

# A SPATIAL SOUND PROCESSOR FOR LOUDSPEAKER AND HEADPHONE REPRODUCTION

GARY S. KENDALL, WILLIAM L. MARTENS, AND MARTIN D. WILDE

*Auris Corporation, 1801 Maple St., Evanston, IL 60201 (708) 491-6594*

## I. INTRODUCTION

When an observer listens to an acoustical event in a reverberant environment, a single auditory event is heard at a given direction and distance. When the same event is recorded with a single microphone and reproduced over loudspeakers, the observer hears an event that is spatially impoverished. None of the spatial information from the original acoustic event has been preserved in the recording. This lost spatial information is captured in a binaural recording, but it is still quite difficult to reproduce, even over headphones. What excites most people about listening to a binaural recording is the sensation of the recorded event as a "first hand" experience. The problematic feature of binaural recording is that it can only capture first hand events, providing a kind of acoustic documentary. Once spatial information is recorded, it cannot be modified in any of its essential elements.

As we enter the 1990s, the rapid evolution of digital signal processing (DSP) technology promises to close the gap between natural acoustic events and synthesized events. This is just as true for spatial sound synthesis as it is for musical tone synthesis. A DSP-based spatial sound processor fills the need for a practical post-production tool to process individual sound tracks and produce a result that creates three-dimensional (3D) spatial imagery. Artistic decisions about the spatial placement of sound can be completely decoupled from the original performance setting. Rather than being limited by the loudspeakers, sound sources appear to be positioned at arbitrary locations in the space surrounding the listener. This facilitates the exploration of new sonic possibilities driven by creative needs which are not limited by the physical constraints of the recording process. With the development of this DSP technology there will be little reason for using binaural recording.

This paper focuses on the characteristics of a three-dimensional spatial sound processor under development for use in loudspeaker applications such as music, television, and film post-production, and in headphone applications such as computer-human interfaces (see Fisher, 1989). The development began at Northwestern University's Computer Music Studio nearly nine years ago (see Kendall and Martens, 1984). The original goal of this work was to provide composers and oth-

er sound artists with a revolutionary creative tool. At that time, the work was carried out entirely in software on a large-scale digital computer. Today, a real-time hardware device is being developed which completes and extends the work begun at Northwestern.

A primary goal for this project has been the creation of auditory spatial images that have all the complexity and richness of everyday experience in typical environments. Achieving this goal requires the synthesis and control of the complete ensemble of auditory cues used in human spatial hearing. The computational model upon which this is based is called the "spatial reverberator" (Figure 1 is a very general instantiation). It is used to simulate the acoustics of the head and a model room. Given specifications of the room dimensions, sound source position, and listener position, it produces a very close approximation of the signals that would reach the listener's ears in a real environment.

The development and evolution of the spatial reverberator have been guided not only by acoustical considerations, but

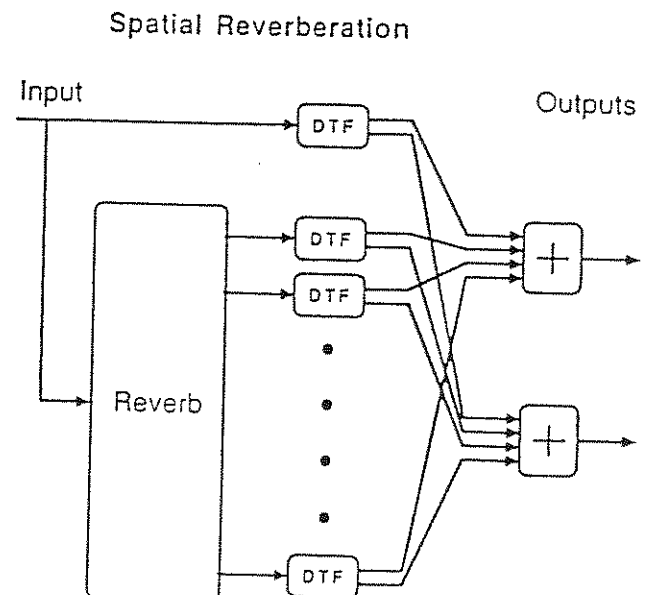


Figure 1. Simple network for spatial reverberation with two audio output channels.

## Time Domain

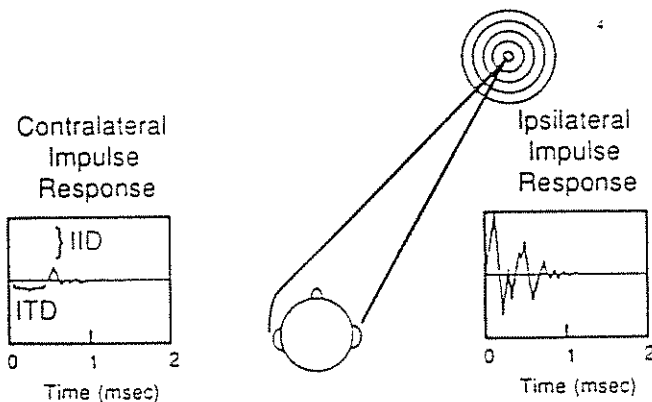


Figure 2. Time domain representation of impulse responses at the ipsilateral and contralateral ears showing IID and ITD.

also by the requirements of the perceptual system. Rather than concentrate on extremely accurate simulation of physical acoustics, the emphasis is placed on spatial image formation, the perceptual process by which acoustic information is translated into the experience of events surrounding the listener. There are two essential elements to a practical system that manipulates spatial imagery. The first is the directionalization of sound sources. Synthetic directional transfer functions (DTFs) provide comprehensive control over perceived direction by capturing the idealized features of measured head-related transfer functions (HRTFs). The DTFs are implemented as finite impulse response (FIR) filters and their perceptual effectiveness is refined through an iterative procedure, generating and testing new sets of FIR filter coefficients. The second es-

## Cone of Confusion

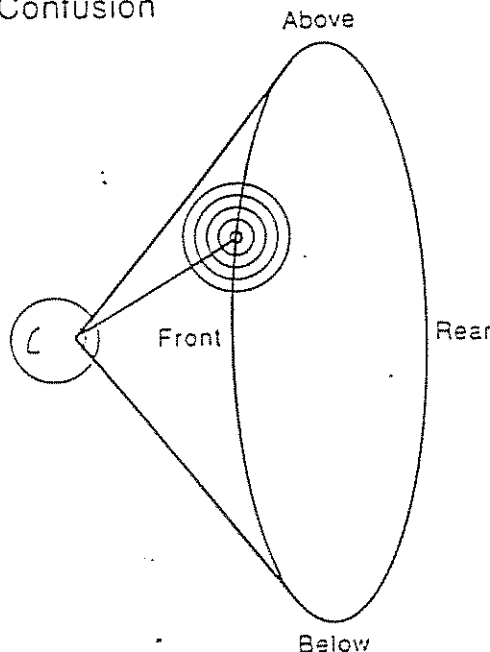


Figure 3. The cone of confusion describes the set of all possible directions from which sounds can arrive when those sounds share a given degree of lateralization. Four sectors of the cone are identified, and a sound source located to the front and above the listener is illustrated.

ential element is the simulation of environmental sound, which provides the acoustic basis for the perception of distance, spaciousness, definition and spatial texture. Environmental sound is also important for the naturalness of the composite spatial image, especially when the sound source is moving dynamically.

## II. DIRECTIONALIZATION

### A. Binaural Cues

The physical properties of the torso, head and pinna determine the acoustic information available at the two ears from which directional judgements are made. (The influence of posture and other non-acoustic factors on localization has been reviewed by Lackner [1983].) The separation of the two ears provides the acoustic basis for the traditional psychoacoustical cues of interaural time difference (ITD) and interaural intensity difference (IID). A time domain representation of the responses measured at the ipsilateral and contralateral ears for an impulsive sound source located to the right are shown in Figure 2. It can be easily seen that the the contralateral response is later in time and less intense than the ipsilateral response.

These cues provide the auditory system with information for judging the lateral position of a sound source along the interaural axis, a one-dimensional, left-right axis defined by the two ears. For example, given an ITD and IID favoring the right side, a person will judge the sound event to be located on the right. In headphone reproduction with ITD and IID alone, subjects report hearing images that are lateralized along the interaural axis, but located inside the listener's head.

With only ITD and IID, a person cannot judge whether an acoustic event is in front, above, behind, or below. Woodworth (1954) and many others have called this circular dimension at a given degree of lateralization the "cone of confusion" (see Figure 3). It turns out that the ambiguity of spatial position on a cone of confusion is resolved by the complex acoustic profiles at the ears created by the torso, head and pinna. While the overall ITD and IID are at a relatively constant value around a cone, the magnitude and phase functions for each direction are unique. Figure 4 shows the magnitude response at the ipsilateral and contralateral ears measured at the eardrum for a single spatial location. The information in the frequency domain representation corresponds exactly to the time domain representation in Figure 2. An examination of all these cues leads to the conclusion that azimuth and elevation angles, though convenient for specifying direction, may not be the best for characterizing directional cues. Figure 5 illustrates the two dimensions that seem better suited to describing human directional perception: the linear dimension of left-right laterality and the circular dimension of front / above / rear / below positions on a cone.

**Role of Head Movements.** It should be pointed out that dynamic ITD and IID can provide effective cues to front / back position when synchronized with head movements. Figure 6 illustrates what happens when the head is turned to the right for two opposing sound locations. If the sound is located in front of the listener, a right head turn introduces a lateral shift in the sound source toward the left ear. The opposite happens when the sound is directly behind the listen-

# Frequency Domain

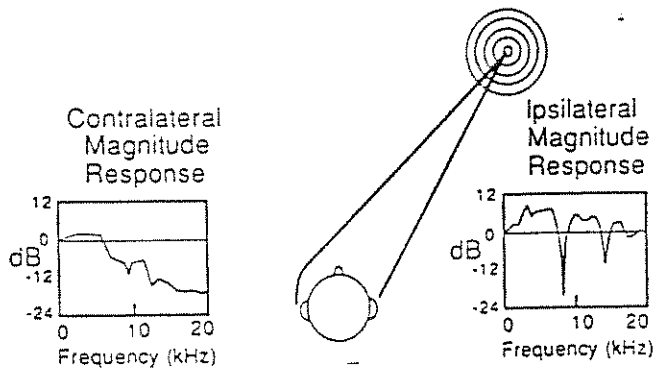


Figure 4. Frequency domain representation of impulse responses at the ipsilateral and contralateral ears.

er: IID and ITD shift the sound source toward the right ear.

Wallach (1940) demonstrated the dominance of dynamic ITD / IID over other directional cues in a series of classic experiments with multiple loudspeakers. He used a mechanical head-tracking system to control the stimulus presentation. As illustrated in Figure 7, if a sound starts on the right and is then shifted through the loudspeaker array at twice the rate of the listener's head rotation, the time and intensity differences shift as they would for a rearward shifting sound source. The listener perceives an auditory image that shifts rearward, even though the actual sound source has moved toward the front. Though it is clear that the pinna plays an important role in distinguishing between front and rear directions when the head is immobile (Gardner & Gardner,

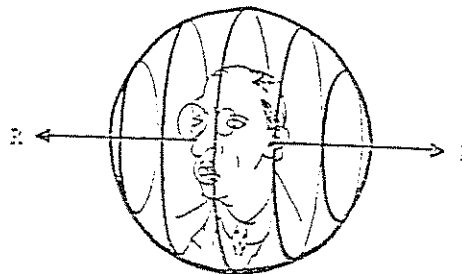


Figure 5 Lateralization and cones of confusion combine to provide a system for identifying a complete sphere of sound source directions.

1973), the conflict between pinna cues and dynamic ITD / IID is resolved in favor of ITD and IID.

In headphone reproduction, manipulation of IID and ITD typically moves an auditory image only left and right inside the head along the interaural axis. Strong front / back distinctions can be supported in headphone reproduction when a head-tracking mechanism is used to sense the movement of the listener's head and IID and ITD are changed accordingly. Figures 8 and 9 show how head-tracking can disambiguate dynamic interaural cues. Figure 8 shows that without head-tracking, an image lateralized to the left ear will shift toward the center of the head as the IID is gradually decreased to zero (see figure caption for explanation). When, however, the IID decreases to zero in response to a 90 degree turn of the head to the left, then the image will unambiguously shift toward the front, as illustrated in Figure 9 (see caption). This observation, taken together with Wallach's results, suggests an important caveat with regard to the evaluation of DTFs used in headphone systems. Good front / back cuing is expected on the basis of dynamic IID and ITD alone. Only by turning head-tracking off, or by steering the source from front to rear without turning the head, can the potency of a particular set of DTFs front / back distinction be assessed.

## B. Data Development

One approach to producing directionalized sound images would be to make non-reverberant impulse response measurements at the eardrums of a dummy head or human subject and

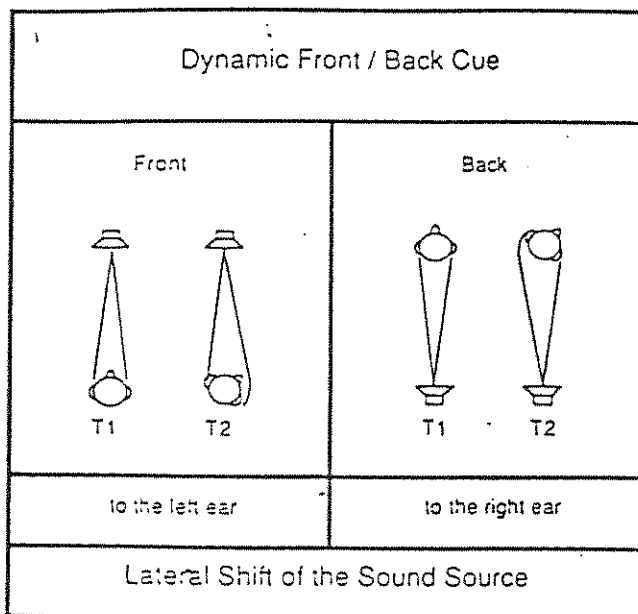


Figure 6. Dynamic head movement and front/back disambiguation.

## Wallach's (1940) Headtracking Display

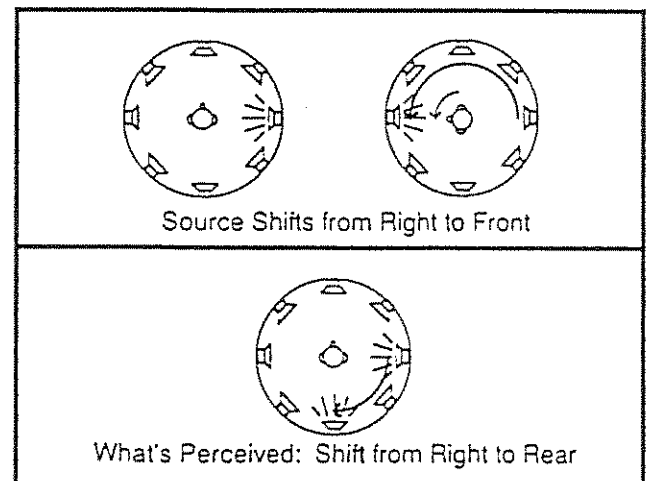


Figure 7. Wallach (1940) presented a broadband sound source to a listener via an array of loudspeakers, the signals to which were linked to head movements. The upper panel shows the sound starting at the loudspeaker located to the right of the listener and shifting to the loudspeaker located in front of the listener as the listener's head is turned 90 degrees to the left. The lower panel shows the perceptual result, a shift in the image toward the rear of the listener.

## Ambiguous Dynamic Auditory Cue

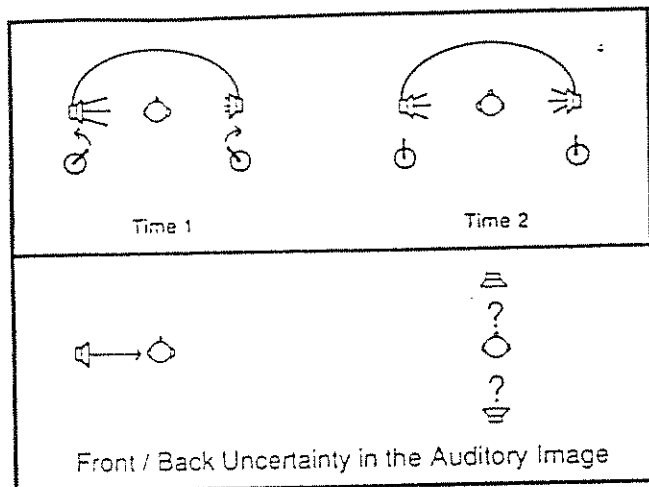


Figure 8. Headphone presentation of changing IID results in front / back uncertainty when this dynamic cue is not associated with head rotation. The upper panel depicts an IID favoring the left ear at time 1, and gradually shifted to zero interaural difference at time 2. The lower panel represents the perceptual result of the presentation. Between time 1 and time 2, the image shifts from the left ear toward the center of the head.

use these recorded data directly as the coefficients of an FIR filter. This approach would yield exactly the same directional cues as a binaural recording without the environmental sound. Our experience with direct use of such measurements suggests that they provide a basis for directional judgements but their success varies greatly among individuals and reproduction strategies.

An alternative approach is to develop idealized DTFs that deviate from measured HRTFs but support comparable or improved localization. Figure 10 illustrates the process through which DTFs are developed for the spatial sound processor. Direct measurements are made of HRTFs which are then transferred to a digital computer which is used to abstract general features of the data. Spectral band data is submitted to a statistical procedure called principal components analysis. The analysis results are manipulated to create "idealized" DTFs. These new DTFs are evaluated over headphones and loudspeakers and the results lead to revised strategies for idealization.

Measurements. All of our measurements of head-related transfer functions over the last eight years have been made through Time-Delay Spectrometry (TDS) using the Crown TEF-10 analyzer. A sine-wave sweep was delivered from an MDM TA-2 loudspeaker at a fixed location relative to the reference microphone or the head of the subject. We have found it very important, and at times very difficult, to remove environmental sound from the measurements, and have therefore made most measurements in the anechoic chamber at Northwestern University. Figure 11a shows the measurement setup. The Kemar mannequin was mounted laterally and turns 360-degrees about the mounting arm. When the mounting arm was perpendicular to an imaginary line drawn between the loudspeaker and the center of Kemar's head, the loudspeaker was in Kemar's horizontal plane. Changing the angle between the mounting arm and the imaginary line changes the eleva-

## Unambiguous Dynamic Auditory Cue

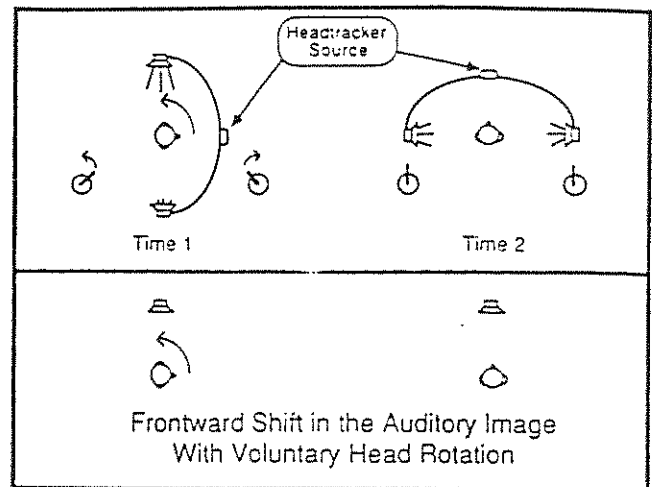


Figure 9. A forward shift in the apparent direction of the sound source occurs when the dynamic auditory cue is generated in response to a voluntary head rotation. The upper panel depicts an IID favoring the left ear at time 1 that is shifted to zero at time 2 in accordance with a head turn sensed by a headtracking system. The lower panel represents the perceptual result of the presentation. Between time 1 and time 2, the listener turns toward a spatially stabilized image that starts on the listener's left and ends in front of the listener.

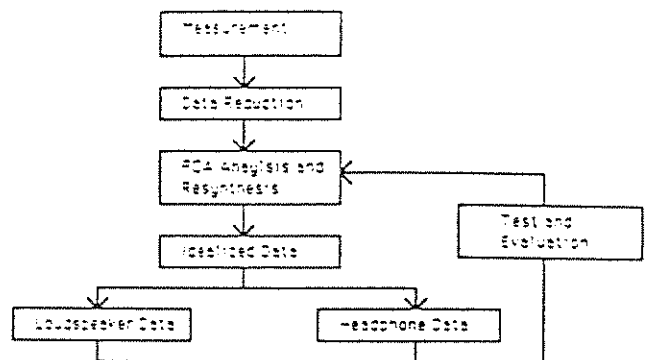


Figure 10. Flowchart of development of data for idealized DTFs.

tion of the loudspeaker relative to Kemar. This procedure enabled us to keep the loudspeaker and Kemar's head in the exact same positions within the room while recording an entire sphere of data. Our measurements of Kemar were collected at 10-degree resolution in both azimuth and elevation. Tests for repeatability showed that measurements made for the same angle at different times varied by no more than two dB.

We have also made measurements with human subjects using a different equipment setup. Subjects were seated on a platform that rotated 360-degrees in 10-degree increments. In this case the vertical position of the loudspeaker was adjusted to provide a complete range of elevation angles. Measurements were taken with a Knowles BT-1759 microphone embedded in a waxy cotton plug placed into the subject's ear canal and molded flush to the ear-canal entrance. This is referred to as the "blocked meatus" measurement. This method was chosen over the traditional probe tube technique because it provided repeatable results. Probe tube measurements often suffer from the problem that it is

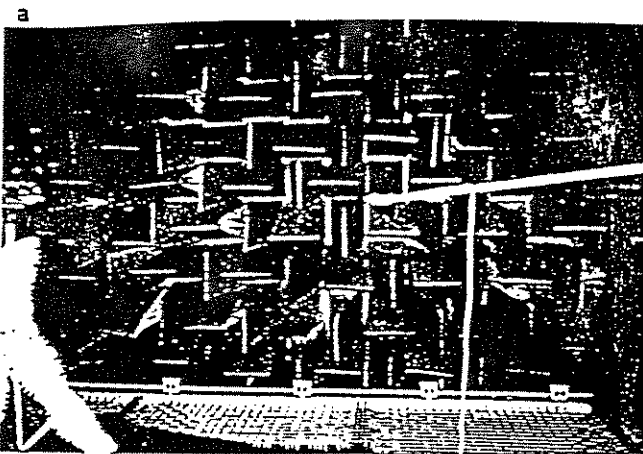
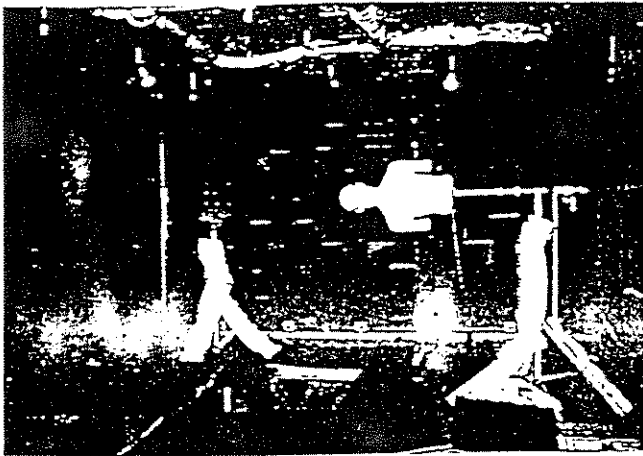


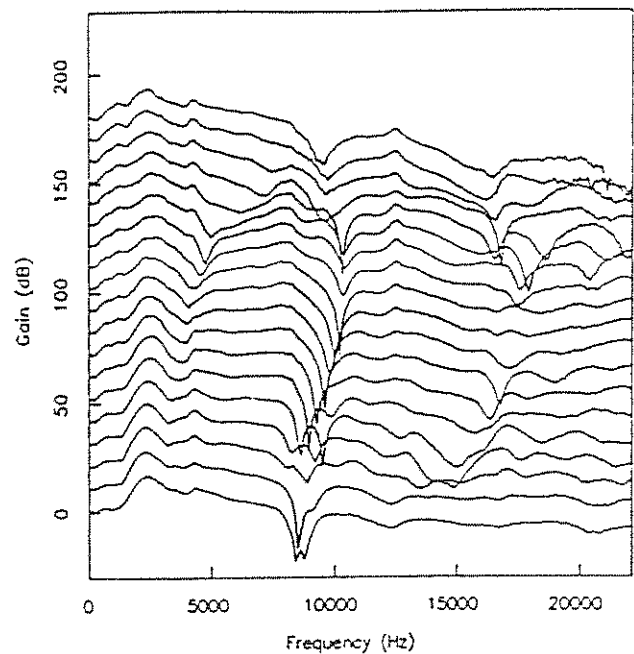
Figure 11. a) Experimental setup for HRTF measurement with the Kemar mannequin mounted laterally in the Northwestern University Anechoic Chamber b) Experimental setup for reference measurement.

very difficult to repeat the exact positioning of the end of the probe tube near the ear drum and the measurements vary considerably with changes in position. Our blocked meatus measurements proved to be nearly as repeatable as the Kemar measurements.

In both measurement setups, reference measurements were taken with the microphone held in free space at the same position as the subsequent measurement (see Figure 11b). Measurements were transferred from the Crown TEF analyzer to a network of general-purpose computers (Pyramid 90X and SUN workstations). In order to obtain free-field HRTFs, the reference measurement for the combined microphone-loud-speaker-room response were divided out, and impulse responses were obtained through application of the discrete inverse Fourier transform.

Acoustic characteristics. The resulting measurements can be displayed in various formats. Figure 12a shows a two dimensional view of a series of frequency domain measurements for the horizontal plane in 10-degree increments from 0-degrees azimuth at the bottom of the plot to 180-degrees azimuth at the top. Figure 12b shows a contour plot of the same series of frequency domain measurements. Despite the fact that the pinna transfer functions are highly complex, they exhibit some easily identifiable spectral features. A quick examination reveals spectral notches and peaks whose

Ipsilateral Magnitude Response



Contour Plot of Ipsilateral Magnitude Response

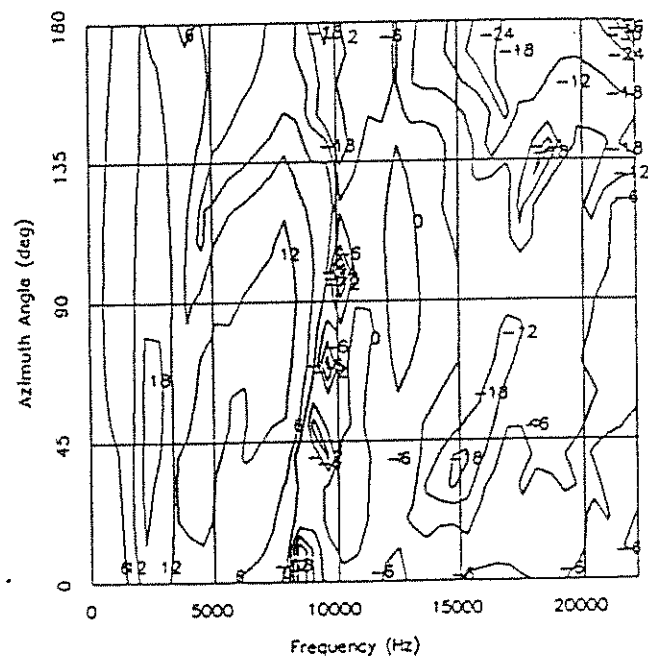


Figure 12. a) Free-field magnitude response measured at the eardrum position for 18 sound sources azimuth angles on the horizontal plane. The curve at the bottom of the graph was measured at 0 degrees azimuth (front) and the curve at the top of the graph was measured at 180 degrees azimuth (rear). Each successive curve is shifted up by 10 dB for graphical comparison. b) Contour plot of the magnitude response curves plotted in "a." Isomagnitude contours are plotted at 6 dB intervals.

frequencies are dependent on the incidence angle of the source signal and migrate in a coherent manner with changing azimuth. It is important to examine the temporal behavior of the HRTFs, since it has been long supposed that there

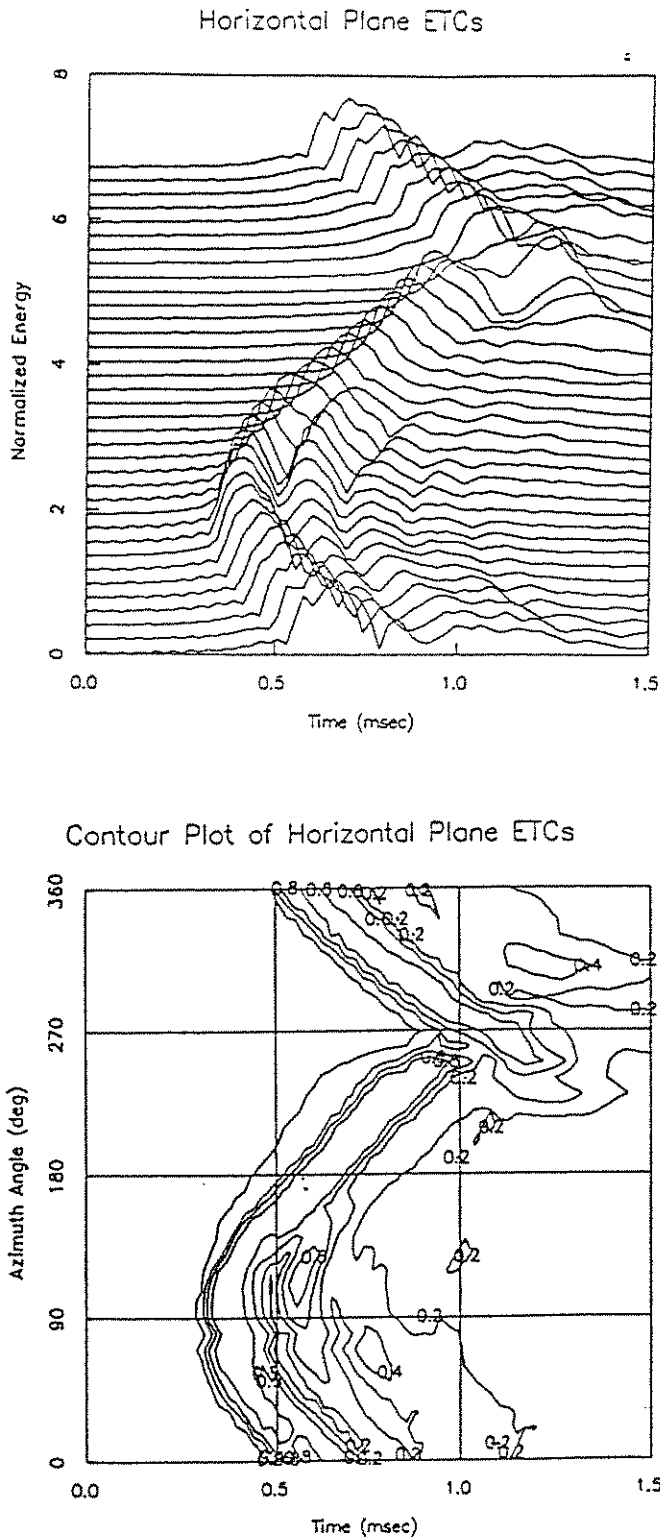


Figure 13. a) Energy-Time Curves (ETCs) measured at the eardrum position for 36 sound source azimuth angles on the horizontal plane. The curve at the bottom of the graph was measured at 0 degrees azimuth (front) and the curve halfway up the graph was measured at 180 degrees azimuth (rear). Each successive curve is normalized then shifted up for graphical comparison. b) Contour plot of the ETCs plotted in "a." Isoenergy contours are plotted at intervals of 0.2 normalized energy units.

are also directionally-dependent features in the time domain (Batteau, 1967).

Figure 13a shows a set of energy-time curves (ETCs) for HRTFs in the horizontal plane similar to those measured by Hiranaka and Yamasaki (1983). Figure 13b shows a contour plot for the same set of ETCs. Notice how the primary arrival time of the main energy varies as a function of direction. The arrival time is shortest when the ear is closest to the sound source at 90-degrees and longest when the ear is turned away on the contralateral side at 270-degrees. In fact, just beyond 270-degrees one can observe the combination of delays wrapping around the front and the back of the head. Secondary energy peaks are also shown here to vary across azimuth, but again in a coherent manner with the primary peaks.

Individual differences. Almost every researcher has noted that HRTFs vary tremendously from one individual ear to the next. Careful examination reveals that despite the variety of details, there are numerous common trends. For example, on the frontal plane (the plane defined by the left / right dimension and the above / below dimension), the frequencies of the two most prominent spectral notches generally increase with increasing elevation. The exact shapes of the HRTFs can differ, as shown in Figure 14 for subject MDL and subject GSK, but both show the same trend in the migration of these spectral notches. This might suggest that the directional information supplied by the pinna can be largely characterized in terms of these spectral notches, although there are several other observable trends involving spectral peaks and the overall spectral contour. In fact, one can separate to some extent the individual spectral features contributed by the head and the pinna.

Numerous authors (Butler & Belendiuk, 1977; Morimoto & Ando, 1983; Wightman & Kistler, 1989) have demonstrated that it is quite possible for one person to utilize the directional hearing cues recorded with another person's ears. Results indicate that some ears provide better basis for directional judgements than others. This also suggests that there is some universality in the acoustical basis of directional hearing, but the issues of how the auditory system evaluates the complex spectral profile at the two ears has not been adequately investigated at this time.

Spectral Band Analysis. In preparation for statistical analysis, the HRTF data were down-sampled by partitioning the data in the frequency domain representation, both magnitude and phase, into 34 bands that approximate critical band spacing. The values within these bands were averaged to provide a sequence of 34 values of magnitude and of phase for each HRTF. We have also used techniques that modeled critical-band filter characteristics more accurately (Petersen, 1980; Stautner, 1983), but have found this extra complexity to be unnecessary.

For completeness, we should also mention that we have used pole-zero filter design techniques to reduce HRTFs to a set of pole-zero locations in the complex  $z$ -plane. Fourth-order approximations were described by Kendall and Rodgers (1982). In the period from 1984 to 1987, eighth- to twelfth-order approximations were used. An automatic filter design program was interactively supervised by a user who could stop the automatic process and intervene when it got into trouble.

Eventually, the user began to design filters with very little program assistance. The pole-zero filters were constrained by the need to assure smooth migration of spectral features from one directional filter to the next, which was accomplished by interpolating pole and zero positions continuously. We eventually found this approach unwieldy since every new filter had to be designed with consideration of its neighbors. It was also the case that the pole-zero specifications did not easily lend themselves to further analysis and redesign.

**Idealized DTFs.** Principal components analysis (PCA) is a multivariate statistical technique which is well suited to the task of quantifying directionally-dependent trends in measured HRTFs (Martens, 1987). The variation in HRTF magnitude and phase across spectral bands was submitted to PCA, and the obtained principal components typically captured the directional distinctions made by the acoustic information contained in a set of HRTFs. Generally, a few components captured all of the directional variation on a given plane. Each principal component determines a spectral band weighting sequence that defines the contrast in spectral features that account for one of the directional distinctions made by HRTFs. When the scores on all components are matrix-multiplied by the obtained weighting sequences, the original HRTF spectral band data is reconstructed exactly.

PCA results were used to produce experimental DTFs making idealized directional distinctions. Novel DTFs were synthesized from a unique set of scores on the three or four principal components that supported primary directional distinctions such as ipsi vs. contra, front vs. back, etc. Component scores were weighted so as to emphasize or deemphasize particular directional distinctions, eliminating others, etc. This process is summarized in Figure 15. The resulting novel spectra were based upon measured acoustics, but were not derived from manipulating an acoustic model. Strategies taken toward the synthesis process depended upon the target direction, and the resulting idealized spectra could vary considerably from measured data.

Once these novel spectra were created they were evaluated subjectively. Our evaluation included both headphone and loudspeaker reproduction. Neither mode of reproduction can be considered neutral to the data. Headphone reproduction minimally requires that the novel spectra be equalized to remove the spectral characteristics of the headphones and their coupling to the head. In our experience, more active processing is required to place images reliably in front of listeners. This experience is consistent with the difficulty obtaining frontal imagery experienced throughout the long history of binaural recording (Sunier, 1986). Loudspeaker reproduction requires equalization and an active approach to deal with crosstalk. For these reasons, our spatial sound processor stores separate sets of DTFs for headphones and for loudspeakers.

### C. The Loudspeaker Reproduction Setting

The effect of the reproduction environment has been an often overlooked aspect of spatial sound. Auditory research requires a controlled environment in which the acoustics are understood and quantified. An anechoic chamber is a controlled environment, but subjects find it uncomfortable and its relationship to more typical reproduction spaces is questionable.

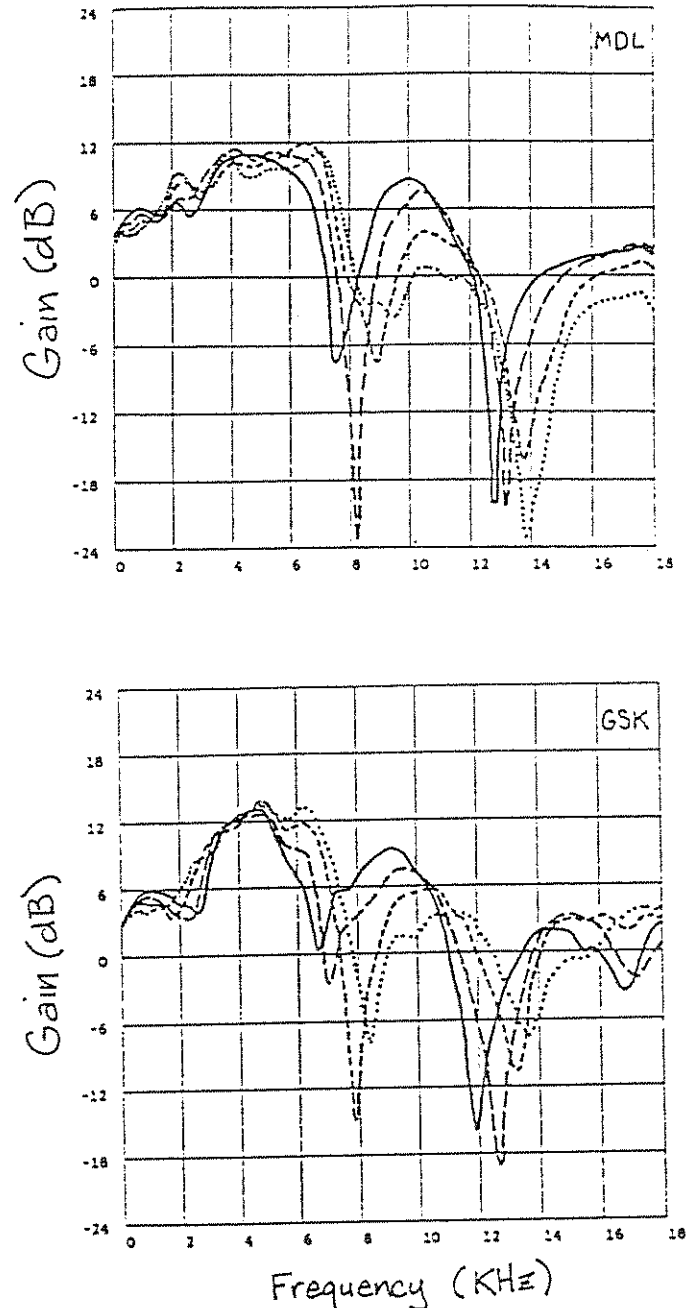


Figure 14. Head-related transfer functions for two subjects illustrating trends in the spectral features for increasing source elevation in the frontal plane. The source source was located 2 meters to the left of the subject and was moved from ear level (0-degrees) to an elevation of 30 degrees above ear level (solid line - 0 deg., long dashes - 10 degrees, short dashes - 20 degrees, and dotted line - 30 deg.).

We designed a different kind of controlled listening environment at the Northwestern University Computer Music Studio (Jones, Kendall, & Martens, 1984). Without actually constructing a "live-end, dead-end" (LEDE) studio monitoring room, we set about controlling the early reflected sound in order to insure a long initial time gap between the arrival of the direct sound from the speaker and the arrival of reflected sound from the room. Our goal was a room that was anechoic between the loudspeakers and the listening position while otherwise reverberant and perceived by subjects as a "normal" acoustic environment.

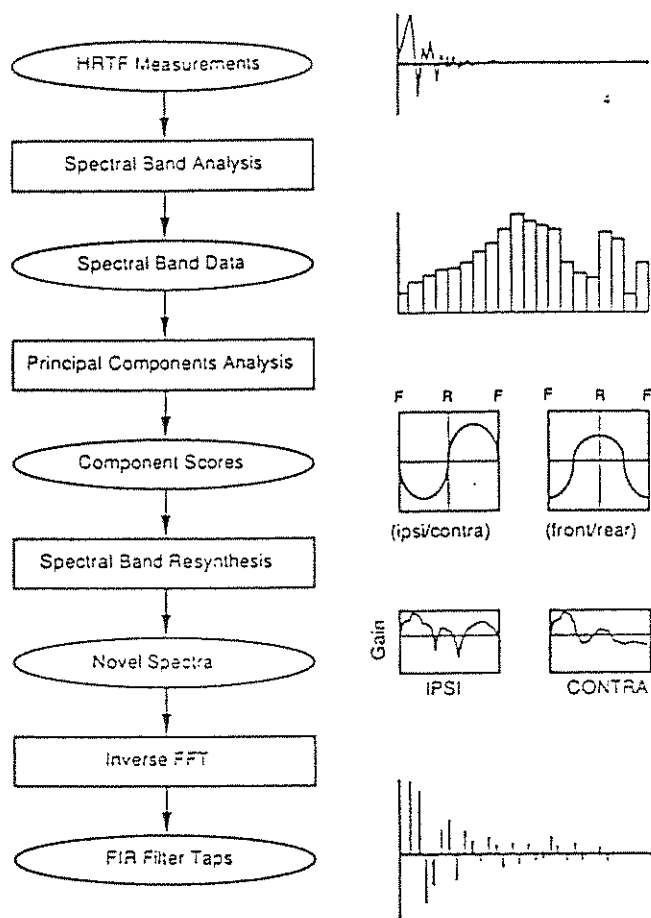


Figure 15. Flowchart of principal components analysis and resynthesis for generating novel DTFs.

Prior to construction, a computer program was used to select dimensions that fit within the available space and minimized the effect of room modes. After the room was constructed, sound absorbent material was attached to stiff foam-core panels. Both the walls of the room and the backs of absorbent panels were covered with Velcro strips that made it possible to position sound absorption anywhere within the room.

The next step in the process was to determine the exact placement of the panels that would eliminate early sound reflections between the loudspeakers and the listening position. Figure 16a shows the initial state of the room with one loudspeaker in the corner, the microphone in the listening position, and no absorption on the walls. Figure 16b shows the corresponding energy-time curve (ETC) measured with the Crown TEF analyzer. The time window starts at 6.3 ms and ends at 26.2 ms. The first spike represents the energy of the direct sound arriving at 7.05 ms. This is the time required for sound to travel from the loudspeaker to the microphone. The remainder of the spikes represent room reflections arriving at the microphone position. Figure 16c is an energy-frequency curve (EFC) which shows the comb filtering produced by the early reflections.

The sound absorption panels were placed in the room so as to eliminate individual early reflections. The general areas on the walls and ceiling from which reflected sound was arriving were determined by simple geometry. The exact positions

were confirmed by comparing the ETCs with and without a particular panel in place. Figure 16d shows the initial placement of the panels in the room while Figures 16e and 16f show the respective time- and frequency-domain responses of the room in this configuration. Through trial and error the placement of subsequent panels was determined so as to eliminate all reflections inside the time window within 30dB of the direct sound. This required a relatively small number of panels, as shown in Figure 16g. The ETC and EFC, shown in Figures 16h and 16i, respectively, then confirmed the improved performance of the room at the listening location at the end of this procedure. Having completed the process for one loudspeaker, new absorbent panels were placed in symmetric positions which eliminated the reflections for the other loudspeaker. The improvement in sound imagery was obvious to everyone who listened.

While this fully-treated room provides an idealized reproduction setting, it can also be used with the sound-absorption panels removed, repositioned, or replaced by sound-diffusing panels (D'Antonio & Konner, 1985). This re-enables selected early reflections and permits judgements of sound imagery under non-idealized yet still quantifiable circumstances (Jones, Martens, & Kendall, 1985).

### III. ENVIRONMENTAL SIMULATION

The spatial reverberator algorithm is designed to provide directional cues for a sound source and for environmental sound. Environmental sound provides the acoustical basis for much of the perceptual richness of spatial hearing. A sound source in a natural environment is accompanied by an ensemble of reflections whose intensity, time of arrival and direction are dependent on the position of the sound source and listener within the environment. Figure 17 shows two views of the first-order reflections in a small rectangular room. The upper panel shows a top view of the paths of four first-order reflections bouncing off the walls and arriving at the listener's location. The lower panel shows a side view with the front and back walls and the ceiling and floor. The intensity, time of arrival and direction of these reflections uniquely characterize a pair of source and listener positions in the room. The spatiotemporal distribution of reflections provides the context in which the spatial image is formed. Not only does environmental sound provide support for perceived direction, but it also provides the primary cues to distance, to the spaciousness of the environment and other characteristics of the auditory spatial image.

Figure 18 illustrates how the spatiotemporal distribution changes in response to changes in the location of the sound source. The side wall reflections are leading in intensity and time toward the side on which the sound source is located (toward the right in the upper panel and toward the left in the lower panel). The front and rear reflections shift direction toward the side on which the sound source is located. This sort of pattern is particularly important in supporting perceived direction when other cues are weak such as the cues for distinguishing between front and rear positions. Thus, a primary goal of the spatial reverberator is to provide the kind of spatially distributed reflected sound that will help listeners localize sound.

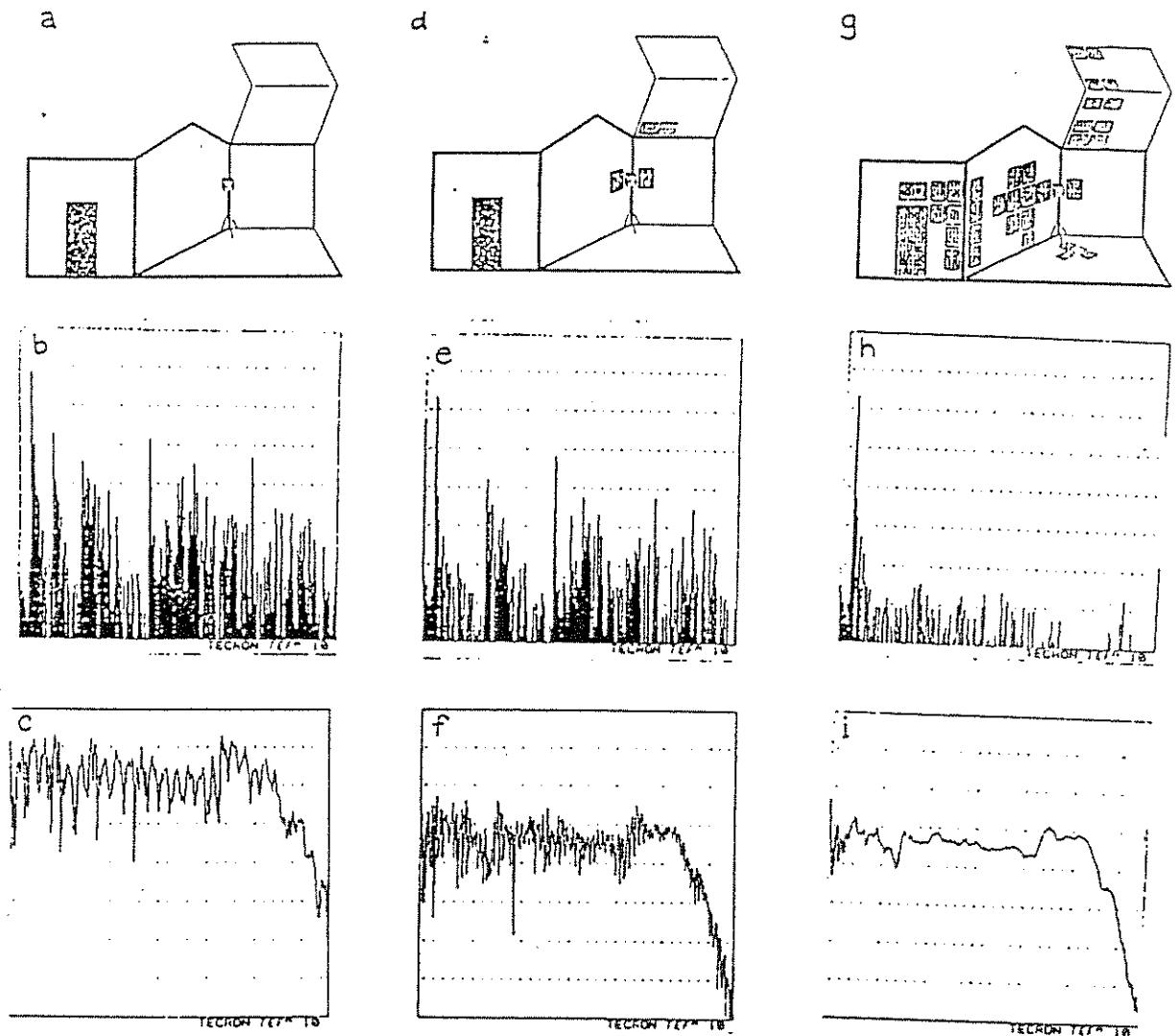


Figure 16. Sound room configuration and measurements: The top panels show an unfolded view of the soundroom with a) no sound absorption, d) the sound absorption placed for the initial measurement, and g) the final configuration of sound absorption on the walls, floor, and ceiling of the soundroom. Panels b,

e, and h show energy-time curves (ETCs) for configurations a, d, and g, respectively. The bottom panels c, f, and i show the corresponding energy-frequency curves (EFCs) for those configurations.

The electronic simulation of reflected sound has been evolving ever since the publication of Manfred Schroeder's (1961, 1962) pioneering articles. Most contemporary approaches to reverberation generation share a number of basic assumptions about the particular quality of the reverberation they intend to produce. First, they attempt to replicate the kind of global reverberation that is typical of large reverberant rooms such as concert halls. Second, they attempt to capture the general characteristics of reverberation without attempting to replicate any of the exact characteristics that distinguish one room from another. Third, they make no attempt to localize the reverberant sound anywhere other than at the loudspeakers. Thus, another goal of the spatial reverberator is that it impart to the listener a strong spatial impression of the exact environment being simulated.

There has been relatively little research on the perceptual effects of environmental sound and consequently the underlying relationships are still not well understood. The stimulus is extremely complex, and it is difficult to isolate the acoustical parameters are most important in predicting what the listener will perceive. The relation between stimulus and response is difficult to address because both the stimulus and the perceptual response to natural environmental sound is multidimensional (Yamaguchi, 1972). Though many terms have been used by researchers of subjective room acoustics to describe the various properties of spatial images (see Rasch and Plomp, 1982, for a review), Kendall and Martens (1984) identify only four basic properties in addition to source direction which are important in audio production and reproduction:

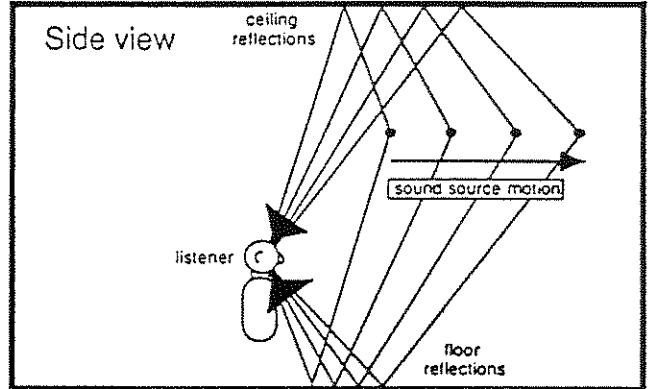
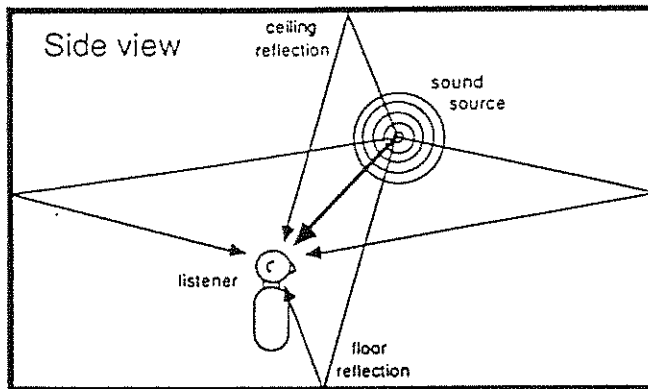
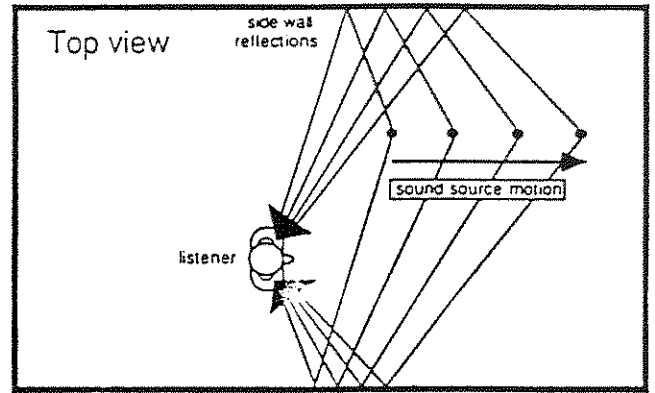
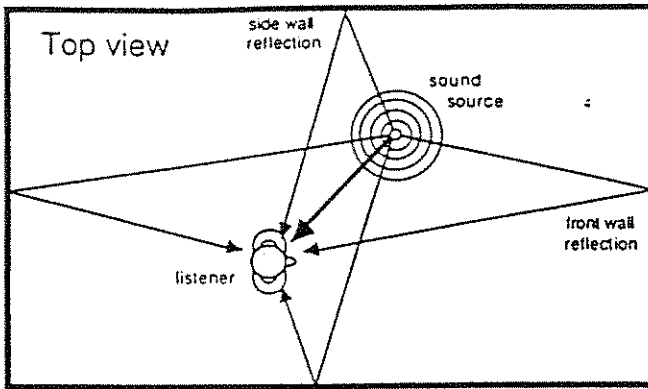


Figure 17. A top view (top panel) and a side view (bottom panel) of discrete early reflections arriving at the listener's location, each with a distinct angle of incidence and delay relative to the direct sound.

Figure 18. A top view (top panel) and a side view (bottom panel) of changes in early reflections that occur as a sound source moves away from the listener through four distinct distances.

- 1) Distance - the immediate percept of ego-centric distance;
- 2) Definition - perception of sound-source characteristics such as spatial extent or focus;
- 3) Spaciousness - perception of environmental characteristics such as liveness, size, and shape;
- 4) Spatial texture - the perception of the interaction of the sound with its environment.

**Distance.** The best understood cue to apparent distance is the level ratio of the direct sound arriving from the source and the indirect sound arriving from the environment. Generally, as a sound source moves further away from the listener, the level of the direct sound diminishes while the reverberation level stays constant. This is illustrated in the graph at the bottom of Figure 19. Thus, the indirect to direct ( $i/d$ ) ratio is a good predictor of apparent distance. A signal processing network to control this parameter is shown at the top of Figure 19. It provides simultaneous and continuously-variable control over the intensity and the apparent distance of the sound source.

This simple perspective glosses over factors that make significant contributions to auditory distance perception, especially in non-reverberant environments which are dominated by a few discrete early reflections. Sheeline (1982) showed

that  $i/d$  with a reverberant sound field provides only relative cues to apparent distance, in contrast to the absolute sense of distance listeners experience in everyday listening. Mershon & King (1975, 1979) showed that absolute distance judgments were very nearly accurate, even in a small space with relatively short reverberation time.

Figure 20 illustrates why this could be the case. As the sound source moves away from the listener there are changes in the spatiotemporal pattern of reflections. The propagation time from source location to listener position increases (as shown by the nearest four concentric circles) as does the propagation time for the first reflection off the left side wall (as shown by the outermost concentric circles). But the propagation time for the reflection does not grow as quickly as for the direct sound. The initial time gap between the direct and reflected sound will decrease with increasing source distance as is shown in the four graphs at the bottom of the figure. (The initial time gap has often been regarded as a cue only to the size of the space in which the listener is located.) A sound source moving away from the listener is also accompanied by some reflections that become less lateralized, shifting toward the source as it recedes into the distance. Thus, the apparent direction is directly influenced by discrete reflections.

The initial software for the spatial reverberator utilized Allen and Berkeley's (1979) image model of room reverber-

Indirect to Direct Ratio  
Cues Distance

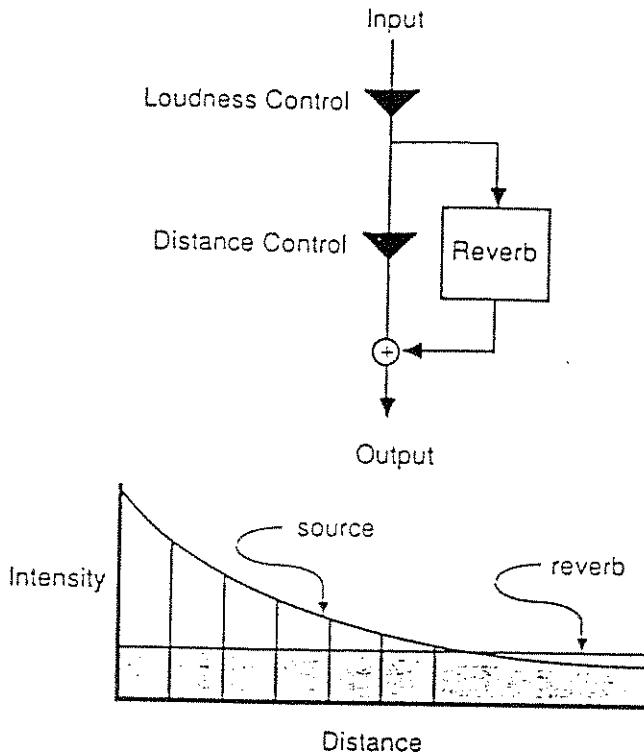


Figure 19. A signal processing network to control apparent distance independent of the loudness of the input sound. The graph at the bottom shows how the intensity of the source decreases with distance relative to a reverberation level that does not vary with the distance of the sound source.

ation to produce a comprehensive and coherent description of early reflection patterns in a simulated rectangular room. Figure 21 shows a top view of how a second-order specular reflection within a room is mapped through multiple mirror image rooms to reveal a virtual source location. If striking the wall had no effect upon the reflection, then there would be no basis for discriminating whether the sound arriving at the listener had been reflected within the model room (dark grey path), or had originated at the virtual source position outside the room (light grey path). Each  $n$ th-order reflection is captured by a virtual image of the sound source in a virtual room that is an  $n$ th-order mirror image. The spatial reverberator effected continuous variation of the intensity, delay, and direction of all first-order and second-order reflections as the specified location of the sound source and/or the listener was varied (Kendall & Martens, 1984). This image model is a simple method of approximating environmental sound and can be extended to solve for the reflection patterns resulting within polyhedral enclosures of arbitrary configuration (Borish, 1984).

**Definition and Spaciousness.** These two subjective properties of room acoustics are often described as complementary to one another; that is, a high degree of definition precludes a high degree of spaciousness (Rasch & Plomp, 1982). Definition is generally defined as the property of a

Initial Time Gap Decreases  
with Increasing Distance

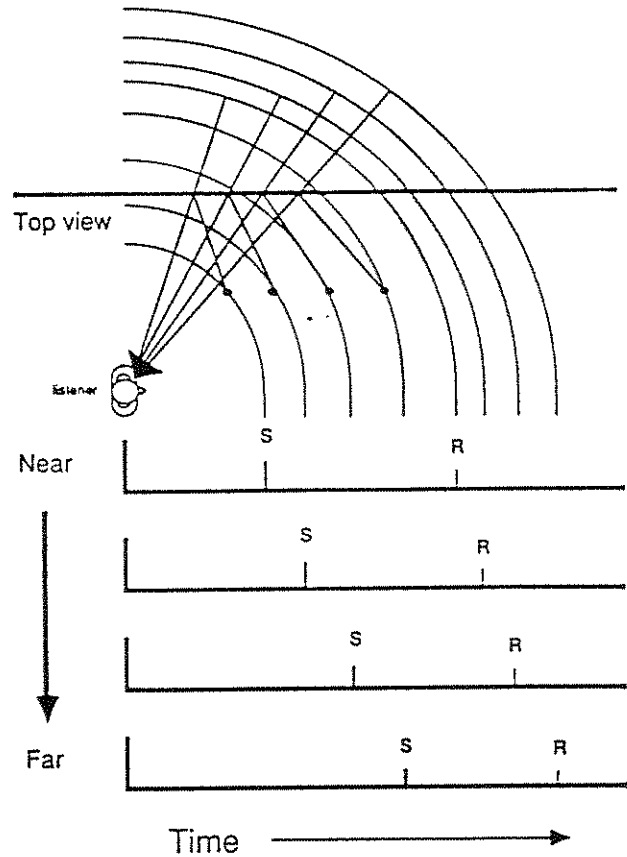


Figure 20. A top view of changes in a single early reflection that occur as a sound source moves away from the listener through four distinct distances. Differences in the propagation time for sound reaching the listener directly from the source and from a side-wall virtual source (represented by the concentric circles centered at the listener's position). The time of arrival of direct sound and reflected sound are graphed below for each of the four source distances.

sound object having to do with the width or focus of its image (also referred to as "spatial extent" by Blauert, 1982). Spaciousness is defined as a property of the sound environment itself. There are many terms that have been used interchangeably with spaciousness such as "ambience," "presence," and "reverberance." They all refer to the perceived size of the space and how live or reverberant it seems.

The apparent size of the space is most strongly influenced by the reverberation time, while the apparent shape is determined by the spatial distribution of reflected sound. The sense of breadth is most strongly influenced by the interaural cross-correlation (IACC) of the two ear's signals (Kurozumi & Ohgushi, 1983), but all three of these factors interact in forming the overall impression of the space. For example, apparent size can be manipulated without any change in the reverberation time through compensatory adjustments of the spatial distribution and IACC. Definition also depends strongly upon IACC. The degree to which the definition of the source is separated from the spaciousness of the environment, depends on the degree to which the lis-

## Second-order Reflection

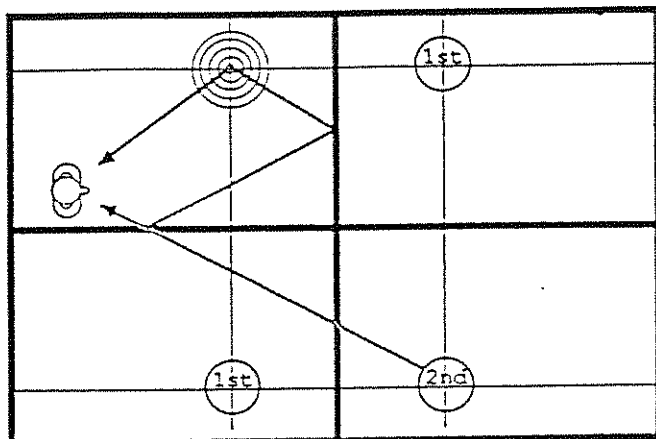


Figure 21. A top view of virtual rooms and the path taken by a virtual source whose spatial position can be used to calculate the direction and delay of a second order reflection.

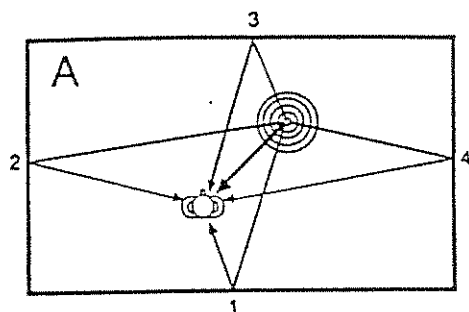
tener is able to segregate the sound image of the direct sound from reflected sound.

**Spatial texture.** Our experience with the spatial reverberator has led us to expand the perceptual dimensions described in the room acoustics literature. We have identified an additional dimension which we refer to as "spatial texture" which captures the quality of the sound source within its environment. Although this property is a function of environmental sound (like definition), it is perceived as a quality of the sounding object interacting with its environment. It is a quality imbued to the sounding object due to its position in the environment and changes with different locations and with different kinds of rooms. We can give it the kind of negative definition often given to timbre: Whereas timbre has been described as the quality which differs between two tones having the same pitch and loudness, spatial texture may be described as the quality which differs between two spatial images having the same distance, definition, and spaciousness. Figure 22 illustrates a situation in which a sound source revolves 90 degrees around the listener in a fixed environment at a fixed distance. But the spatiotemporal pattern of the reflected sound changes constantly and induces changes in spatial texture. For example, the idiosyncratic reflection patterns that occur in the corners of rooms creates an idiosyncratic spatial texture.

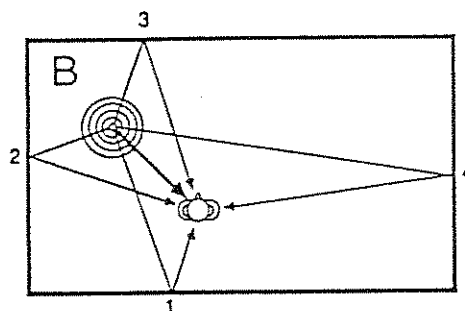
## IV. CONCLUSIONS AND OBSERVATIONS

At the beginning of this paper, it was stated that the goal of this project was the "creation of auditory spatial images that have all the complexity and richness of everyday experience in natural environments." To that end, we have discussed a computational model called the "spatial reverberator" which incorporates two essential elements: directionalization and environmental simulation. A process has been described whereby idealized directional transfer functions (DTFs) are created through application of principal components analysis. These DTFs can deviate quite substantially from measured HRTFs and still provide the auditory system with all the infor-

## Moving Sound Source



Sound source rotates counter-clockwise around stationary listener



Spatial and temporal pattern of reflections shift

Figure 22. A top view of changes in a early reflection patterns that occur as a sound source moves around the listener's position at a fixed egocentric distance.

mation necessary to the formation of a robust spatial image. Also described has been the perceptual importance of simulating the spatiotemporal distribution of reflected sound. This simulation supplies the richness of spatial imagery typical of real world experience. The combination of these two elements within a digital sound processing device enables sound artists and recording engineers to create auditory images that capture the quality of "first hand" experiences and strongly suggests the development of new artistic idioms.

## V. REFERENCES

- Allen, J. B., & Berkeley, D. A. (1979) Image model for efficiently simulating small room acoustics. *J. Acoust. Soc. Am.*, 65, 943-950.
- Batteau, D. W. (1967) The role of the pinna in human localization. *Proceedings of the Royal Society of London*, 168 (series B), 158-180.
- Blauert, J. (1983) *Spatial hearing*, trans. J. S. Allen. MIT Press (Cambridge, Mass.) 1983. Originally entitled *Ra'o(u,)mliches horen*, S. Hirzel Verlag, Stuttgart, 1974.
- Borish, J. (1984) Extension of the image model to arbi-

rary polyhedra. *J. Acoust. Soc. Am.*, 75, 1827-1839.

Buder, R. A., & Belendiuk, K. (1977) Spectral cues utilized in the localization of sound in the median sagittal plane. *J. Acoust. Soc. Am.*, 61, 1264-1269.

D'Antonio, P., & Konner, J. H. (1985) The role of reflection phase grating diffusers in critical listening and performing environments. Presented at the 78th AES Convention, Anaheim.

Fisher, S. S. (1989) Virtual environment, personal systems, and telepresence. In: H. Thwaites (Ed.), *Proceedings of the 3D Media Technology Conference*, Montreal.

Gardner, M. B., & Gardner, R. S. (1973) Problem of localization in the median plane: Effect of pinna cavity occlusion. *J. Acoust. Soc. Am.*, 53, 400-408.

Hiranaka & Yamasaki (1983) Envelope representations of pinna impulse responses relating to three-dimensional localization of sound sources. *J. Acoust. Soc. Am.*, 73, 291-296.

Jones, D., Kendall, G., & Martens, W. (1984) Designing a sound room for auditory research using the TEF. Poster presented at the 1984 International Computer Music Conference, Paris, October, 1984.

Jones, D., Martens, W., & Kendall, G. (1985) Optimizing control rooms for stereo imagery. Paper presented for the Acoustical Society of America, Nashville, Tennessee, November 1985.

Kendall, G. S., & Martens, W. L. (1984) Simulating the cues of spatial hearing in natural environments. In: W. Buxton (Ed.), *Proceedings of the 1984 International Computer Music Conference*, Paris, Oct.

Kendall, G. S. & Rodgers, C. A. P. (1982) The simulation of three-dimensional localization cues for headphone listening. *Proceedings of the 1982 International Computer Music Conference*.

Kurozumi, K., & Ohgushi, K. (1983) The relationship between the cross-correlation coefficient of two-channel acoustic signals and sound image quality. *J. Acoust. Soc. Am.*, 74, 1726-1733.

Lackner, J. R. (1983) The influence of posture on the spatial localization of sound. *J. Aud. Eng. Soc.*, 31, 650-661.

Martens, W. (1987) Principal components analysis and resynthesis of spectral cues to perceived direction. *Proceedings of the International Computer Music Conference 1987*, S. Tabei & J. Beauchamp, eds.

Morimoto, & Ando (1983) On the simulation of sound

localization. *J. Acoust. Soc. Jap.*, 74, 873-887.

Mershon, D. H. & Bowers, J. N. (1979) Absolute and relative cues for auditory perception of egocentric distance. *Perception*, 8, 311-322.

Mershon, D. H. & King, L. E. (1975) Intensity and reverberation as factors in auditory perception of egocentric distance. *Perception & Psychophysics*, 18, 409-415.

Petersen, T. L. (1980) Acoustic signal processing in the context of a perceptual model. PhD. dissertation, Department of Computer Science, University of Utah.

Rasch, R. A. & Plomp, R. (1982) The listener and the acoustic environment. In: D. Deutsch (Ed.), *The Psychology of Music*, New York: Academic Press, 135-147.

Schroeder, M. R. (1961) Improved quasi-stereophony and "colorless" artificial reverberation. *J. Acoust. Soc. Am.*, 33, 1061-1064.

Schroeder, M. R. (1962) Natural-sounding artificial reverberation. *J. Aud. Eng. Soc.*, 10, 219-223.

Sheeline, C. W. (1982) An investigation of the effects of direct and reverberant signal interaction on auditory distance perception. PhD dissertation, Department of Hearing and Speech Sciences, Stanford University.

Stautner, J. P. (1983) Analysis and synthesis of music using the auditory transform. Masters Thesis, Department of Electrical Engineering and Computer Science, MIT.

Sunier, J. (1986) A history of binaural sound. *Audio*, March, 36-46.

Wallach, H. (1940) The role of head movements and vestibular and visual cues in sound localization. *J. Exp. Psychol.*, 27, 339-368.

Wenzel, E. M., Wightman, F. L., & Foster, S. H. (1988) Development of a three-dimensional auditory display system. Paper presented at the Computer-Human Interaction Conference, May 15-19, Washington D.C.

Wightman, F. L. & Kistler, D. J. (1989) Headphone simulation of free-field listening. II: Psychophysical validation. *J. Acoust. Soc. Am.*, 85, 858-867.

Woodworth, R. S. (1954) *Experimental Psychology*, Revised Edition. New York: Holt, Rinehart, & Winston.

Yamaguchi, K. (1972) Multivariate analysis of subjective and physical measures of hall acoustics. *J. Acoust. Soc. Am.*, 73, 291-296.