverberation decorrelates the left and right ear signals; when added to the interaurally correlated masker, the target produces a binaural advantage (e.g., see solid and dashed gray lines in Fig. 3a).

While classroom reverberation does not necessarily destroy all spatial unmasking effects, the combination of reverberation and masking noise degrades spatial unmasking when listeners identify isolated phonemes [16]. These results are consistent with the hypothesis that listeners can at least partially compensate for destructive effects of reverberant energy under some circumstances [17], but are less able to “average out” or perceptually calibrate to random distortion caused by reverberation when signals are relatively short.

Classroom reverberation also does not provide robust information to the listener about their location within the room. In virtual-auditory-space experiments using the BRIRs analyzed in Fig. 2, trained listeners performed poorly when asked to identify their location in the room from the binaural signals they heard, despite the fact that the BRIRs differ from each other acoustically [18]. In other words, at least in this classroom environment, both the positive and negative consequences of reverberant energy are mild.

Overall, these perceptual results demonstrate that in order to predict whether reverberation will disrupt or help performance on a given task depends in complex ways on the location of the sound sources and listener in the environment as well as the environment itself.

2.3. Effects of Reverberation on Neural Processing

A simple neural model was developed to begin to quantify the effects of reverberant energy on the cues reaching the listener’s ears (see [20] for more detail). The model takes left- and right-ear acoustic signals and simulates the time-varying activity of neurons in the brainstem that are sensitive to interaural time differences. Acoustic signals are first processed by a model of the auditory periphery [19]. The left and right auditory nerve model outputs are then cross-correlated using a short (roughly 100-ms-long), sliding time window and the resulting activity evaluated as a function of center frequency and time.

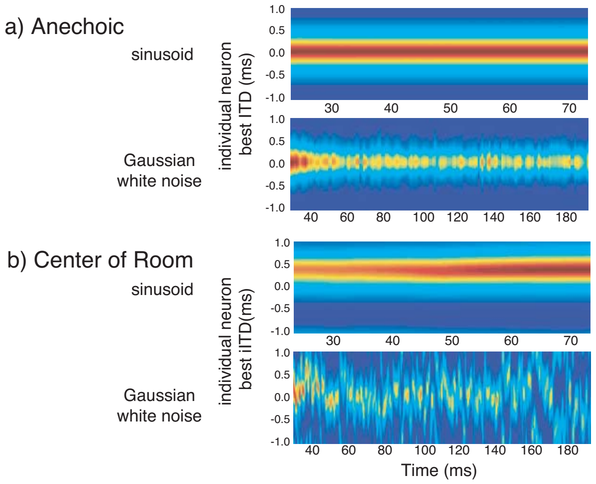

Figure 4: Output of the 500-Hz channel of a neural ITD-processing model as a function of time for tone and noise sources. a) Anechoic results. b) Center of the classroom results.

Fig. 4 plots the model output as a function of time for neurons tuned to 500 Hz. Results are shown for both anechoic (panel a) and the room center (panel b) conditions in response to a 500-Hz tone and Gaussian noise (top and bottom plots in each panel, respectively) presented from (0˚, 1 m). The vertical axis represents the ITD of the cross-correlation function. Strong activity signifies that the neurons tuned to the corresponding ITD fire robustly to the input signals.

For the 0˚ source simulated in Fig. 4, the “true” ITD is 0. For anechoic simulations (Fig. 4a), the neurons tuned to 0 (vertical center of each panel) respond strongly to the input with relatively constant output over time (although the across-time variability is larger for the noise than for the tone; compare top and bottom panels). For the listener in the room center (Fig. 4b), the results depend strongly on the source characteristics. For a tone burst (top panel of Fig. 4b), the activity is constant over time; however, the peak activity is displaced relative to the peak in the anechoic condition. In the case of Gaussian noise (bottom panel of Fig. 4b), peak activity varies dramatically over time, but the expected value of the peak is essentially unchanged compared to anechoic results.

Analysis of the cross-correlation model output determined how accurately source azimuth could be computed from the model population response [20]. Results show that by integrating the model activity over time in an optimal manner, the true azimuth of the source can be extracted, despite the distortion of the ITD information caused by reverberant energy. Such analysis further demonstrates that for a listener in a room, such across-time integration is critical for achieving localization accuracy, whereas such integration is not particularly helpful in anechoic conditions where optimal performance is already very good even for short stimuli.

3. Summary

For normal-hearing listeners, everyday reverberant energy may only modestly degrade directional hearing accuracy, speech intelligibility, and performance on other tasks. In a 3-D virtual auditory display, echoes and reverberation can provide robust cues for source distance, improve the perceived realism of the display, and increase the externalization of simulated sources. Further behavioral and theoretical work is necessary to delineate the conditions in which reverberation truly harms performance and those in which it causes little degradation. Overall, the results of acoustical, behavioral, and computational studies of the effects of reverberation suggest that reverberation increases response uncertainty (e.g., in localization tasks) by increasing the variability in acoustic cues as a function of time. However, listeners may be able to compensate for any destructive effects of such variability if they have 1) sufficient time to adjust how they process and interpret the acoustic signals reaching the ears to compensate for the distortion due to reverberation (perhaps by averaging out random fluctuations in acoustic cues through temporal integration) and 2) sufficient experience with the reverberation to realize that such compensation is necessary. The decision of whether or not to include reverberation in a 3-D sound reproduction system should be made only after carefully considering the kinds of stimuli, environments, and configurations that the system will normally be used to simulate and the tasks that a listener is likely to perform when using the system.

4. Acknowledgements

This work was supported in part by grants from AFOSR, the Sloan Foundation, and NIH. S. Constant, S. Devore, K. Kawakyu, and N. Kopco helped gather the results reviewed here.

5. References

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

[2] J. S. Bradley and G. A. Soulodre, "The influence of late arriving energy on spatial impression," *J. Acoust. Soc. Am.*,. 97, pp. 2263-2271, 1995.

[3] J. S. Bradley, H. Sato, and M. Picard, "On the importance of early reflections for speech in rooms," *J. Acoust. Soc. Am.*, vol. 113, pp. 3233-3244, 2003.

[4] 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.

[5] J. S. Bradley, R. D. Reich, and S. G. Norcross, "On the combined effects of signal-to-noise ratio and room acoustics on speech intelligibility," *J. Acoust. Soc. Am.*,. 106, pp. 1820-1828, 1999.

[6] A. K. Nablek, T. R. Letowski, and F. M. Tucker, "Reverberant overlap- and self-masking in consonant identification," *J. Acoust. Soc. Am.*, vol. 86, pp. 1259-1265, 1989.

[7] W. M. Hartmann, "Localization of sound in rooms," *J. Acoust. Soc. Am.*, vol. 74, pp. 1380-1391, 1983.

[8] P. M. Zurek, "Binaural advantages and directional effects in speech intelligibility," in *Acoustical Factors Affecting Hearing Aid Performance*, G. Studebaker and I. Hochberg, Eds. Boston, MA: College-Hill Press, 1993.

[9] R. Plomp, "Binaural and monaural speech intelligibility of connected discourse in reverberation as a function of azimuth of a single competing sound source (speech or noise)," *Acustica*, vol. 34, pp. 200-211, 1976.

[10] B. G. Shinn-Cunningham, N. Kopco, and T. J. Martin, "Acoustic spatial cues contained in binaural room impulse responses from a classroom," *J. Acoust. Soc. Am.*, submitted.

[11] B. G. Shinn-Cunningham, "Distance cues for virtual auditory space," *Proc. IEEE-PCM 2000*, Sydney, Australia, pp. 227-230, 2000.

[12] B. G. Shinn-Cunningham, "Learning reverberation: Implications for spatial auditory displays," *Proc. Int. Conf. Auditory Displays*, Atlanta, GA, pp. 126-134, 2000.

[13] N. Kopco and B. G. Shinn-Cunningham, "Auditory localization in rooms: Acoustic analysis and behavior," *Proc. 32nd Int. Acoust. Conference EAA Symposium*, Zvolen, Slovakia, pp. 109-112, 2002.

[14] M. Schoolmaster, N. Kopco, and B. G. Shinn-Cunningham, "Effects of reverberation and experience on distance perception in simulated environments," *J. Acoust. Soc. Am.,* vol. 113, pp. 2285, 2003.

[15] B. G. Shinn-Cunningham, "Speech intelligibility, spatial unmasking, and realism in reverberant spatial auditory displays," *Proc. Int. Conf. Auditory Displays* Atlanta, GA, pp. 183-186, 2002.

[16] S. Devore and B. G. Shinn-Cunningham, "Perceptual consequences of including reverberation in spatial auditory displays," *Proc. Int. Conf. Auditory Displays,* pp. 75-78, 2003.

[17] A. J. Watkins, "Perceptual compensation for effects of reverberation on amplitude-envelope cues to the 'slay'-'splay' distinction," *Proc. Inst. Acoust.*, vol. 14, pp. 125-132, 1992.

[18] B. G. Shinn-Cunningham and S. Ram, "Identifying where you are in a room: Sensitivity to room acoustics," *Proc. Int. Conf. Auditory Displays,* pp. 21-24, 2003.

[19] Z. Zhang, M. G. Heinz, I. C. Bruce, and L. H. Carney, "A phenomenological model for the responses of auditory-nerve fibers: I. Nonlinear tuning with compression and suppression," *J. Acoust. Soc. Am.,* vol. 109, pp. 648-670, 2001.

[20] B. Shinn-Cunningham and K. Kawakyu, "Neural representation of source direction in reverberant space," *Proc. IEEE-WASPAA*, New Pfalz, New York, pp. 79-82, 2003.