Binaural Auralization and Perceptual Veridicality

3421 (M-3)

Durand R. Begault
NASA-Ames Research Center
Aerospace Human Factors Research Division
Moffett Field, CA 94035, USA

Presented at
the 93rd Convention
1992 October 1–4
San Francisco

AES

AN AUDIO ENGINEERING SOCIETY PREPRINT

BINAURAL AURALIZATION AND PERCEPTUAL VERIDICALITY

Durand R. Begault
Aerospace Human Factors Research Division
Mail Stop 262-2
NASA-Ames Research Center
Moffett Field, CA 94035

ABSTRACT

Early reflection patterns calculated from a room design program were filtered with measured HRTFs within a "hypothetical auralization system". With reference to a particular set of room/listener/sound source configurations, the system is shown to produce both perceptible and imperceptible results. The results of this exploratory analysis are illuminating for future design considerations of HRTF-based auralization systems.

SUBJECT CLASSIFICATIONS: psychoacoustics; 3-D sound ; binaural sound; auralization systems; virtual acoustic displays

0. Introduction

Last year, at the 91st Audio Engineering Society Convention in New York, a set of papers were presented during the "Auralization" session that shared the common aim of relating a physical model, in this case, room models, to aural perception (see [1,2] for an overview). The ideal function of an auralization system is to allow an arbitrary sound input to be processed so that a listener may experience it in real time in a variety of room/listener/sound source configurations. An important difference between the current approaches that are called "auralization" systems and computer methods introduced by Schroeder and Atal at Bell Laboratories about thirty years ago is the inclusion of "binaural synthesis" or "3-D sound" techniques, and real time simulation (see [3,4,5] for examples). Rather than capturing the binaural transfer function of actual concert halls from a dummy head, as done by Schroeder, Gottlob and Siebrasse in the 1970s [6], auralization involves measuring the binaural "head related transfer function" (HRTF) separately. These transfer functions are then convolved with the results of a computer model of reverberation, a technique that was done "off-line" in the 1970s -1980s before the ready availability of real-time digital signal processing chips [7,8,9,10].

The following reports on some perceptual anomalies found with a hypothetical auralization system that uses binaural synthesis for simulating early reflections. The room model used with the system presented here is very simple, and could not possibly reflect the level of complexity useful to the end users of auralization systems. In fact, the only room that this model could represent is an anechoic chamber with 7 speakers. Nevertheless, the models were chosen as a platform for illustrating some basic problems with binaural synthesis of early reflections. Contrasting the specificity of the room model, the binaural synthesis techniques described below were of very high fidelity, and have been used previously in several basic research studies into the perception of "virtual" audio sources.

A ray tracing program was used to calculate the angle of incidence, intensity and attenuation of a direct sound path and 6 early reflection paths, one from each interior surface of a modeled room. This data was used to calculate an appropriate binaural impulse response. Listener position was varied in several implementations of the room-listener model, and in one case the room size was varied. These iterations represent typical interface manipulations of an auralization system. An acoustician, for example, would iteratively examine a design solution by varying parameters such as listener position, building materials, or enclosure size. A real time auralization system could allow immediate comparative "A-B" evaluation between these variations over headphones.

The following is a report on some predicted perceptual results of this simple auralization system model, along with some informal listening tests. First, a hypothetical auralization system is presented, for the purpose of illustrating how the essential features of a more complex, "physically matched" system would interface with psychoacoustic results. Second, the details of four room/sound source/listener configurations used in this study are presented, along with the details of how the hypothetical auralization system was actually realized. Finally, the relationship between this hypothetical system and perceptual data, including some pilot listening evaluations, are presented.

1. Hypothetical, real-time auralization system

Figure 1 illustrates the components of a hypothetical real-time auralization system used for designing the off-line signal processing and room modeling of the current investigation. This emulates the essential features of many auralization and 3-D sound systems. Alternative implementations of auralization systems not considered here are loudspeaker based systems that use cross talk cancellation ("transaural" processing) [11], and the use of head-tracking devices that create virtual acoustic images independent of head position [12].

A commercially available, interactive graphic-based room modeling program is used as the user interface for driving "translation" software that converts graphic display parameters into signal processing parameters. Typical room modeling programs allow specification of sound source and listener orientations and positions within an enclosure. The enclosure itself can be altered by the user to change its dimensions, complexity, and materials. The sound source's early reflection response is traced to the listener position by means of ray tracing or image models methods. The translation software is used to calculate the time delay, attenuation, and filtering parameters of delayed "copies" of the input sound. An important feature is the translation software's capacity to calculate the angle of incidence of the reflections to the listener, and use this information to select an appropriate binaural filter for signal processing.

The use of a ray tracing or image model presupposes the fact that early reflections behave in a manner akin to light, i.e., in a specular manner. In actuality, an impending waveform will scatter energy in many directions in a diffuse manner, especially at lower frequencies. Life would be easy if it were possible to determine a "cut-off" point between specular and diffuse reflection behavior for a signal processing model,: many engineers, including the author, have had to make simplifying assumptions in this regard. Another inherent difficulty is modeling the transfer function of a wall surface. A common procedure is to model a wall as a filter, which indeed it is. One can obtain building material characteristics and have an auralization system adjust filtering parameters to match these specifications. Unfortunately, especially in the case of ray tracing, the published specifications are appropriate for a single angle of incidence. The transfer function of building materials can vary substantially as a function of angle of incidence of the incident waveform [13].

The translation software can be used to load appropriate parameters to real-time digital signal processing chips such as the Motorola 56001. In the hypothetical system shown here, a separate stage of processing is used for each reflection. The bottom of Figure one shows signal modification for attenuation and time delay based on the path length of each ray, a frequency dependent attenuation to represent the absorptive characteristics of the reflective boundary, and a binaural filter containing FIR filter implementations of the HRTF for a given angle and elevation of incidence. The output of each stage is then mixed and multiplexed for output to stereo headphones.

2. Present realization of the hypothetical system; room models

Figures 2 and 3 show the room/listener/sound source specifications referred to in the discussions below. Figure 2 shows a small room with dimensions of 18 x 13 x 9.5, a typical size for an office or small control room, with listener positions 1, 2 and 3 (the sound source position is labeled **A**). Figure 3 shows a large room with dimensions of 85 x 40 x 20, with listener position 4. A room modeling program (Bose Corporation's Modeler+, version 4.0) was used to calculate the angle of incidence, intensity and attenuation of a direct sound and 6 early reflections. The sound source was modeled in both rooms as an omnidirectional impulse generator; a summary of the wall materials and other specifications to the Modeler+ program are given within Figures 2 and 3.

Figure 4 shows examples of the ray tracing produced by this program. A direct path reaches the listener directly, while six separate "first order" early reflections are ray traced to each interior surface of the enclosure. It should be noted that, for the given models discussed here, the ray tracing and image model algorithms available within this program produce identical results for these six reflections. Each surface was frequency dependent within certain bands; all calculations made within this paper are based on data for the 1 kHz center frequency band.

This data was used to calculate a binaural impulse response for subsequent digital filtering in the signal processing stage. The time delay, overall intensity adjustment, and binaural filtering of each reflection was accomplished off-line on workstation computers (Apple Mac II si, cx, fx) with signal processing software (Zola Technologies' DSP Designer) and sound recording/playback software and hardware (Digidesign's Sound Tools). The impulse responses used for the binaural filters were obtained from the spatial map source files used in a real-time 3-D audio display device (Crystal River Engineering's Convolvotron). The binaural impulse responses were originally measured by Wightman and Kistler (pinnae of subject SDO) at every 15 ° degrees of azimuth starting at 0 ° azimuth (directly ahead), at six elevation angles : 0 ° (directly ahead of the listener), down 18 and 36 degrees, and up 18, 36, and 54 degrees [14]. Although the Convolvotron has the capability of interpolating between these measured positions [15], the current investigation rounded values obtained from the Modeler+ program to the nearest measured azimuth and elevation point. It is doubtful that the results presented here would differ substantially if a reflection were modeled, for example, at an angle of incidence of 12 instead of 15 degrees azimuth. In addition, the frequency and phase response of the headphones used (Sennheiser HD-430) were divided out of each transfer function.

3. Early reflection thresholds.

In the context of spatial hearing and psychoacoustic investigations, the measurement of thresholds can be a challenging task. Perhaps the most difficult challenge for the experimenter is giving subjects a definition of a particular threshold that will be consistent amongst subjects. The easiest is the absolute or masked threshold, where the listener indicates at what point any qualitative difference is heard with the sound. There are also image shift thresholds, which are usually higher than absolute thresholds, involving perception of sound source displacement. This threshold is often difficult to ascertain since it is easy to confuse image displacement with image broadening. Specifically, it is possible for one listener to feel that an image has simply broadened but maintained its same "center of gravity" as to its location, while another listener may associate the image broadening with spatial displacement. Even more difficult to insure consistency among subjects are thresholds associated with adjectives such as "disturbing" or "annoying".

There is evidence that the absolute threshold for an early reflection changes as a function of angle of incidence, and as a function of the spectral similarity between the direct and reflected sound [16]. These observations are important inasmuch that an auralization system could hypothetically expend valuable computation power to directionalize modeled early reflections that would effectively be imperceptible. The situation is complicated by the fact that thresholds change as functions of the time delay of the reflection and of the source material used.

A study by Olive and Toole showed that, within a single reflection experimental paradigm, a 60 ° vertically displaced reflection or a 65 ° laterally displaced reflection has a threshold around 5 to 10 dB lower than a reflection originating at a direction near the direct sound source [17]. In addition, nearly the same thresholds were observed for the vertical and laterally displaced reflection, although the audible effect was different (spatial qualities changed in the lateral case, and timbral qualities changed in the vertical case). Absolute thresholds for a single reflection as a function of lateral angles between 30 ° and 65 ° were summarized by Olive and Toole (from their own work and comparable data sources) for music, speech and a variety of test signals. For time delays between roughly 2 to 25 msecs (the range most relevant to the models used in the current investigation), the absolute threshold is between -25 and -15 dB below a 0 dB "direct sound" , with outliers at -10 dB (musical stimuli) and -45 dB (pulsed click stimuli).

Investigating the data from the source/listener/room models shown in Figures 2 and 3, it would appear that some reflections would definitely be below threshold. Figure 5 shows the combined binaural magnitude response of the summed HRTF-filtered impulse responses for listener position 4. Because of the size of the enclosure, the path length of the rear wall reflection is relatively long, resulting in an impulse response with a relative amplitude peak of -30 dB. Because of arrival time at around 110 milliseconds, the addition of higher order reflections and "late reverberation" into an auralization implementation would contribute to masking this particular reflection.

Figure 6 shows the dB SPL reflectogram summary given by the room modeling program for listener position 0 (ref. Figure 2), and the resulting binaural impulse response. The reflectogram shows the second and last early reflections to have the lowest relative SPL level, around 20 dB below the direct sound, and the first reflection to be relatively strong, around 10 dB below the direct sound. However, based on the data cited above, a perceptual scale would show that the 1st reflection from the forward wall would be as weak as the 2nd reflection, since the 1st reflection comes from the same direction as the direct sound, and the 2nd one is directionalized from below the listener The addition of obstructing surfaces or more complex geometries would further attenuate these reflections, possibly below threshold.

Of particular relevance to a system using HRTF spatialization of early reflections is the following question: are spatial cues masked- specifically, frequency dependent, interaural amplitude differences? Consider the hypothesis that a reflection must be above the absolute threshold at both ears in order for the spatial hearing cue of interaural level differences to function. The change in these interaural level differences across frequency are a key component of binaural HRTF measurements, inasmuch as they function as a cue to spatial position of a sound source. A conclusion could then be made: if the interaural amplitude differences are masked for a particular reflection, then binaural HRTF convolution of the reflection is a tautological use of computation power. This is not really true because the binaural HRTF also contains interaural time differences that are important to spatial hearing. But computation could be reduced if HRTF spectral shaping were no longer perceptually relevant.

Referring back to Figure 6, the left wall reflection is shown to be 12 dB below the level of the direct sound. Based on the research cited above, the reflection would be above the absolute threshold, to the extent that thresholds for a single reflection can be extended to the current situation with six reflections. But investigation of the relative level at each ear yields a different result for perception of interaural level differences. Figure 7, top, shows the magnitude of the appropriate HRTFs for the reflection, reduced by -12 dB in relationship to the direct sound HRTF. Figure 7, bottom, shows the difference in the magnitude between the direct sound and the reflected sound at the right ear. The graphs suggest that many of the interaural, frequency dependent amplitude differences, i.e. the spectral "peaks" and "valleys", of the left reflection HRTF at the right ear would be masked by the relative strength of the 0 degree direct sound HRTF. Assuming a -25 dB threshold for the reflection, the bottom of Figure 7 shows masking to occur at the right ear between 2 kHz - 5 kHz and 9 kHz- 14 kHz.

It should be pointed out that there are no definite conclusions that can be cited for assessing the perceptual significance of these measured spectral characteristics. Regarding the spectral modification of the HRTF, one study has shown greater sensitivity at spectral peaks in lower frequencies (around 1 kHz) than to troughs at higher frequencies (around 8 kHz) [18], and another recent study has shown that interaural time difference cues of relatively low frequencies (below around 2 kHz) dominate interaural level differences in localization [19].

4. Spatial ambiguity: localization error for free-field and headphone binaural synthesis

Research into the spatial perception of binaurally synthesized stimuli over headphones is a relatively recent area of study. The general conclusion to be made is that the simulation is veridical compared to free-field localization of actual sound sources [20, 21]. However, some trends do seem to be suggested by the existent literature: (1) with long-term training, people localize more accurately than without long-term training [20]; (2) that, although there are exceptions, people localize more accurately with their own HRTFs than with those of selected individuals [21,22]; (3) that for some people, headphone localization using binaural synthesis is worse than free-field localization, especially for elevation [20,21]; (4) that localization accuracy can vary widely between subjects [20,21,23]; (5) that speech is localized less accurately than noise [23]; and finally, (6) that artificial reverberation can degrade localization accuracy [24].

The angles of incidence measured for the reflections for listener positions 0, 2 and 4 in Figures 2 and 3 are summarized in Table I, and Table II summarizes absolute headphone localization error for speech stimuli without reverberation, in terms of "tolerances", at positions near the angles given in Table I. The data in Table II summarizes the mean percentage of judgements for left and mirror-image right positions (e.g., left and right 60 degrees), taken from 11 unexperienced subjects, listening through non-individualized HRTFs [25]. For the positions shown, the azimuth judgement error is greater than 30 degrees in 44% of the subjective judgments that were analyzed. Comparing the perceptual data in Table II with the modeled data for early reflections in Table I, it seems that a perceptual ambiguity would exist, to the extent that the "perceptual identity" of these listener positions within a simulation is revealed by the angle of incidence of the reflection. Specifically, Table I shows only a 30 degree difference between the left and right wall early reflection angles- 45 degrees vs. 75 degrees. This difference is smaller than the size of the "tolerance angle" of > 30 degrees (column 4 of Table I).

Another observation made in studies of HRTF filtered speech stimuli is a subjective tendency towards elevated judgements. Specifically, in one study, the mean value was up 17 degrees for target elevations of 0 degrees at various azimuths [23]; in another study, the mean was up 11 degrees. In the latter study, the mean elevation increased to up 28 degrees when artificial reverberation that included HRTF-filtered reflections was added to the stimuli [24]. With reference to Table I, this data implies that the user of an auralization system might not be able to discriminate between the ceiling and front wall reflections on the basis of spatial perception.

Localization of HRTF-filtered speech can also be investigated in terms of the number of positional "reversals" between the front and rear hemispheres between the front and to the rear of the listener; e.g., hearing a sound at left 120 degrees azimuth instead of a target position of 60 degrees. These reversed judgements are a confounding variable in both free-field and headphone localization investigations . (In fact, if reversed judgements are "corrected", the data shown in Table II shows a much higher degree of localization accuracy; most all judgements would then be within a 0 - 10 degree "tolerance range"). For speech, one study determined the reversal rate to be around 47 % for front-back reversals, and 11 % for back-front reversals [23]. The literature has also shown that the rate and directional trend for reversals is a highly individual matter. But with reference again to Table I, there is strong possibility that the user of an auralization system would not be able to spatially discriminate between the back and front wall reflections.

6. Informal listening tests evaluating HRTF-filtered early reflections

Some informal listening evaluations were conducted, using four to five "expert listeners". In these listening tests, no visual interface could be seen; hence, the judgements were made on the basis of aural cues, and verbal information supplied by the experimenter.The test material consisted of a 10 second portion of male speech with tabla accompaniment, recorded under very dry conditions (Robert Ashley's CD "Perfect Lives: The Park/The Backyard", Lovely Music label). The informality of these tests must be emphasized; only three of several evaluations are reported here.

In the first listening evaluation, the directions of the reflections and direct sound were evaluated for listener position 4. Subjects auditioned the convolution of this test material with three pattern of reflections: The first "facing" the sound source, as derived from the room model; the second, a version with the listener turned 180 degrees; and the third, with a random spatial distribution of reflections. The same timings and amplitudes were used in each case. Subjects could listen to the sound examples as many times as they wanted, and in any order; they were asked to state what differences they heard between the files, in terms of timbre, spatial positioning, or loudness of the source. They were also asked to state anything else they wanted to about the sound they heard.

Because most of the subjects used were "expert listeners", all reported the sensation that the sound source was in some kind of room- i.e., they noticed that reverberation was present, although only early reflections were used. But none of the subjects reported the image switching back and forth from behind to in front of them while listening to the first and second patterns. In fact, it was extremely difficult for anyone to discriminate <u>any</u> difference between these three examples, on <u>any</u> type of basis: timbral, spatial, or loudness. This was verified with some simple two-alternative, forced choice discrimination tests.

A second listening evaluation was conducted which allowed subjects to compare listener position 2 in the small room to listener position 4 in the large room (refer to Figures 2-3 and Table I). Both listener positions were specified as 12 feet from the sound source; the intensity was equalized between the two sound files to eliminate any cues that might be derived from absolute loudness. Normally, the reverberation time resulting from "early" <u>and</u> "late" reflections would be the main cue for simulating relative room size (see the Rt graphs at the bottom of Figures 2 and 3). For the present simulation, the "early reflection time" (the duration of the impulse responses used for the convolution) was about 14.2 msecs long for listener position 2, and about 31 msecs long for listener position 4, not including the low-amplitude, possibly masked reflection at 110 msecs (ref. Figure 5 and discussion in Section 3). All listeners could hear a timbral difference between the two positions, and when asked which room was "bigger", all chose the larger room's impulse response. This suggests that some aspect of the pattern of 6 early reflections, probably the temporal spacing, was sufficient to suggest relative room size.

Some subjects also compared listener positions 2 and 3. Position 3 was the only off-axis position with which stimuli was convolved; the direct sound was at an angle of incidence of right 15 degrees for this situation. These listeners were first asked if their orientation to the sound source was the same or different in the two examples. All detected some type of change in orientation. However, when told to imagine they were facing the sound source, and then asked to indicate where the sound source was in each example by pointing, results were very inconsistent between and even within subjects, over several trials. This can be explained in terms of the effect of image broadening (ref. [24] and the discussion of image shift thresholds above in section 3).

7.Conclusions

What are the real implications of this perceptual data to the design of auralization systems? First, it seems allowable to accept a level of localization error that is present in actual listening, and many subjects that have been tested localize real and virtual sources with nearly equivalent accuracy [21]. Second, the above data is based on experimental paradigms where most of the spatial and environmental information normally available through multiple sensory pathways and orientation-search (e.g., head movement) is missing [26]. Third, auditory localization judgements are highly malleable as a function of expectation or memory; localization errors and reversals are therefore not really a problem, because the auralization system's user "knows" the position since she or he indicated this information via the human interface. For example, in a concert hall design, the user has a cognitive, visual picture of seats facing the sound source. Finally, the perceptual importance of early reflections probably lies more in their net effect on the perceived intelligibility and timbre of the direct sound than in the ability to sense their direction. Nevertheless, the spatial distribution of reflections is perceptually relevant , as evidenced by the research into the importance of lateral reflections in real and simulated concert hall acoustics [6,13,27].

While the goals of virtual auditory displays differ somewhat from that of auralization systems, it is clear that the development of better systems for both domains could benefit from perceptually veridical simulation of room acoustics. For example, the percentage of unexternalized virtual acoustic sources has been shown in one study to decrease from 25 % to 3 % with the addition of synthetic reverberation [24]. To attain better auralization systems, future perceptual studies should include manipulation of separate independent variables related to early reflections. For instance, it could be that the perception of relative hall size in the second listening test described above could be due merely to the "initial time delay gap" cited by Beranek [28], or that the final reflection at 110 msecs really was a determining factor.

Once it is gathered, perceptual data could be used to interpret the results of the room model into a "listening model"; i.e., an additional element in the "translation software" shown in Figure 1. Such an algorithm could have multiple functions. One would be to **eliminate** spatial cues for computational efficiency; i.e., HRTF processing of reflections when they are likely to be masked. Another function would be to **exaggerate** particular spatial auditory cues so that a "net result" could be obtained across many subjects . Such exaggeration could be worked into to the HRTFs themselves. But the exaggeration could also be used to heighten different spatial aspects of a room selectively, perhaps by means of the user interface. This may be particularly important in a context where the sensitivity of the designer using the auralization system happens to differ from the sensitivity of the client or the critic!!

As a final note, the following anecdote seems appropriate. I once knew a former tenor with the New York Metropolitan Opera who was quite erudite on the matter of concert hall sound. I asked him how he knew, when he was an audience member, that he was in a first rate concert hall. He said that one clue was that "...**the sound seems to sizzle in the air above the people in front of you, up ahead a few rows.**". While I can't speculate as to whether or not this is a dependable qualitative measure, I've heard the effect that he's referred to, and I've been fascinated by this attention to spatial-auditory imagery. Unfortunately, I've yet to hear any auralization or 3-D sound system that comes close to creating this or a similar type of veridical frontal imagery over headphones. Why is unclear; the reasons could lie with the complexity and realism of the room model, the nature of the HRTFs used, and/or the lack of scanning ability provided by head tracking devices, as mentioned previously. Through some basic perceptual research, it may be possible to determine how much the 'realism' of room simulation lies with spatialized early reflection patterns, and how these reflections might be manipulated for a particular goal, such as improving virtual acoustic displays or auralization systems.

8. Acknowledgments

This work was produced under the auspices of the San Jose State University Foundation. Thanks to the "expert listeners", to my colleagues at NASA-Ames, and to Christopher B. Ickler, Senior Research Engineer in Acoustic Research at BOSE Corporation, for consulting with me on several issues.

9. References

[1] Kleiner, M., Dalenbäck, B.-I., & Svensson, P. (1991). Auralization-an overview. In 91st Convention of the Audio Engineering Society, Preprint 3119 . New York, New York: Audio Engineering Society.

[2] Kleiner, M., Svensson, P., & Dalenback, B.-I. (1990). Auralization: Experiments in Acoustical CAD. In 89th Convention of the Audio Engineering Society, Preprint 2990 . Los Angeles, California: Audio Engineering Society.

[3] Ahnert, W., & Feistel, R. (1991). Binaural Auralization from a sound system simulation program. In 91st Convention of the Audio Engineering Society, . New York, New York: Audio Engineering Society.

[4] Jørgensen, M., Ickler, C. B., & Jacob, K. D. (1991). Judging the speech intelligibility of large rooms via computerized audible simulations. In 91st Convention of the Audio Engineering Society, preprint 3126 . New York, New York: Audio Engineering Society.

[5] Xiang, N., & Blauert, J. (1991). Computer-aided tenth-scale modeling for binaural auralization in room acoustic design. In 91st Convention of the Audio Engineering Society, Preprint 3120 . New York, New York: Audio Engineering Society.

[6] Schroeder, M. R., Gottlob, D., & Siebrasse, K. F. (1974). Comparative study of European concert halls. Journal of the Acoustical Society of America, 56, 1195-1201.

[7] Schroeder, M. R. (1970). Digital simulation of sound transmission in reverberant spaces. Journal of the Acoustical Society of America, 47(2), 424-431.

[8] Borish, J. (1984) Electronic Simulation of Auditorium Acoustics. Ph.D. Dissertation, Stanford University.

[9] Kendall, G. S., & Martens, W. L. (1984). Simulating the cues of spatial hearing in natural environments. In Proceedings of the 1984 International Computer Music Conference, (pp. 111-125).San Francisco: Computer Music Association.

[10] Begault, D. R. (1987) Control of auditory distance. Dissertation, University of California San Diego.

[11] Cooper, D. H., & Bauck, J. L. (1989). Prospects for transaural recording. Journal for the Audio Engineering Society, 37(1-2), 3--19.

[12] Foster, S. H., Wenzel, E. M., & Taylor, R. M. (1991). Real-time synthesis of complex acoustic environments (Summary). In Proceedings of the ASSP (IEEE) Workshop on Applications of Signal Processing to Audio and Acoustics, IEEE Press: New Paltz, NY:

[13] Ando, Y. (1985). Concert Hall Acoustics. Berlin: Springer-Verlag.

[14] Wightman, F. L., & Kistler, D. J. (1989). Headphone simulation of free-field listening. I : Stimulus synthesis. Journal of the Acoustical Society of America, 85(2), 858-867.

[15] Wenzel, E. M., Wightman, F. L., & Foster, S. H. (1988). A virtual display system for conveying three-dimensional acoustic information. In Proceedings of the Human Factors Society 32nd Annual Meeting, (pp. 86-90). Santa Monica, California: Human Factors Society.

[16] Divenyi, P. L., & Blauert, J. (1987). On creating a precedent for binaural patterns: When is an echo an echo? In W. A. Yost & C. S. Watson (Eds.), Auditory Processing of Complex Sounds Hillsdale, NJ: Lawrence Erlbaum.

[17] Olive, S. E., & Toole, F. E. (1989). The detection of reflections in typical rooms. Journal of the Audio Engineering Society, 37(7/8), 539-553.

[18] Moore, B. C. J., Oldfield, S. R., & Dooley, G. J. (1989). Detection and discrimination of spectral peaks and notches at 1 and 8 kHz. Journal of the Acoustical Society of America, 85(2), 820-836.

[19] Wightman, F. L., & Kistler, D. J. (1992). The dominant role of low-frequency interaural time differences in sound localization. Journal of the Acoustical Society of America, 91(3), 1648-1661.

[20] Wenzel, E. M., Arruda, M., Kistler, D. J., & Wightman, F. L. (1992) Localization with someone else's ears. Manuscript submitted to the Journal of the Acoustical Society of America.

[21] Wightman, F. L., & Kistler, D. J. (1989). Headphone simulation of free-field listening. II: Psychophysical validation. Journal of the Acoustical Society of America, 85(2), 868-878.

[22] Wenzel, E. M., Wightman, F. L., Kistler, D. J., & Foster, S. H. (1988). Acoustic origins of individual differences in sound localization behavior. Journal of the Acoustical Society of America, 84(S79).

[23] Begault, D. R., & Wenzel, E. M. (1991). Headphone Localization of Speech Stimuli. In Proceedings of the Human Factors Society 35th Convention, (pp. 82-86). San Francisco: Santa Monica: Human Factors Society (accepted for publication in Human Factors).

[24] Begault, D. R. (1991). Perceptual effects of synthetic reverberation on 3-D audio systems. In 91st Convention of the Audio Engineering Society, Preprint 3212 . New York, NY: Audio Engineering Society

[25] Begault, D. R. (1991). Challenges to the successful implementation of 3-D sound. Journal of the Audio Engineering Society, 39(11), 864-870.

[26] Gibson, J. J. (1966). The Senses Considered as Perceptual Systems. Boston: Houghton Mifflin.

[27] Barron, M., & Marshall, A. H. (1981). Spatial impression due to early lateral reflections in concert halls: the derivation of a physical measure. Journal of Sound and Vibration, 77(2), 211-232.

[28] L.Beranek, L. (1954). Acoustics New York: McGraw-Hill.

TABLE I: Summary of spatial positions given by room modeling program, rounded within 5 ° for 3-D sound processing.

surface	LISTENER POSITION 0 source at 3 ft.	LISTENER POSITION 2 source at 12 ft.	LISTENER POSITION 4 source at 12 ft.
floor	0 , - 36	0, - 36	0, - 36
left wall	left 75, 0	left 45, 0	left 75, 0
front wall	0, 0	0, 0	0,0
right wall	right 75,0	right 45, 0	right 75, 0
back wall	180, 0	180,0	180,0
ceiling	0, + 54	0, + 54	0, + 54
DIRECT	0, 0	0,0	0,0

TABLE II: Azimuth error for dry speech, 0 degree elevation.

target azimuth	% of judgements with < 10 degrees error	% of judgements within 11-30 degrees error	% of judgements with > 30 degrees error
0	33	2	65
30	9	17	64
60	5	58	37
90	74	10	16
180	56	7	37
Mean	**36.3**	**19.3**	**44.3**

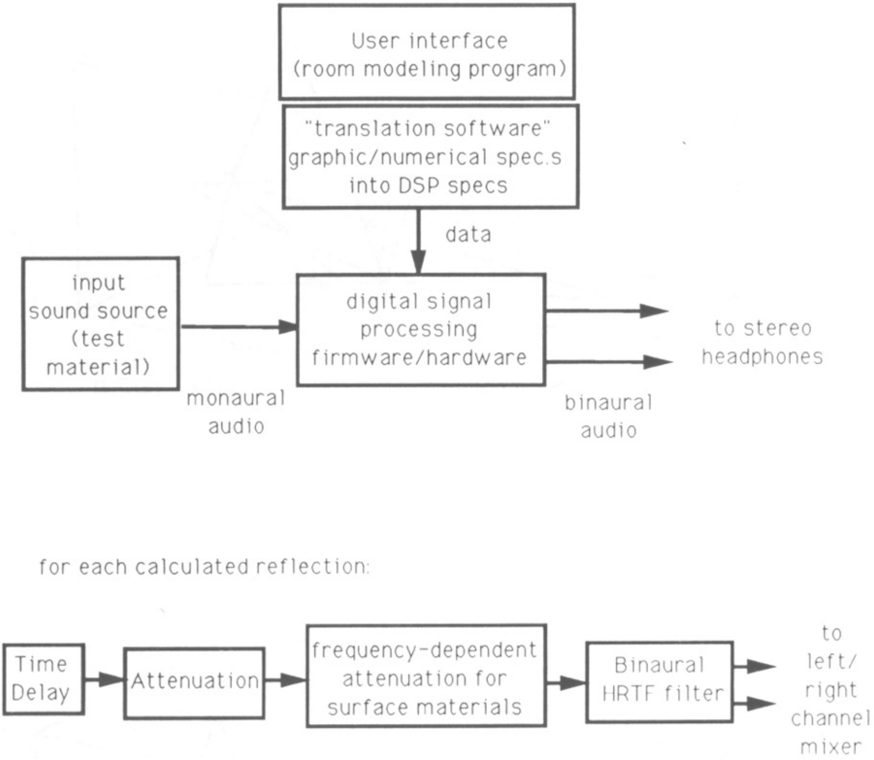
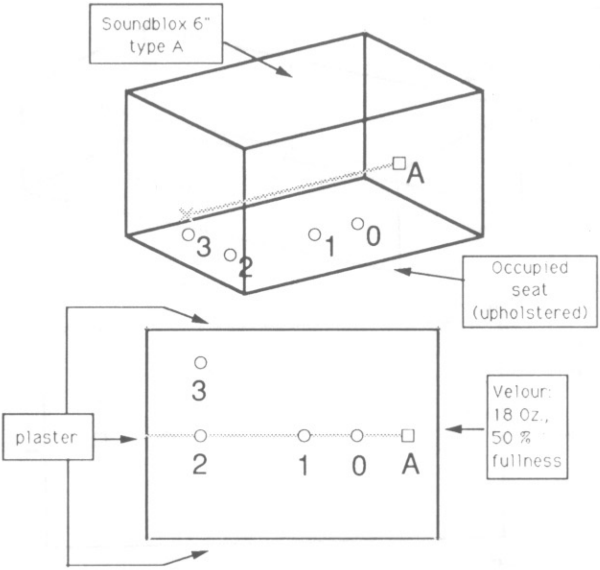


FIGURE ONE. Components of a hypothetical real-time auralization system, used as the model for off-line signal processing in the current paper.



MODELER+ parameter summary:

room dimensions: 18 x 13 x 9.5 speaker 2 ft. from back wall
listener positions: 0, 1 and 2 seated at 3, 6 and 12 feet from speaker, 0 degree orientation, listener position 3 at 20 ° off axis, 2 feet from side wall
reflection parameters: no obstructions, smoothing, specular diffusion
speaker type: omnidirectional, 1 watt applied, height at seated ear level (3.3 feet)

reverberation time:

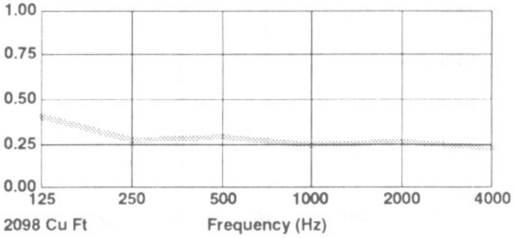
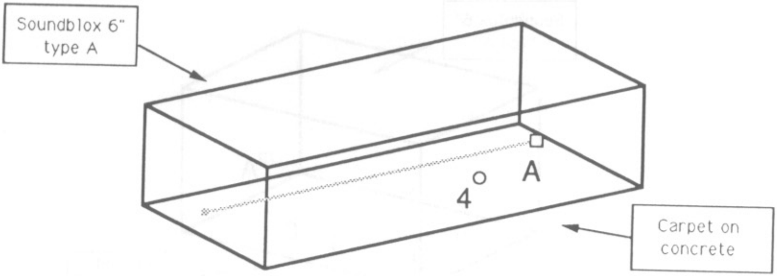
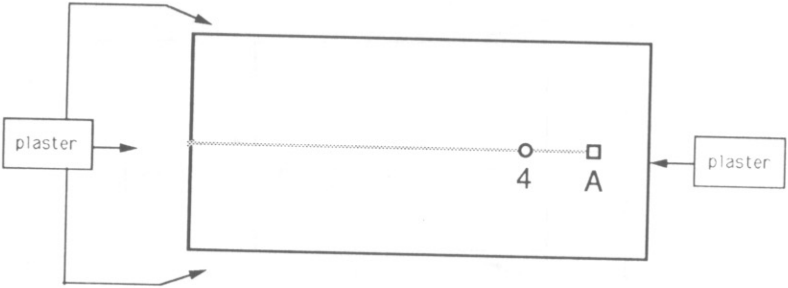


FIGURE TWO: small room design parameters





MODELER+ parameter summary:

room dimensions: 85 x 40 x 20, speaker 10 ft. from back wall
listener positions: 4 seated at 12 feet from speaker, 0 degree orientation (same distance as listener position 2 in the small room)
reflection parameters: no obstructions, smoothing, specular diffusion
speaker type: omnidirectional, 1 watt applied, height above seated ear level (4.9 feet)

reverberation time:

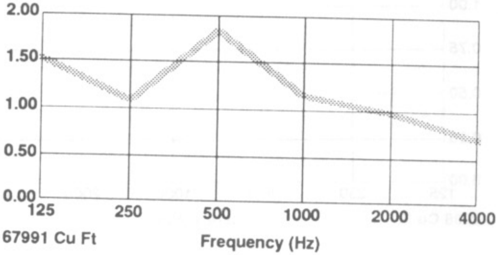
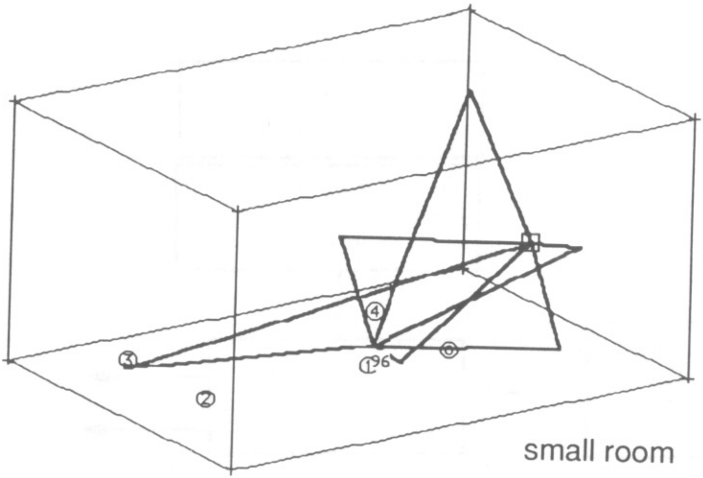
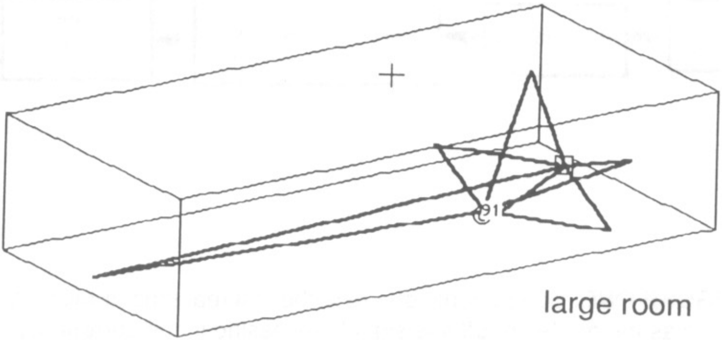


FIGURE THREE: large room design parameters



small room



large room

FIGURE FOUR. Examples of ray tracing results for specifications in Figures 2 and 3.

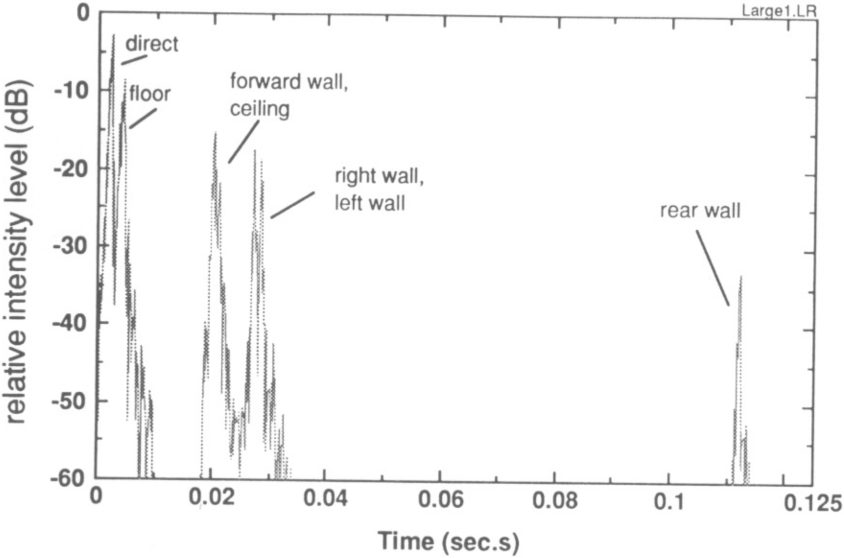


FIGURE FIVE. The combined binaural magnitude response from summed HRTF-filtered early reflection impulse responses for listener position 4 (see Figure 3).

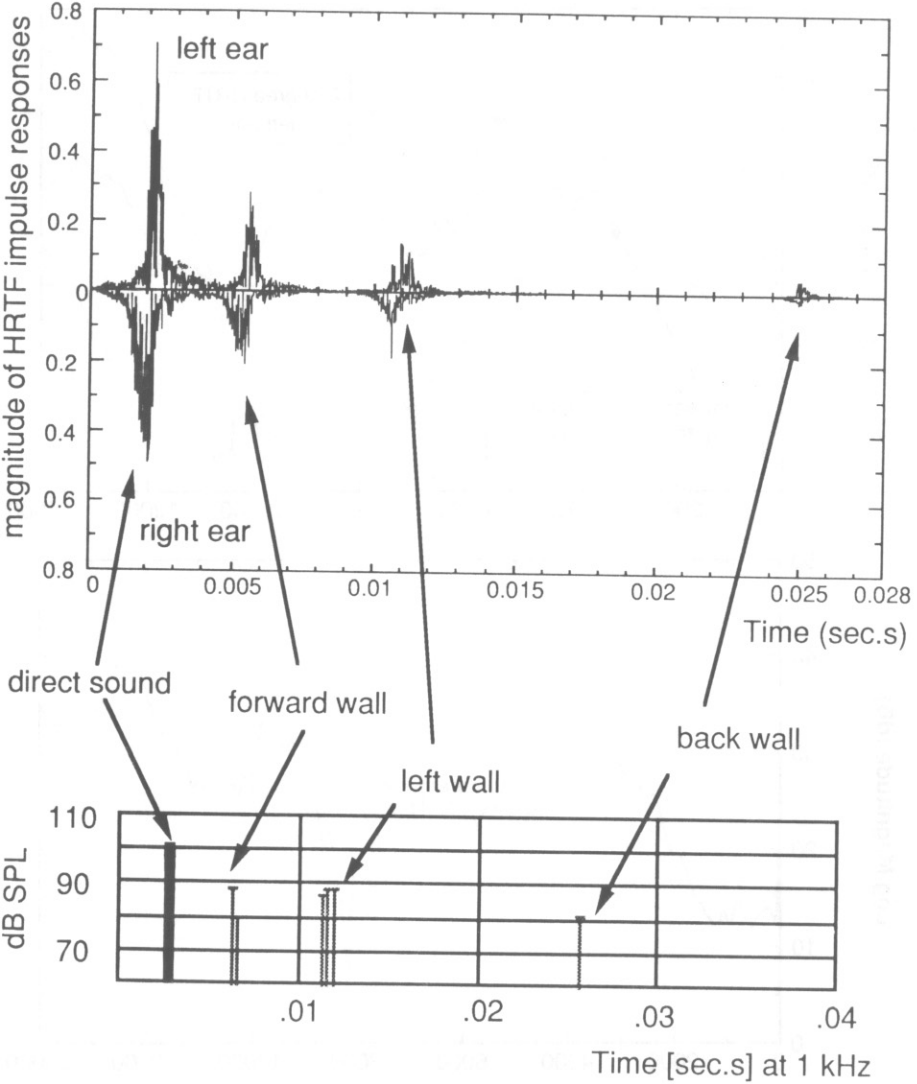


FIGURE SIX: Direct sound and early reflections for listener position 0 (figure 2). **Bottom:** reflectogram (dB SPL/time) given by the <u>Modeler+</u> program. **Top:** reflection data normalized in amplitude, direct arrival time = 0, and filtered by HRTFs. Peak amplitudes in binaural version occur about .002 seconds after onset time as given in reflectogram.

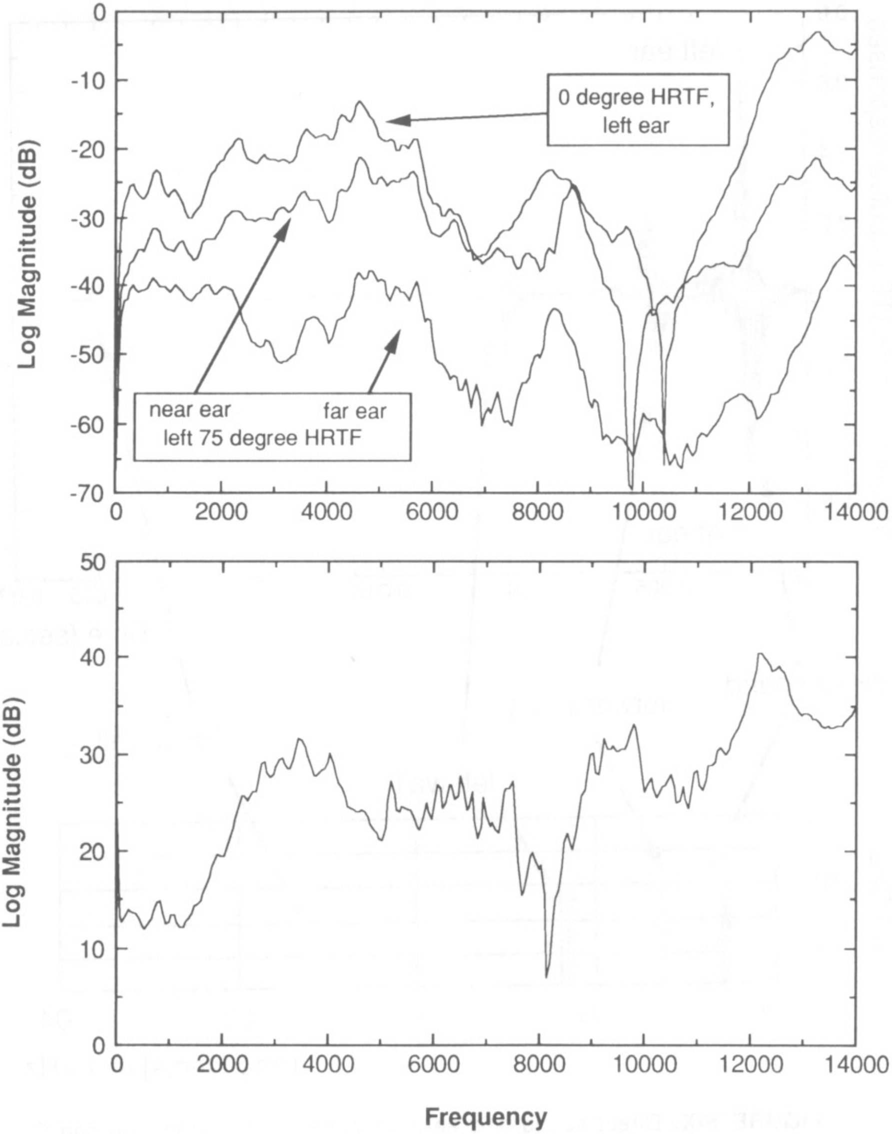


FIGURE SEVEN. Top: HRTF magnitudes based on listener position 0: reflection from left wall (left 75 degrees) -12 dB down, with direct sound HRTF shown for comparison (0 degrees). **Bottom:** The difference in magnitude at the right ear for the 0 degree HRTF, and the 75 degree HRTF, -12 dB down