In Usability Evaluation and Interface Design: Cognitive Engineering, Intelligent Agents and Virtual Reality, MJ Smith, G Salvendy, D Harris, & RJ Koubek (eds.), (Proc HCI International 2001, New Orleans, 5-10 August), NJ: Erlbaum, 604-608.

[illegible]

Barbara Shinn-Cunningham

Boston University, Depts. of Cognitive and Neural Systems and Biomedical Engineering
677 Beacon St., Boston, MA 02215

[illegible]

In order to create a three-dimensional virtual auditory display, both source direction and source distance must be simulated accurately. Echoes and reverberation provide the most robust cue for source distance and also improve the subjective realism of the display. However, including reverberation in a virtual auditory display can have other important consequences: reducing directional localization accuracy, degrading speech intelligibility, and adding to the computational complexity of the display. While including an accurate room model in a virtual auditory display is important for generating realistic, three-dimensional auditory percepts, the level of detail required in such models is not well understood. This paper reviews the acoustic and perceptual consequences of reverberation in order to elucidate the tradeoffs inherent in including reverberation in a virtual auditory environment.

[illegible]

The main feature distinguishing virtual auditory displays from conventional displays is their ability to simulate

In Usability Evaluation and Interface Design: Cognitive Engineering, Intelligent Agents and Virtual Reality, MJ Smith, G Salvendy, D Harris, & RJ Koubek (eds.), (Proc HCI International 2001, New Orleans, 5-10 August), NJ: Erlbaum, 604-608.

at a particular location (relative to the listener) in a room. For sources in anechoic space, these impulse responses are called Head-Related Impulse Responses (HRIRs; in the frequency domain, the filters are called Head-Related Transfer Functions or HRTFs; e.g., see Wightman & Kistler, 1989; Wenzel, 1992; Carlile, 1996). For sources in a room, these impulse responses are the summation of the anechoic HRIR for a source at the corresponding position relative to the listener and later-arriving reverberant energy. In order to illustrate these effects, measurements of the impulse response describing the signals reaching the ears of a listener in the center of an ordinary 18 x10 x12 conference room are shown in the following figures. The sample room is moderately reverberant; when an impulse is played in the room, it takes approximately 450 ms for the energy to drop by 60 dB. While the room in question is not atypical, the listener is always positioned in the center of the room, far from any reflective surfaces, and the sources are at a maximum distance of one meter for all of the measurements shown. These results show what occurs for moderate levels of reverberation. The effects would be much greater for more distant sources, more reverberant rooms, or even different listener positions in the same room.

Figure 1 shows different portions of a time-domain impulse response for the right ear when a source is at directly to the right at a distance of 1 m. The top panel shows a five-ms-long segment containing the direct-sound response (the anechoic HRIR); the central portion shows both the direct sound and some of the early reflections (the first 120 ms of the impulse response); the bottom-most portion shows a close-up (multiplying the y-axis by a factor of 100) of the first 300 ms of the impulse response. Figure 1 illustrates that the reverberation consists both of discrete, early echoes (note the discrete impulses in the middle panel of Figure 1 at times near 11, 17, 38 ms, etc.) and an exponentially-decaying reverberant portion (note the envelope of the impulse response in the bottom of Figure 1).

side (Figure 3a) and straight ahead (Figure 3b) of a listener in the sample room. Transfer functions are shown for both near (15 cm) and relatively distant (1 m) source positions. It is clear that for a source to the right, the energy at the right ear is greater than that at the left (Figure 2a); similarly, the HRTFs for near sources have more energy than those for far sources (compare top and bottom rows in Figures 3a and 3b). As a result, the effect of reverberation varies dramatically with source position. For a near source at 90˚ azimuth, the right ear reverberant transfer function is essentially identical to the corresponding anechoic HRTF (top right panel, Figure 2a). However, the effect of reverberation on the far (left) ear is quite large for a source at 90˚ azimuth (left column, Figure 3a).

In anechoic space, HRIRs depend only on the direction and distance of the source relative to the listener. In contrast, nearly every aspect of the reverberant energy varies not only with the position of the source relative to the listener, but also with the position of the listener in the room. The effects of reverberation shown in Figures 1-3 arise when a listener is located in the center of a large room, far from any walls. In such situations, the most obvious effect of reverberation is the introduction of frequency-to-frequency variations in the magnitude (and phase) transfer function compared to the anechoic case. For the far ear, there is also a secondary effect in which notches in the source spectrum are filled by the reverberant energy (e.g., see the notches in the left and right ear anechoic spectra for a 1-m source, bottom row of Figure 3b). However, different effects arise for different listener positions in the room.

We find that in addition to adding frequency-to-frequency fluctuations to the spectral content reaching the ears, reverberation can lead to pronounced comb-filtering effects (non-random deviations in the long-term spectrum as a function of frequency) when a listener is close to a wall (Brown, 2000; Kopco & Shinn-Cunningham, 2001). These effects cause larger distortion of the basic cues underlying spatial perception (spectral shape, ITD, and ILD) than those that arise when a listener is relatively far from any large reflective surface. In particular, strong, early reflections can lead to dramatic nulls and peaks in the magnitude spectra, rapid shifts in the phase spectra as a function of frequency, and concomitant distortions of interaural differences. Finally, it is clear that essentially every measurable effect of reverberation depends on the relative energy in the direct and reverberant portions of the HRTFs, which depends on the position of the source relative to the listener, the position of the listener in the room, and the room itself.

[illegible]

Reverberation leads to clear physical effects on the signals reaching the ears. However, when considering whether or how to incorporate reverberation in a virtual environment, the critical question is how reverberation influences perception and performance on different tasks of interest.

Reverberation dramatically improves the subjective realism of virtual auditory displays (e.g., see Durlach, listening to headphone simulations using individualized HRTFs do not accurately perceive source distance despite the large changes in ILDs present in their anechoic HRTFs, but do extremely well at judging distance when presented with reverberant simulations (Shinn-Cunningham, Santarelli & Kopco, 2000a). Further, in reverberant simulations, changes in distance are still perceived accurately for monaural presentations of lateral sources (turning off the far-ear signal), suggesting that the cue provided by reverberation is essentially monaural (Shinn-Cunningham et al., 2000a). While much work remains to determine how source distance is computed from reverberant signals, these results and results from other studies (e.g., Zahorik, Kistler & Wightman, 1994) suggest that simplified simulations of room effects may provide accurate distance information.

Results of previous studies of “the precedence effect,” in which directional perception is dominated by the location of an initial source (i.e., the direct sound) and influenced only slightly be later-arriving energy (see Litovsky, Colburn, Yost & Guzman, 1999), suggest that reverberation should have small effect on directional localization accuracy. However, few studies have quantified how realistic room reverberation affects directional hearing. We find that reverberation causes very consistent, albeit small, degradations in directional accuracy compared to performance in anechoic space (Shinn-Cunningham, 2000b). Further, localization accuracy depends on the listener position in a room (Kopco & Shinn-Cunningham, 2001). When a listener is near a wall or in the corner of the room, response variability is greater than when in the center of the room. Based on analysis of the room acoustics, these results are easy to understand; reverberation distorts the basic acoustic cues that convey source direction, and this distortion is greatest when a listener is near a wall (Brown, 2000; Kopco & Shinn-Cunningham, 2001). We also observe that, over time, directional accuracy improves in a reverberant room and (after hours of practice) approaches the accuracy seen in anechoic settings (Santarelli et al., 1999a; Shinn-Cunningham, 2000b). Figure 4 shows this learning for an experiment in which listeners judged the postion of real sources in the room in which reverberant impulse responses were measured. In the figure, the mean left/right localization error (computed based on the difference in ITD caused by a source at the true and response positions, in ms) is shown. The error was computed both for the initial 100 trials (after 200 practice trials in the room, to accustom the listener to the task), and for the final 100 trials of a 1000-trial-long experiment. For each subject, error decreased by the end of the 1000 trials. These results suggest that any detrimental effects of reverberation on directional localization, which are relatively minor at worst, disappear with sufficient training.

Finally, reverberation can interfere with the ability to understand or analyze the content of acoustic sources in the environment (e.g., see Nomura, Miyata & Houtgast, 1991). For instance, one of the most important acoustic signals that humans encounter is speech and much of the information in speech signals is conveyed by amplitude modulations. However, as shown in Figure 2, these modulations are reduced by reverberation. Although moderate amounts of reverberation do not degrade speech intelligibility severely, reverberation can degrade intelligibility. In addition, it is likely that reverberation will degrade signal intelligibility even more when there are competing signals than it will in quiet. Specifically, reverberation decorrelates the signals at the two ears and tends to reduce differences in the level of a signal reaching the two ears. Both of these factors can improve signal intelligibility in the presence of an interfering sound (Zurek, 1993). Thus, we predict that reverberation will have a particularly adverse impact on speech intelligibility in the presence of a masking source, a hypothesis we are currently exploring.

[illegible]

The computational complexity in simulating realistic reverberant signals is prohibitive. In order to allow real-time, interactive environments to be simulated, many current virtual auditory environments do not include any reverberation. Those that include reverberation often use algorithms that simplify the computations (e.g., by accurately simulating the directional information only for a small number of discrete echoes, and generating decorrelated noise to simulate later-arriving reverberation). The perceptual consequences of these computational simplifications are not well understood. More research is needed to quantify how sensitive the human listener is to these simplifications and to determine how they influence the subjective realism of the display, the ability to judge source distance accurately, and the ability to learn to accurately judge source direction with training.

There are inherent tradeoffs in including reverberation in a virtual environment. As with any complex design decision, the appropriate choice (of whether to include reverberation, how accurately to simulate room acoustics, etc.) depends upon the goals of the display. If the main goal of the auditory display is to provide speech input to the listener, it may be best to exclude any reverberation. If the goal is to provide distance information about arbitrary sound sources, including some form of reverberation is critical; however, one may be able to provide distance

(Santarelli et al., 1999a), it may be possible to reduce the energy of simulated echoes and reflections and still