

Creating three dimensions in virtual auditory displays

Barbara Shinn-Cunningham

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

ABSTRACT

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.

1. SOUND LOCALIZATION CUES

The main feature distinguishing virtual auditory displays from conventional displays is their ability to simulate the location of an acoustic source. In this section, the basic spatial auditory cues that convey source position are reviewed in order to gain insight into how reverberation influences spatial auditory perception.

The physical cues that determine perceived sound direction have been studied extensively for over a century (for reviews, see Middlebrooks & Green, 1991; Gilkey & Anderson, 1997). Most of these studies were performed in carefully controlled conditions with no echoes or reverberation and focused on directional perception. Results of these anechoic studies identified the main cues that govern directional perception, including differences in the time the sound arrives at the two ears (interaural time differences or ITDs), differences in the level of the sound at the two ears (interaural level differences or ILDs), and spectral shape. ITDs and ILDs vary with the laterality (left/right position) of the source, whereas spectral shape determines the remaining directional dimension (i.e., front/back and left/right; e.g., see Middlebrooks, 1997).

In contrast with directional localization, relatively little is known about how listeners compute source distance. In the absence of reverberation¹, overall level can provide relative distance information (Mershon & King, 1975). However, unless the source is familiar, listeners cannot use overall level to determine the absolute distance (Brungart, 2000). For sources that are within a meter of the listener, ILDs vary with distance as well as direction (Duda & Martens, 1997; Brungart & Rabinowitz, 1999; Shinn-Cunningham, Santarelli & Kopco, 2000b), and appear to help listeners judge distance in anechoic space (Brungart, 1999; Brungart & Durlach, 1999). However, ILD cues are not useful for conveying distance information unless a source is both off the mid-sagittal plane and within a meter of the listener (Shinn-Cunningham et al., 2000b).

The most reliable cue for determining the distance of an unfamiliar source appears to depend upon the presence of reverberation (Mershon & King, 1975; see also Shinn-Cunningham, 2000a). While the direct sound level varies inversely with distance, the energy due to reverberation is roughly independent of distance. Thus, to a first-order approximation, the direct-to-reverberant energy ratio varies with the distance of a source (Bronkhorst & Houtgast, 1999). While many studies show the importance of reverberation for distance perception, there is no adequate model describing how the brain computes source distance from the reverberant signals reaching the ears. Of course, in real environments, reverberation does not just improve distance perception; it influences other aspects of performance as well. The remainder of this paper explores acoustic and perceptual effects of realistic reverberation and discusses the tradeoffs to consider when adding reverberation in virtual auditory environments.

2. ACOUSTIC EFFECTS OF REVERBERATION

Reverberation has a dramatic effect on the signals reaching the ears. Many of these effects are best illustrated by considering the impulse response that describes the signal reaching the ear of the listener when an impulse is played

This work was supported by the AFOSR and the Alfred P. Sloan Foundation. N. Kopco and S. Santarelli helped with data collection.

¹For simplicity, the term "reverberation" is used throughout this paper to refer to both early, discrete echoes and later reflections. In contrast, in much of the literature, "reverberation" refers only to late arriving energy that is the sum of many discrete echoes from all directions (and is essentially diffuse and uncorrelated at the two ears).

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 x 10 x 12 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).

To compute the total signal at the ears for an arbitrary signal, the impulse response must be convolved with the signal emitted by the distal source (Wightman & Kistler, 1989; Wenzel, 1992; Carlile, 1996). This process can distort and temporally smear the signal reaching the ears. Figure 2 shows the time-domain waveform for a recorded speech utterance (the word “bounce”) in its raw form (Figure 2c) and processed through impulse responses to recreate the total signal that would reach the listener’s left and right ears (Figures 2a and 2b, respectively) for a source at a distance of 1 m and directly to the right. In the figure, the black lines plot the anechoic HRIR and the gray lines show the reverberant impulse response for the same source position relative to the listener. To ease comparisons, the waveforms are scaled (normalized) so that the maximum amplitude is 1.0 in the anechoic signals. In anechoic space, the envelope of the waveform reaching the ears is fairly similar to that of the original waveform. Of course, the HRIR processing does cause some spectral changes. For instance, for the right-ear signal (Figure 2b), the sibilant at the end of the utterance (the second energy burst in the unprocessed waveform; i.e., “boun-CE”) is emphasized relative to the initial energy burst because high frequencies are boosted by the right-ear’s HRIR for this source position. Nonetheless, the general structure of the waveform is preserved. Since the direct sound energy to the near (right) ear is much greater than for the far (left) ear, the effect of reverberation is much more pronounced for the left ear signal. For the left ear, the waveform envelope is smeared in time and the modulations in the waveform are reduced (Figure 2a); in the right ear, the waveform envelope is well preserved (Figure 2b). While these graphs only show the distortion of the total waveform envelope, similar modulation distortion occurs for narrow band energy as well (i.e., such as the representation in an auditory nerve).

Figure 3 shows the anechoic and reverberant HRTFs at the left and right ears (left and right columns) for sources both to the right

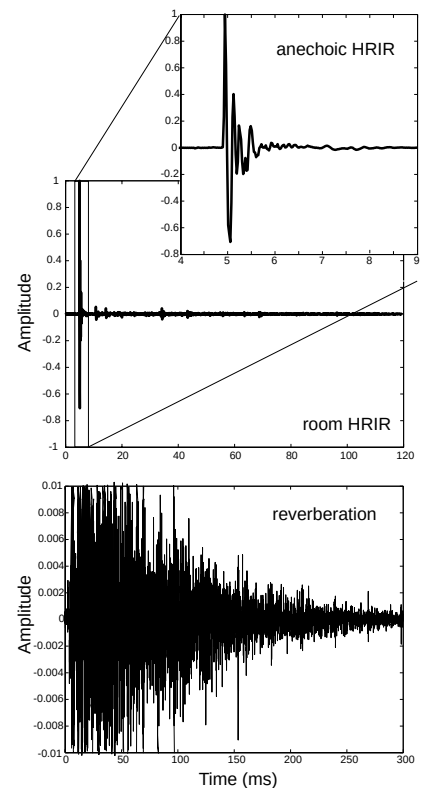


Figure 1: Impulse response to a listener’s right ear for a source at 1 m, 90° azimuth in the horizontal plane in a reverberant room.

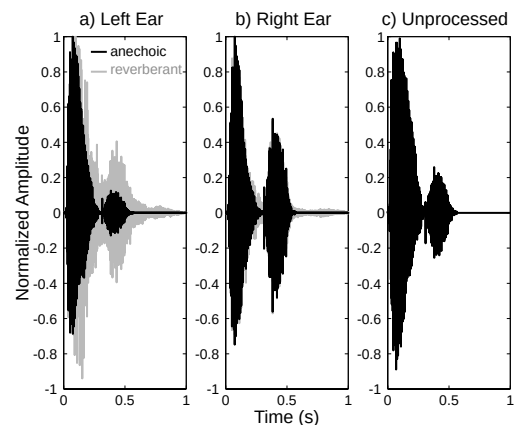


Figure 2: Speech waveform at the ears (anechoic, black, or reverberant, gray) for a source in the horizontal plane (at 1 m, 90° azimuth) and unprocessed waveform.

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.

3. PERCEPTUAL EFFECTS OF REVERBERATION

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, Rigapulos, Pang, Woods, Kulkarni, Colburn & Wenzel, 1992). In many auditory displays that do not include reverberation, sources are often heard in roughly the correct direction, but near or even inside the head. In such anechoic simulations, painstaking care to provide simulations tailored to the individual listener and to compensate for characteristics of the headphone delivery system can ameliorate this “lack of externalization” (e.g., see Wenzel, Arruda, Kistler & Wightman, 1993; Pralong & Carlile, 1996). However, in contrast, good externalization is usually obtained when subjects listen to recordings made from a head in a reverberant setting, even without compensating properly for the headphone characteristic and when the playback is not tailored to the individual listener. Reverberation also provides information about the characteristics of the space itself, conveying information about the room size (e.g., see Bradley & Soulodre, 1995). While realism and environmental awareness are dramatically increased by reverberation, the extent to which these benefits depend on the fine structure of the reverberation has not been quantified. In other words, it may be possible to provide simplified reverberation cues that are less costly to include in a virtual environment but which still convey this information to the listener.

As noted above, distance perception is dramatically improved with the addition of reverberation (Mereshon & King, 1975). We find that even when sources are within a meter of the head, where the relative effects of reverberation are small and robust ILDs should provide distance information, subjects are much more accurate at judging source distance in a room than in anechoic space (Santarelli, Kopco & Shinn-Cunningham, 1999a; Santarelli, Kopco, Shinn-Cunningham & Brungart, 1999b; Shinn-Cunningham, 2000b). In addition, subjects

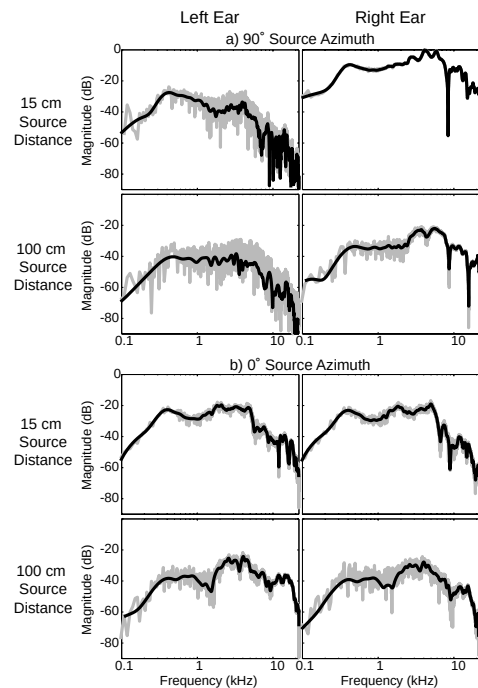


Figure 3: Magnitude spectra of anechoic (black) and reverberant (gray) transfer functions at the two ears for different source positions in a room.

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 by 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 position 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.

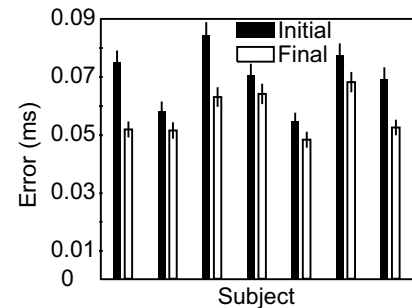


Figure 4: Mean left/right errors at the beginning and end of an experiment run in a reverberant room for each of seven subjects.

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.

4. DISCUSSION

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 information using a very simplified algorithm for generating “reverberation.” Along a similar vein, because even moderate levels of reverberation can provide useful distance information and increase the realism of a display

(Santarelli et al., 1999a), it may be possible to reduce the energy of simulated echoes and reflections and still provide the perceptual benefits of reverberation while limiting its destructive influence (e.g., on directional localization or on speech intelligibility). Of course, recent work suggests that subjects are sensitive to the agreement of visual and auditory environmental cues (see the article by Gilkey, Simpson, and Weisenberger in this volume); thus, if the main goal is to produce a virtual environment that is subjectively realistic and produces the sense that the listener is truly “present” in the virtual world, the simulation should include the most realistic model of room acoustics that can be integrated into the system.

REFERENCES

- Bradley, JS and GA Soulodre (1995). “The influence of late arriving energy on spatial impression.” *J Acoust Soc Am*, **97**, 2263-2271.
- Bronkhorst, AW and T Houtgast (1999). “Auditory distance perception in rooms.” *Nature*, **397**, 517-520.
- Brown, TJ (2000). Characterization of Acoustic Head-Related Transfer Functions for Nearby Sources. Electrical Engineering and Computer Science. Cambridge, MA, Massachusetts Institute of Technology.
- Brungart, DS (1999). “Auditory localization of nearby sources III: Stimulus.” *J Acoust Soc Am*, **106**, 3589-3602.
- Brungart, DS (2000). A speech-based auditory distance display. 109th Convention Audio Eng Soc, Los Angeles.
- Brungart, DS and NI Durlach (1999). “Auditory localization of nearby sources II: Localization of a broadband source in the near field.” *J Acoust Soc Am*, **106**, 1956-1968.
- Brungart, DS and WM Rabinowitz (1999). “Auditory localization of nearby sources I: Head-related transfer functions.” *J Acoust Soc Am*, **106**, 1465-1479.
- Carlile, S (1996). *Virtual Auditory Space: Generation and Applications*. New York, RG Landes.
- Duda, RO and WL Martens (1997). Range-dependence of the HRTF for a spherical head. IEEE ASSP Workshop on Applications of Digital Signal Processing to Audio and Acoustics.
- Durlach, NI, A Rigapulos, XD Pang, WS Woods, A Kulkarni, HS Colburn and EM Wenzel (1992). “On the externalization of auditory images.” *Presence*, **1**, 251-257.
- Gilkey, R and T Anderson (1997). *Binaural and Spatial Hearing in Real and Virtual Environments*. Hillsdale, New Jersey, Lawrence Erlbaum Associates, Inc.
- Kopco, N and BG Shinn-Cunningham (2001). Effect of listener location on localization cues and localization performance in a reverberant room. 24th mid-Winter meeting Assoc Res Otolaryng, St. Petersburg Beach, FL.
- Litovsky, RY, HS Colburn, WA Yost and SJ Guzman (1999). “The precedence effect.” *J Acoust Soc Am*, **106**, 1633-1654.
- Mershon, DH and LE King (1975). “Intensity and reverberation as factors in auditory perception of egocentric distance.” *Percept Psychophys*, **18**, 409-415.
- Middlebrooks, JC (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, JC and DM Green (1991). “Sound localization by human listeners.” *Ann Rev Psych*, **42**, 135-159.
- Nomura, H, H Miyata and T Houtgast (1991). “Speech-intelligibility and subjective MTF under diotic and dichotic listening conditions in reverberant sound fields.” *Acustica*, **73**, 200-207.
- Pralong, D and S Carlile (1996). “The role of individualized headphone calibration for the generation of high fidelity virtual auditory space.” *Jornal of the Acoustical Society of America*, **100**, 3785-3793.
- Santarelli, S, N Kopco and BG Shinn-Cunningham (1999a). Localization of near-field sources in a reverberant room. 22nd mid-Winter meeting Assoc Res Otolaryng, St. Petersburg Beach, FL.
- Santarelli, S, N Kopco, BG Shinn-Cunningham and DS Brungart (1999b). “Near-field localization in echoic rooms.” *J Acoust Soc Am*, **105**, 1024.
- Shinn-Cunningham, BG (2000a). Distance cues for virtual auditory space. Proceedings of the IEEE-PCM 2000, Sydney, Australia.
- Shinn-Cunningham, BG (2000b). Learning reverberation: Implications for spatial auditory displays. International Conference on Auditory Displays, Atlanta, GA.
- Shinn-Cunningham, BG, S Santarelli and N Kopco (2000a). Distance perception of nearby sources in reverberant and anechoic listening conditions: Binaural vs. monaural cues. 23rd mid-Winter meeting Assoc Res Otolaryng, St. Petersburg Beach, FL.
- Shinn-Cunningham, BG, S Santarelli and N Kopco (2000b). “Tori of confusion: Binaural localization cues for sources within reach of a listener.” *J Acoust Soc Am*, **107**, 1627-1636.
- Wenzel, EM (1992). “Localization in virtual acoustic displays.” *Presence*, **1**, 80-107.
- Wenzel, EM, M Arruda, DJ Kistler and FL Wightman (1993). “Localization using nonindividualized head-related transfer functions.” *J Acoust Soc Am*, **94**, 111-123.
- Wightman, FL and DJ Kistler (1989). “Headphone simulation of free-field listening. I. Stimulus synthesis.” *J Acoust Soc Am*, **85**, 858-867.
- Zahorik, P, DJ Kistler and FL Wightman (1994). Sound localization in varying virtual acoustic environments. Second International Conference on Auditory Display, Santa Fe, NM, Santa Fe Institute.
- Zurek, PM (1993). Binaural advantages and directional effects in speech intelligibility. *Acoustical Factors Affecting Hearing Aid Performance*. G. Studebaker and I. Hochberg. Boston, MA, College-Hill Press.