AFIT/GAM/ENG/93D-1

Binaural Room Simulation

THESIS

Brian A. Smith
1LT, USAF

AFIT/GAM/ENG/93D-1

Approved for public release; distribution unlimited

AFIT/GAM/ENC/93D-1

Binaural Room Simulation

THESIS

Presented to the Faculty of the School of Engineering
of the Air Force Institute of Technology
Air University
In Partial Fulfillment of the
Requirements for the Degree of
Master of Science (Mathematical Statistics)

Brian A. Smith, B.S. Chemistry
1LT, USAF

December, 1993

Approved for public release; distribution unlimited

Acknowledgements

In most thesis papers, the acknowledgements section is typically one of the shorter sections in the document. In my case, I would have no trouble making it the longest section; but in staying with tradition, I will keep this as short as possible. If not for the help of dozens of people, this work would have never been completed. In thanking the many people who helped me, I will start at the very top. No one deserves more credit for this work than the Good Lord–Jesus Christ. It is only by the grace of God that I had this opportunity. In His great wisdom He provided me with an ideal research topic and support group.

My thesis committee, composed of Dr. Mark Oxley, Barbara McQuiston, and Dr. Steve Rogers, as well as employees of Armstrong Aerospace Medical Research Laboratories–Marc Ericson, Lt. Bill D'Angelo, and Richard McKinley–were the source of ideas, and technical expertise for the research in this thesis. A number of Electrical Engineering students were very helpful in teaching me better ways to use the Sun Workstations in the Signal Processing Lab. I would like to thank John, Neil, Curtis, Gary, Bob, Kim, and Bill for all their help. Also, Capt John Colombi was extremely helpful with answering my questions regarding the use of ESPS. Also, thank you to Janet Slifka for her help in developing the filters for this research. If not for the contributions of each of these individuals, my work load would have been even heavier. Finally, I would like to thank my fiancee, Karen, for her support throughout this research effort. Her willingness to spend her weekends at AFIT made the task of completing this work possible.

Brian A. Smith

Table of Contents

Page

List of Figures

List of Tables

AFIT/GAM/ENG/93D-1

Abstract

Research in binaural and spatial hearing is of particular interest to the Air Force. Applications in cockpit communication, target recognition, and aircraft navigation are being explored. This thesis examines human auditory localization cues and develops a mathematical model for the transfer function of a sound signal traveling from an isotropic point source through a rectangular room to both ears of a listener. Using this model as a guide, non-head coupled binaural sound signals are generated in a binaural room simulation. Reflection and attenuation cues included in the computer generated signals are varied in order to determine which cues enhance the listener's degree of extracranialization. Results of this research indicate that the addition of three or more attenuated reflections into a non-head coupled binaural signal provide the listener with a binaural sound that is localized extracranially.

Binaural Room Simulation

I. Introduction

1.1 Background

Human auditory localization is the ability of a human listener to acoustically identify the azimuth, elevation, and range of a sound source (19:1). A human's ability to locate a sound stems primarily from the fact that humans hear with two ears. In a process called binaural fusion, the brain extracts information from the sound at each eardrum to aid in determining the exact location of a sound source (15:66). In the ideal reproduction, a listener hears the signal played through headphones and perceives the sound as eminating from a three-dimensional point in space (23:171-172).

The development of a system that enables humans to accurately localize sound in three-dimensional space is of particular interest to the Air Force (19) (32). The implementation of an auditory localization system into future Air Force cockpits would provide pilots a new source of spatial information that might assist them in performing complicated tasks intuitively. For example, the otherwise tedious task of keeping track of the location of a wingman could be accomplished simply through auditory localization of radio transmissions. Crew aircraft would also benefit from the use of this technology. A pilot, co-pilot, navigator, and bombadier could all identify who is speaking through spatial sound localization. Other applications exist in the areas of target recognition, and aircraft navigation (20) (38).

The foundation for an accurate auditory cue synthesizer has been under construction for more than sixty years (23:172). During this time, theory and implementation of the theory have evolved continuously. When a sound is generated and travels to a listener, the acoustic signal that is received at the listener's eardrums is significantly different from the acoustic signal that left the source. These changes result from the acoustic signal's interaction with the acoustic environment, and from the design of the human head and ears. (32) (19) (18) (23).

Past research indicates that the design of the human head acts to transfer the original sound from a given direction and distance into two signals that are interpreted by the brain (33). If the transfer functions to each ear can be found, they can then be modeled mathematically and simulated electronically as filters (23:208). If the transfer functions for all locations were known, it would be possible to artificially place any sound at any location in three-dimensional space.

This thesis mathematically models the overall transfer function for sound traveling from a source to a listener in a rectangular room and implements the model electronically on a SUN Sparc 2 Workstation using the Entropic Signal Processing System (ESPS) designed by Entropic Research Laboratories and copyrighted in 1992. Models for this type of simulation already exist; however, research in this area is still necessary since there is still no real-time working system available (11) (4). This thesis develops a non-real-time binaural room simuiation, and consequently does not incorporate head motion. Past research indicates that head coupling is essential to extracranialization of synthesized sound. This thesis attempts to improve upon existing auditory cue synthesizers that do not account for head motions by ensuring that the listener does not experience the common problem of in-head localization. With the in-head localization problem, a listener perceives the sound as coming from inside the headphones or from some location along the surface of the head (intracranial lateralization) as opposed to coming from the distance simulated by the electronic filter (extracranialization) (23:173). At a recent conference on Binaural and Spatial Hearing, the importance of research in the area of in-head versus out-of-head localization was emphasized (8).

1.2 Definitions

This section provides definitions of key terms that will be used in this thesis.

Binaural sound is sound that arises from two separate audio signals–one at each ear. The signal that arrives at the left ear is different from the signal at the right ear. The brain uses the differences between the two signals to determine the direction and distance of the source. When sound is played through headphones, the sound is binaural if the signal sent to the left headphone is different from the signal sent to the right (32:3). When binaural sounds are presented without incorporating head motion, the listener normally can lateralize the sound intracranially. Binaural sounds synthesized in this thesis are non-head coupled.

Binaural room simulation is a simulation which incorporates cues from the acoustic environment into the synthesized signal. In this thesis, these cues are added to s[illegible]te a sound traveling from a specific location in a rectangular room to each ear of a listener at another specific location in the same room (23:209).

Telepresence is the degree to which a human perceives that he or she is present in a natural environment even though the environment is artificial or virtual.

Virtual audio is synthetically produced audio signals which enable a listener to acheive auditory telepresence. Virtual audio differs from binaural sound in that listeners experiencing virtual audio can localize sound extracranially as opposed to simply lateralizing the sound.

Extracranialized sound is sound that is presented binaurally to a listener and is perceived as coming from some distance from the listener–i.e. outside the head.

Pinna(e) is(are) the human outer ear(s). The design of each person's pinnae is unique, and each set of pinnae transform sounds differently. In fact, the human head and ears form an antenna system characteristic to the individual (33). This system is frequency and angle dependent.

Head related transfer function (HRTF) is the transfer function which accounts for the filtering effects of the pinnae. Because the filtering of sound by the pinnae is dependent upon direction, there is a different HRTF for each angle of incidence. The angle of incidence is composed of an azimuth angle and an elevation angle. HRTF's will differ for each persons pinnae. Accurate localization of sound sources can be made by some subjects using "someone else's pinnae" (42). In nearly all cases, however, listeners localize better using their "own ears".

Auditory Localization Cues are anything that allows a listener to determine the location of a sound source. Several cues are addressed in this thesis. These cues are Interaural Time Delay (ITD), Interaural Intensity Difference (IID), Pinna Effects, Head Motion, Reflections, Reverberation, Attenuation, Sound Soure Directivity, and Doppler.

1.3 Problem

Auditory localization cue synthesizers with head coupling produce extracranialized sound, but the distance cue is not controllable. Systems which do not account for head motion fail to produce extracranialized localization of binaural sound.

1.4 Research Objectives

This thesis develops a complete mathematical model representing the overall transfer function for a sound traveling from a source to a listener. This transfer function accounts for direct path and all reflected paths in a rectangular room. The model is implemented electronically in the form of a binaural room simulation to generate a binaural signal that is extracranialized

1.5 Scope

This thesis develops a binaural room simulation that enables a listener to accurately determine the azimuth, elevation, and range of a sound source presented over headphones without head coupling. A mathematical model is developed for an arbitrary sized rectangular room with each of the four walls, the ceiling, and the floor having a known absorption coefficient. By simulating a room, cues from the acoustic environment can be included in the binaural signal. This creates a more realistic auditory experience for the listener and results in extracranialization of synthesized sound. The coefficient of absorption of the air is represented by a constant for a given temperature. The sound source is modelled as a point source emitting isotropic sound and the listener is assumed to have two ears-each with a characteristic HRTF-separated by a given distance. The case of a stationary source and listener is examined. This is not a real-time simulation due to the extensive amount of processing required to obtain the final binaural sound signals.

Armstrong Aerospace Medical Research Laboratories(AAMRL) provides the HRTF and ITD data used in this simulation (19). The headphones used are intra-aural or insert Etymotic ER-2 headphones; standard Walkman type headphones are used on a comparative basis. Monophonic sound recordings are made and processed in time using filters representing the transfer functions for the head and pinna. The signals are processed in time using the ITD data. Furthermore, the cues resulting from the acoustic environment (reflections, and attenuation) are included in the simulation in varying degrees in order to determine how each cue affects the extracranialization of the binaural sound signal.

1.6 Approach

The approach to this research problem consists of two major parts, 1)mathematical modeling of the overall transfer function, and 2)electronically generating the signal described by the mathematical model by means of a binaural room simulation.

1.6.1 Mathematical Approach. The literature will be reviewed; existing models will be adopted into this thesis' model and modified appropriately. The transfer functions for the headphones and the head are obtained from direct measurements. Both of these transfer functions along with the room transfer function are combined to get a single overall transfer function for sound traveling from a point source to each of the two ears of a listener.

1.6.2 Computational Approach. Portions of the mathematical model are implemented through digital signal processing using a Sun Sparc 2 workstation and the ESPS software. Filters designed from the AAMRL HRTF data are used to synthesize sounds from different angles. The transfer function for the headphones is ignored in the implementation of the mathematical model since the transfer function of the headphones to be used is essentially flat (9). Finally, reflections and attenuation are included in the binaural room simulation. Different combinations are examined to determine which factor(s) enhance the extracranialization of the sound.

1.7 Thesis Outline

The following chapter in this thesis is a literature review of many of the topics mentioned in Chapter I. Chapter III details the mathematical analysis done in this thesis. Chapter IV describes the methods used for implementing the mathematical models on the computer. Chapter V provides the results of this research effort, and Chapter VI summarizes this thesis and presents recommendations for follow-on research in sound localization.

II. Literature Review

2.1 Introduction

In this chapter, a literature review of three main areas will be accomplished. The first section will cover the changes that occur in the signal as it reaches the receiver; these are the free-field cues to human auditory localization. Next, a review of changes in the characteristics of sound as it propagates from source to receiver will be accomplished. The focus in this section will be on room acoustics and the changes that occur in a sound signal as it travels through an acoustical environment. The concluding section will review the aspects of sound as it leaves the source; topics include sound source directivity and doppler.

2.2 Receiver Cues in Auditory Localization

2.2.1 Introduction Receiver Cues. The fundamentals of human auditory localization in the free-field consist of four primary cues. Two cues– interaural time difference (ITD) and interaural intensity difference (IID)–(Note: IID is also called interaural pressure level difference (ILD) in some references)–aid the brain in determining the angle of azimuth to a sound source. ITD and IID emerged from the duplex theory of sound localization; the critical element of this theory is the human's use of two ears to locate sound sources (39) (28) (29). IID and ITD work together in providing the brain with localization information. The third cue is the effects of the pinnae; the pinnae accentuate and suppress various frequencies as a function of the azimuth and elevation of the sound source. The fourth cue is head motion; this cue is believed to be a cue used by humans to disambiguate front/back and up/down confusions (42) (19) (20). The combination of these four cues provides the brain with the necessary information to accurately localize a sound source (36:21-22) (19) (20).

2.2.2 Interaural Time Difference. Past research has established that interaural time difference is the primary localization cue in the azimuthal plane at low frequencies (17:157). ITD is a binaural, temporal cue; in general, research indicates that both the monaural and binaural temporal cues are most salient in the azimuthal plane (21). Low frequency ITD's are generally larger than higher frequency ITD's, and therefore provide a stronger cue (17:160). The exact frequency at which ITD is no longer the major contributor to localization is not clear, but studies have consistently found that this frequency is approximately 1500 Hz. In his study, Kuhn concludes that the trade-off frequency where IID becomes more important to localization than ITD is 1400 Hz. Furthermore, Kuhn finds that there is steady improvement in the ITD cue as the frequency is lowered from 1400 Hz to 500 Hz; but below 500 Hz, there is no improvement in the available ITD cue (17:162).

In a separate study, Abbagnaro *et al* reach similar conclusions. They find that ITD is a function of the geometrical interaural distance between the two ears, and that ITD decreases gradually as frequency increases becoming relatively constant at frequencies above 1000 Hz (1:700). In research done at AAMRL, McKinley and Ericson physically measured the time differences of arrival for sounds of various frequencies. They found that time difference of arrival ranges from 0 to 750 microseconds depending upon the angle of arrival in the horizontal and vertical planes and the frequency of the sound (19:56-63). In experiments which have examined the effects of ITD, listeners wearing headphones have been presented with a binaural signal that contained a time delay to one ear. Subjects percieved that the sound signal originated from the direction corresponding to the ear that received the signal first (31). These types of experiments clearly established ITD as a major cue in sound localization; the following list provides a summary of the important information concerning ITD.

- Interaural Time Difference is a function of the distance between the two ears, the angle of incidence of the incoming sound, and the frequency of the sound.

- ITD is frequency independent below approximately 500 Hz and above approximately 3000 Hz; furthermore, between these frequencies ITD decreases as frequency increases. For azimuth angles of incidence less than 60 degrees left or right of directly in front of the listener (0 degrees), the minimum ITD occurs between 1400 and 1600 Hz. (17:165-166).
- The most significant changes in ITD occur at angles of incidence between 0 and 30 degrees (1:700).

2.2.3 Interaural Intensity Difference. A second concept, which also developed from the duplex theory, is Interaural Intensity Difference (IID). Like ITD, IID results directly from the fact that humans have two ears and is therefore a binaural, spectral cue. Past research indicates that due to head shadowing, the intensity of the sound at one eardrum is different from the intensity at the other (39:83). Furthermore, IID becomes the primary cue in azimuthal localization as frequency increases (17:162). In a recent experiment *testing* sound-pressure levels in the human ear canal, Middlebrooks developed contour diagrams which indicated areas of high and low amplitudes. In this experiment, Middlebrooks is primarily interested in the variation in amplitude of single frequency components as a function of sound source location–he calls this the directionality of the sound (22:93). Results of this experiment show that the patterns of directionality are different at different frequencies (22:97). Furthermore, results suggest that the size and shape of a listener's torso (affects primarily low frequencies), head, and ears affect the amplitude of the signal at different frequencies (22:101). At a recent conference on binaural and spatial hearing, Middlebrooks emphasized the importance of spectral cues in localizing elevation angles as well as in aiding the disambiguation of front/back and up/down reversals (21).

In their research, Abbagnaro *et al* find similar results. They find that at the ear receiving the sound directly the amplitude response sharply increases at high frequencies when the angle of incidence is increased; the maximum rise is observed to be at an angle of incidence of 60 degrees of azimuth which corresponds to when the pinna is in approximate confrontation with the sound source (1:699). Furthermore, they find that as the head continues to turn the response decreases due to shielding by the pinna. At the shadowed ear, the response decreases until approximately 120 degrees where it reaches a minimum (1:700). The following list is a summary of the important information concerning IID.

- Interaural Intensity Difference is a function of the distance between the two ears, the angle of incidence of the incoming sound, and the frequency of the sound. IID can also be affected by size and shape of the torso, head, and ears. The torso affects primarily the low frequencies.
- The most significant changes in IID occur at angles of incidence between 0 and 30 degrees, consistent with the fact that humans perceive a sound best when the source is directly in front of the listener (1:700).
- Improvement in localization capabilities of humans at frequencies above 3000 Hz is a result of improvement in the IID cue. (17:166).
- Although IID provides elevation cues at low frequencies, at frequencies above 8000 Hz, IID cues substancially increase human capabilities to localize in elevation as well as in azimuth (22:101).

2.2.4 The Effects of the Pinnae. Under the assumption that the head is a stationary sphere with symmetrically located ear canals and no pinnae, a given ITD or IID will place the sound source on a conical shell representing all possible locations (39:85-86). In reality, however, the head is not spherical, it does have pinnae, and it does move. To model the head without these important features causes ambiguity in the binaural signals presented to listeners–i.e. front/back and up/down reversals (39:86). Sound arriving at each ear follows a unique path, and part of that path includes the pinna. The pinnae filter mid- and high-frequencies of incoming sound as a function of angle of incidence (36:22). The filtering properties of the human outer ear significantly change the signal that is processed by the brain; these changes, in fact, contain a great deal of information about the location of the sound source (10:18). As a result, the pinnae effects when combined with the ITD, IID, and head motion cues allow accurate localization of sound sources. In addition, the pinna cues seem to enhance the extracranialization of the sound (39:84).

Since the goal of this thesis is to extracranialize sounds, the importance of adding the pinnae cues is clear. There are a variety of different effects on sound as it travels through the pinna to the eardrum. The spectral shaping that occurs as sound travels this path can be accounted for by a group of filters which are dependent upon the incidence angle and frequency–the Head Related Transfer Functions (HRTFs) (39) (19) (32). Measurements of head related transfer functions can be made by producing an impulse at a specific location in an anechoic chamber and measuring the output of microphones placed inside a human subject's or manikin's ears.

In 1989, Wightman and Kistler successfully measured the filtering properties of the HRTF using digital signal processing (41) (40). In an earlier experiment, McKinley and Ericson successfully measured the HRTF's for 272 different angles of azimuth and elevation (19). These HRTF's will be used in this thesis effort. McKinley and Ericson used an acoustically accurate manikin–Knowels Electronics Manikin for Acoustic Research (KEMAR)–to make their initial measurements of the HRTF's. The orientation of the KEMAR manikin for the measurements is shown in figure 1 and figure 2.

90th percentile pinnae and 50th percentile head and torso were used in the experiment (19:21). A list of the speaker locations used in these measurements is found in Appendix B, and examples of the HRTFs used in this research are found in Appendix E. Using the filters developed by McKinley and Ericson, the filtering effects of the human outer ear can be accounted for in the binaural room simulation.

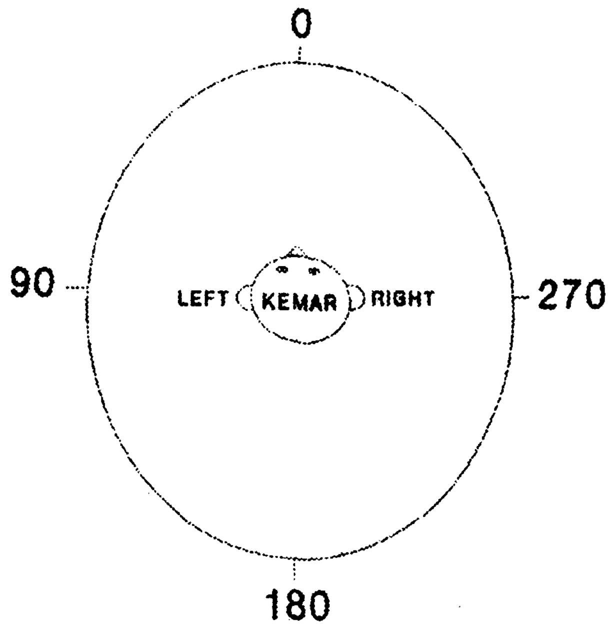

Figure 1. Orientation for Azimuth Data

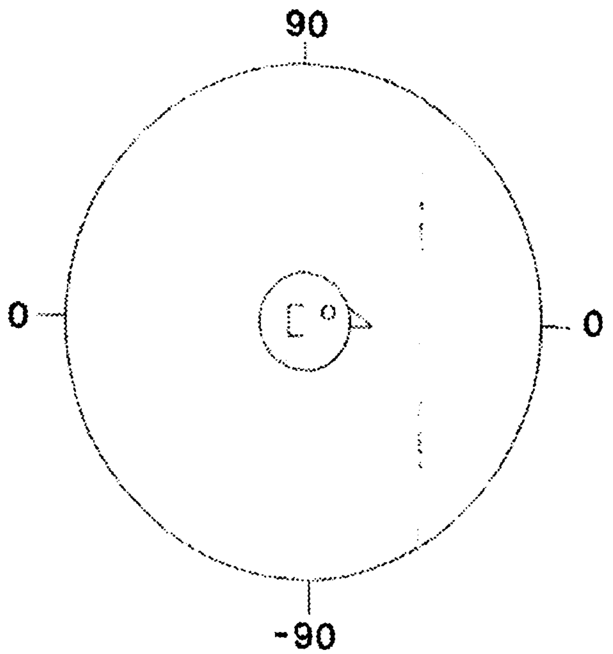

Figure 2. Orientation for Elevation Data

2.2.5 Head Motion. The importance of head motion in human auditory localization, specifically in extracranialization of sound, cannot be overstated (19) (20). Without head motion, a person's head and ears can still act as a directional antenna; however, sounds located at 0 degrees azimuth and various elevations (i.e. on the medial plane) can be extremely difficult to localize. Front/back and up/down confusions result when head motion is not allowed; this occurs because the spectral and temporal cues are very similar at both ears (42) (3). The implementation of this cue into a binaural room simulation requires a real-time system be used. For this reason, head motion will not be implemented into the non-real-time simulation done for this thesis.

2.2.6 Conclusion Receiver Cues. There are many complex cues used by the human brain to determine the location of a sound source. The four primary cues in the free-field are interaural time difference (ITD), interaural intensity difference (IID), effects of the pinnae, and head motion. ITD is the most prominent cue at low frequencies while IID becomes more prominent at high frequencies. The pinnae filter mid- and high-frequencies of incoming sound to provide the brain with localization cues, and head motion aids in localization on or near the medial plane. Together, these four cues act to provide enough information to the brain for listeners to accurately localize sound sources in three dimensional space.

2.3 Room Acoustics

2.3.1 Introduction Room Acoustics. In a normal environment, it is clear that in addition to the free-field cues there are many cues that occur due to the listening environment or medium. Since current technology has yet to produce a binaural listening system without head coupling which allows the listener to extracranialize incoming sounds, it is unclear what characteristics of the sound must be added to solve the in-head localization problem. Current research points to several key cues that arise as a result of room acoustics; these cues result from the sound's interaction with the medium (a room in this case) prior to reaching the receiver. As sound propagates in a room, the changes that occur in the signal that is received aid the listener in localizing sounds. Therefore, room acoustics will be the focus of this section.

2.3.2 Reflections. An acoustical environment is defined by a set of physical properties which determine the changes that will occur in a sound as it excites this space. The task in a binaural room simulation is to model the environment defined by these properties (18:260). The environment may range from an anechoic chamber in which there no reflections to a reverberation room in which there are a large number of reflections. In an anechoic environment, listeners localize sound sources very well, but they are unable to judge distance. In a reverberation room, listeners are unable to localize sound sources. In a normal rectangular room, the environment lies somewhere in between the two extremes. Listeners receive direct source information, but they also receive a variety of reflected sound signals from different directions–each containing different spectral and temporal cues. The brain suppresses many of the early reflections and hears the array of sounds as one entity; this is explained by the precedence effect (13) (12). The information that is contained in these early reflections; however, contains information about the listening environment (3) (6).

The reflections must somehow be represented. The best way of accounting for the reflected sounds is to represent each reflection as a reflected sound source. Assigning a linear filter to each reflected source and delaying the signal appropriately defines the spectral and temporal differences specific to that reflected source. Due to the limitations of the human auditory system (spatial and temporal resolution) and the precedence effect, any sound field can be simulated by a finite number of reflected sound sources (18:261-264). Clearly, the more reflecting surfaces in the acoustic environment, the greater the number of reflected sound sources required to define its reflecting properties. The two common methods used in determining the location of the reflected sound sources are the ray-tracing method and the image source method.

2.3.2.1 Ray-Tracing Method for Reflections. The ray tracing model for determining the location of reflected sound sources in a room is based on geometrical acoustics. Sound sources emit a conical beam of rays; these rays pass in straight lines through air (air can be considered a homogeneous, isotropic medium) and are reflected geometrically when striking a reflective surface (37) (26) (16) (18). Each of the rays that is emitted by the source carries energy, and travels at the speed of sound (37:173). The ray tracing technique consists very simply of emitting rays from a sound source, following their paths through a room, and recording the reflections off each surface (26:787). Using this method, obstacles within a room can be considered. This leads to the development of sophisticated models accounting for reflecting surfaces (desks, windows) and scattering surfaces (screens, computers) within the room (26). The model developed in this thesis does not account for these factors.

2.3.2.2 Image Source Method. The image source method (also called the mirror image method) for determining reflected sources is also based on geometrical acoustics (18) (2) (26) (5). The image source method determines the location of a reflected sound source by considering a point on a reflective surface which represents the sound field generated by the original sound source striking that surface. This point can also be considered a field generated by a secondary (virtual) source determined by mirroring the original source at the plane of the reflecting surface. Figure 3 provides an example of how a virtual source location is determined using the image source method (5). The effects of reflective surfaces in a room can then be represented by a set of image sources symmetrically placed with respect to the reflecting surfaces (24:367). This method is ideal for the case of simulating a rectangular room; however, its use in more complex simulations requires a great deal

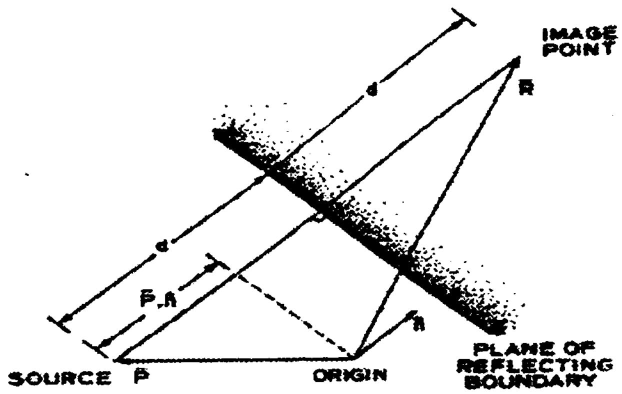

Figure 3. Calculating a Virtual Source Using the Image Source Method (Adapted from Borish)

of computational effort (18:265). Figure 4 shows a two dimensional slice of virtual sources obtained using the image source method in a rectangular room (5). In the case of a complex simulation, the ray tracing method is more efficient; however, given unlimited computational effort, the results of both the ray tracing and image source methods would be identical. For the simple rectangular room simulation accomplished in this research, the image source method is used to determine the location of virtual sound sources.

2.3.3 Reverberation. When a sound wave is generated in a room and decays exponentially over time, this dying out is called reverberation. The length of time it takes the sound to decay to one-millionth of its initial mean value (a reduction of 60 dB) is called the reverberation time (25:558). The reverberation time is dependent upon factors such as the total absorbtion of the room, the volume of the room, and the speed of sound (25:578-579). Reverberation or "the late part" of a room response does not provide significant localization cues for a listener (18). The reflection density in the reverberation component of sound prevents the human auditory system from identifying individual late reflections; however, the brain does use the reverberant part of sound. Reverberation cues contain significant informa-

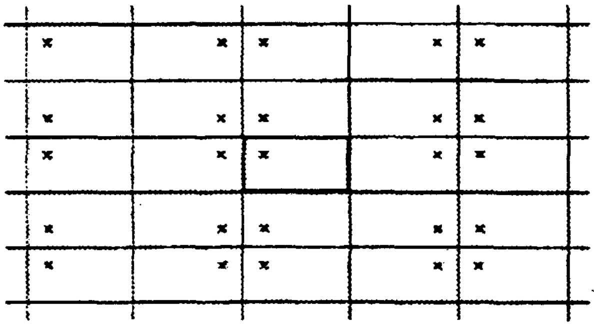

Figure 4. Two-Dimensional View of Virtual Sources Obtained Using the Image Source Method in a Rectangular Room (Adapted from Borish)

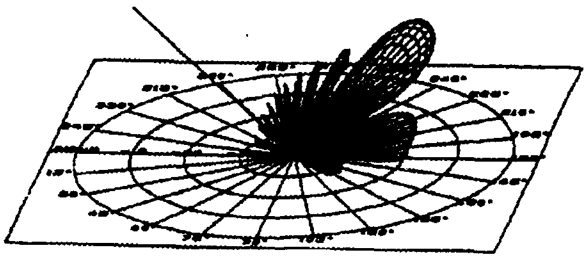

Figure 5. An Example of a Diffuse Reflection Resulting from Sound Contacting a Reflective Surface (Adapted from Blauert)

tion about room size and spatiousness, and these cues help a listener to identify coloration of sounds (35) (18). Methods do exist for sound field modelling of reverberation (18:272). The technique used most often is direct waveform generation controlled by stochastic parameters.

2.3.4 Diffuse Reflections. One of the specific effects that is neglected using geometrical acoustics is the effect of diffuse reflections. When a sound hits a wall, sound energy is reflected in multiple directions–not just the one direction accounted for in the image source method (34) (18) (27). Figure 5 provides an example of the diffusion pattern that occurs when a sound contacts a reflecting surface (18:277). The amount of diffusion and the shape of the diffusion pattern are dependent upon the characteristics of the reflecting surface, the angle of the reflection, and the frequency of the sound. These diffuse reflections may be accounted for in the image source method by generating a cloud of virtual sources for each image source; the cloud of sources is determined from the coefficient of absorption and the coefficient of diffusion of the reflecting surface. Diffuse reflections are not implemented in the binaural room simulation accomplished in this thesis.

2.3.5 Attenuation and Phase Change Due to Reflections. Another factor that is not directly accounted for in the image source method is the absorption of sound that occurs at a reflective surface. The assumption that a reflecting surface is rigid is not accurate when considering many acoustical environments. When a sound strikes a reflecting surface, there is a transfer of acoustic energy into heat energy (35) (24) (30) (25). The amount of absorption that occurs is dependent upon the characteristics of the reflective surface, the angle of reflection and the frequency of the sound. The absorption at a surface can be measured; this measured quantity is called an absorption coefficent. These coefficients represent the average fraction of power absorbed by a surface when sound is falling on it (30) (25). Absorption coefficients can be implemented into a binaural room simulation in order to represent attenuation in the sound signal due to absorption at the reflecting surface.

Another change which occurs in the characteristics of a sound when it is reflected is phase change. Phase change is not considered in the mathematical model or the binaural room simulation completed in this thesis.

2.3.6 Attenuation Due to the Medium. As sound travels through any acoustic environment, it passes through some medium. The speed at which the sound travels through the medium is a function of the medium itself (43:30-35). Furthermore, the amount of acoustic energy that is absorbed by the medium is dependent upon the characteristics of the medium (humidity and temperature in air, for example) and the frequency of the sound (18:282). In this thesis, the approximation for absorption of the sound due to the medium (air) is found by attenuating the sound by the inverse of the distance traveled ($1/d$) (18).

2.3.7 Conclusion Room Acoustics. As a sound propagates in a room, the acoustic environment that it travels through shapes the sound that eventually arrives at the listener. These changes are just as important to a listener's perception of a sound as the changes that take place at the head. A listener receives information about the environment from properties such as reverberation, and absorption. Some portion of the room acoustic cues; however, may be very important in helping a listener localize sounds–in particular distance localization. The objective of this thesis is to determine which of these properties, if any, aid in extracranializing binaural sound signals.

2.4 Sound Source Cues

2.4.1 Directivity of the Sound Source. In this thesis, the sound source will be considered a point source. In nature, however, most sources are directional. When a person speaks, for example, sound is generated at different amplitudes in different directions. A listener standing behind the speaker receives a completely different sound signal than a listener standing in front of the speaker. Sound source directivity is a complex cue to model; directivity can best be modelled through the use of directivity filters (18:279). Amplitude and phase characteristics of a sound source's directivity can be accounted for by these filters. Directivity filters are designed by measuring the emitted sound spectra in the direction of interest and some reference direction, and then dividing the results. The directivity filter is a function of frequency and location (18).

2.4.2 Doppler. The change in apparent frequency of a sound as the source moves past a listener is termed the Doppler effect (25:699). This effect occurs due to the relative motion between the source and the listener. A similar Doppler change in frequency results if the listener is in motion with respect to the source. The Doppler effect is a kinematic phenomenon and for this reason motion by the listener can be represented by an opposing motion of the source (25). The Doppler effect is not included in the binaural room simulation since only stationary source and listener are considered.

2.5 Conclusion

Localization of a sound source by the human auditory system is accomplished through identification of differences in the spectra and times of arrival of sound signals at each ear. Spectral shaping of the sound begins from the moment the sound signal is emitted from the source. Source directivity and relative motion play key roles in the sound that arrives at the listener. As the sound travels to a listener, the acoustic environment continues to shape and delay portions of the sound. Reflected sounds, reverberation, and attenuation of the sound present the listener with a sound signal in the free-field which is drastically different from the emitted sound. Head motion and filtering by the pinnae provide additional spectral and temporal cues. Finally the IID and ITD information is interpreted by the brain. Accurate sound localization in a binaural room simulation requires that the listener receives sufficient psychoacoustic cues. Many of these cues are complex and change as a function of frequency, angle, distance, size, and even the type of material involved.

III. *Mathematical Analysis*

3.1 *Introduction*

This chapter contains a mathematical analysis of the localization cues available to a listener in a rectangular room. A transfer function for a sound traveling from a source to each ear of a listener in a rectangular room is developed. The sound source is modeled as an isotropic point source, and the receivers–the left and right ears–are treated as point receivers. The primary emphasis of the analysis is modelling room acoustics; however, free-field cues (ITD, IID, and pinna effects) are also included in the model.

3.2 *Modelling the Delay Due to Distance Traveled*

The analysis begins by chosing a point in three-dimensional space to be the origin. The origin will be defined as the point where wall 2, wall 3, and the floor intersect (See Figure 6.) A monophonic sound signal generated from a point source located at $X_A = (x_A, y_A, z_A)$ within the room will be denoted by $s_A(t)$. The subscript A denotes that this source is the actual sound source. Suppose this signal then travels to a point receiver located at $X_P = (x_P, y_P, z_P)$ within the room. The signal measured at X_P is different from the signal that was generated at X_A, and will be denoted by $s(t, X_A, X_P)$. Notice that $s(t, X_A, X_A)$ is just the original signal $s_A(t)$. Taking the Fourier transform of the original signal results in the following:

$$\mathcal{F}s_A(t) = S_A(\omega)$$

where $\omega = 2\pi f$ denotes the frequency in *radians/second* and $S_A(\omega)$ is the original signal produced at X_A represented in the frequency domain. This signal does not arrive instantaneously at the receiver. The time delay (τ) of the signal measured at X_P is a function of the distance traveled by the sound (D_{AP}) divided by the speed

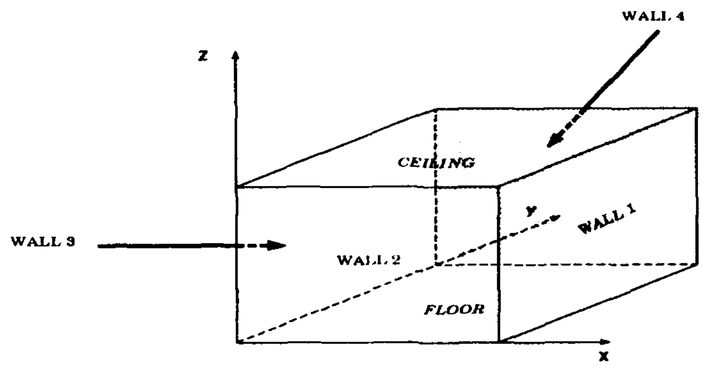

Figure 6. Orientation of the Rectangular Room Used in the Simulation

of sound (c). Thus,

$$s(t, X_A, X_P) = s_A(t - \tau)$$

where $\tau = D_{AP}/c$ and $D_{AP} = dist(X_P, X_A) = \|X_P - X_A\|$ is the Euclidean distance between X_P and X_A. In the frequency domain, the appropriate time delay is modelled as a modulation, that is:

$$\mathcal{F}s(t, X_A, X_P) = S(\omega, X_A, X_P) = S_A(\omega)\exp(\frac{i\omega D_{AP}}{c}) \tag{1}$$

In the case of a sound traveling from a source to a listener, there are two receivers–the left and the right ear. Each ear will be modeled as a point receiver. The location of the left ear is given by the position vector X_L and the position of the right ear is given by X_R. The location of the ears depends upon the location of the center of the head and the orientation of the listener's head with respect to the sound source. In order to determine the location of a listener's ears in free space, a new coordinate system must be defined which represents the listener's viewing angle. The new coordinate system's origin lies on the midpoint of the line segment connecting the the left and right ears. This point is defined as the listener's location and is represented by $X_C = (x_C, y_C, z_C)$. The subscript C denotes the center of the listener's head. The new coordinate system, shown in Figure 7, is related to the original coordinate system by:

$$\begin{aligned}
\hat{e}_1 &= a_1\hat{\imath} + b_1\hat{\jmath} + c_1\hat{k} \\
\hat{e}_2 &= a_2\hat{\imath} + b_2\hat{\jmath} + c_2\hat{k} \\
\hat{e}_3 &= a_3\hat{\imath} + b_3\hat{\jmath} + c_3\hat{k}
\end{aligned}$$

such that $\{\hat{e}_1, \hat{e}_2, \hat{e}_3\}$ is an orthonormal set. That is, $\hat{e}_\alpha \cdot \hat{e}_\beta = 0$ for $\alpha \neq \beta$, and $\alpha, \beta \in \{1, 2, 3\}$, and $\|\hat{e}_\alpha\| = 1$ for $\alpha \in \{1, 2, 3\}$. Therefore, there exists a relationship between the scalars $\{a_\alpha, b_\alpha, c_\alpha | \alpha = 1, 2, 3\}$. Furthermore $\hat{\imath}, \hat{\jmath}$, and $\hat{k}$ are unit vectors

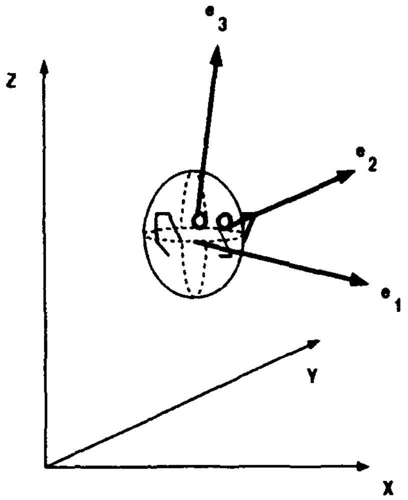

Figure 7. Coordinate System $(\hat{e}_1, \hat{e}_2, \hat{e}_3)$ Defining the Position and Orientation of the Listener's Head

in the x, y, and z directions, respectively, and the $\hat{e}_1$ axis represents the direction the nose is pointing. A single matrix then defines the orientation of the listener's head, with respect to the room. For example, the following matrix represents the case where the listener is looking directly in the $\hat{\imath}$ direction:

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Since the listener's ears lie on the $\hat{e}_2$ axis, the left and right ear positions can be identified by the following vectors:

$$\begin{aligned} X_L &= X_C + (d/2)\hat{e}_2 \\ X_R &= X_C - (d/2)\hat{e}_2 \end{aligned}$$

where d is the length of the line segment connecting the two ears.

Given the position of both ears, the distance from any sound source to either or both ears can be determined. Given the actual source location X_A, the distance to the right ear D_{AR} is given by:

$$\begin{aligned} D_{AR} &= dist(X_R, X_A) \\ &= dist(C - [d/2]\hat{e}_2, X_A) \\ &= \sqrt{(x_C - [d/2]a_2 - x_A)^2 + (y_C - [d/2]b_2 - y_A)^2 + (z_C - [d/2]c_2 - z_A)^2} \end{aligned}$$

For the example above where the listener is looking directly in the $\hat{\imath}$ direction, $\hat{e}_1 = \hat{\imath}, \hat{e}_2 = \hat{\jmath}$, and $\hat{e}_3 = \hat{k}$, hence $a_2 = 0, b_2 = 1$, and $c_2 = 0$, so that:

$$D_{AR} = dist(X_R, X_A) = \sqrt{(x_C - x_A)^2 + (y_C - [d/2] - y_A)^2 + (z_C - z_A)^2}$$

The distance to the left ear is determined using X_L.

Given this distance, the signal at the right ear, accounting for time delay due to distance traveled, is given by substituting X_R for X_P and D_{AR} for D_{AP}into Equation 1, that is,

$$\mathcal{F}s(t, X_A, X_R) = S(\omega, X_A, X_R) = S_A(\omega)\exp(\frac{i\omega D_{AR}}{c}) \tag{2}$$

3.3 Determining the Angle of Incidence

The incidence angle on the receiver (the listener's right ear) is also a function of the listener's viewing angle and the position of the center of the listener's head in free space. Given the coordinate system defined in the previous section, the position and orientation of the head can be defined. Now the angle at which the sound arrives at the listener must be determined.

In dealing with the angle of incidence, the azimuth angle (θ) will range from $-\pi$ to π ($-\pi < \theta \leq \pi$) and is measured with respect to the ($\hat{e}_1, \hat{e}_2, \hat{e}_3$) coordinate system. The nose is $\theta = 0$. Angles on the listener's right (the $-\hat{e}_2$ direction) are positive; this orientation is shown in Figure 8. Elevation angle (ϕ) ranges from $-\pi/2$ to $\pi/2$ ($-\pi/2 < \phi < \pi/2$), and is also measured with respect to the ($\hat{e}_1, \hat{e}_2, \hat{e}_3$) coordinate system, so that $\phi = \pi/2$ corresponds to the $\hat{e}_3$ direction, and $\phi = -\pi/2$ corresponds to the $-\hat{e}_3$ direction. This is identical to the orientation shown in figure 2. A vector from the right ear to the actual sound source is given by $X_A - X_R$, and then corrected to the center of the head by adding $(d/2)\hat{e}_2$ to get:

$$X_D = X_A - X_R + (d/2)\hat{e}_2 = X_A - X_C$$

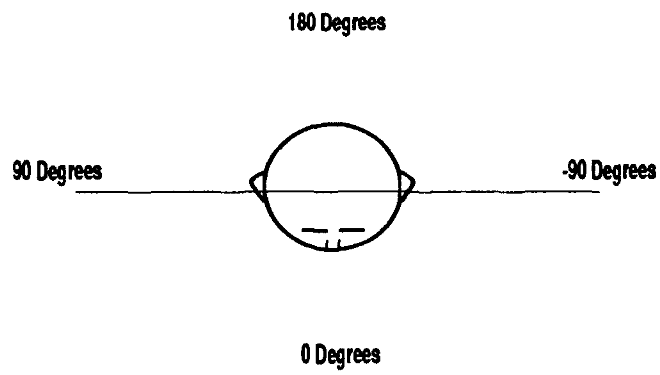

Figure 8. Orientation for Azimuth in the Mathematical Model

The subscript D denotes difference. Now the projection of this vector onto the $\hat{e}_1\hat{e}_2$ plane is defined:

$$X_{D,\hat{e}_1,\hat{e}_2} = \mathcal{P}_{\hat{e}_1,\hat{e}_2}(X_D) = (X_D \cdot \hat{e}_1)\hat{e}_1 + (X_D \cdot \hat{e}_2)\hat{e}_2$$

This vector may point in positive or negative θ direction depending on the location of the virtual sound source; this must be considered when determining θ. Therefore, the following function is defined:

$$f(X_{D,\hat{e}_1,\hat{e}_2}) = \begin{cases} 1 & \text{if } X_{D,\hat{e}_1,\hat{e}_2} \cdot \hat{e}_2 < 0 \\ 0 & \text{if } X_{D,\hat{e}_1,\hat{e}_2} \cdot \hat{e}_2 = 0 \\ -1 & \text{if } X_{D,\hat{e}_1,\hat{e}_2} \cdot \hat{e}_2 > 0 \end{cases}$$

Now θ is calculated:

$$\theta = \Theta(X_A, X_C, \mathcal{E}) = f(X_{D,\hat{e}_1,\hat{e}_2})\text{Arccos}\left(\frac{X_{D,\hat{e}_1,\hat{e}_2} \cdot \hat{e}_1}{\|X_{D,\hat{e}_1,\hat{e}_2}\|}\right)$$

where $\mathcal{E} = \{\hat{e}_1, \hat{e}_2, \hat{e}_3\}$, and Arccos is the principle branch of the arc-cosine function.

Similarly, the elevation angle (ϕ) is calculated:

$$\phi = \Phi(X_A, X_C, \mathcal{E}) = \frac{\pi}{2} - \text{Arccos}\left(\frac{X_D \cdot \hat{e}_3}{\|X_D\|}\right)$$

3.4 Head Related Transfer Function

The head related transfer function (HRTF) denoted by H was discussed in detail in Chapter 2. Recall that the HRTF is a function of frequency (ω), source position $X_A = (x_A, y_A, z_A)$, and the point receiver position $X_P = (x_P, y_P, z_P)$. It is also a function of viewing angle defined by $\mathcal{E}$. For any given source location, there is a separate HRTF for each ear. For the purposes of this analysis, only the right ear is addressed (analysis for the left ear is identical). Given that the point receiver is the listener's right ear and θ and ϕ are determined as shown in the previous section, a head related transfer function for the right ear at these angles can be used to filter the incoming sound from the actual sound source. In fact, the HRTF is measured with respect to these angles θ and φ. In particular, if $\check{H}_R(\omega, \theta, \phi)$ denotes the HRTF of the right ear for a given frequency ω, azimuth θ, elevation ϕ, and head viewing direction $\mathcal{E}$ then:

$$H_R(\omega, X_A, X_C, \mathcal{E}) = \check{H}_R(\omega, \Theta(X_A, X_C, \mathcal{E}), \Phi(X_A, X_C, \mathcal{E}))$$

defines the HRTF in terms of location vectors. This transfer function accounts for the filtering effects of the pinnae. Multiplying the HRTF for the right ear and the delayed sound signal in Equation 2:

$$H_R(\omega, X_A, X_C, \mathcal{E}) S_A(\omega) \exp(\frac{i\omega D_{AR}}{c}) \tag{3}$$

produces a directional monophonic signal for the direct path from the actual source to the right ear. When presented simultaneously with the signal for the left ear, the result is a binaural signal filtered to account for the HRTF data and time delays (including the ITD if its frequency dependence is ignored). This is the direct path sound to the listener.

3.5 Headphone Transfer Function

In general, the headphone transfer function is a function of frequency. For the Etymotic insert headphones used in this thesis, the transfer function is assumed to be constant per the manufacturer's data. As is the case throughout this analysis, the phase portion of the transfer function is ignored. In other words, the output of a given input sound is attenuated equally in amplitude for all frequencies. If the transfer function is not constant, it can still be measured. In this thesis, the headphone transfer function is assumed to be identical for each earpiece and is denoted $P(\omega)$. In many cases, however, the headphone transfer function is different for each ear

($P_L(\omega)$ and $P_R(\omega)$ would be used if this were the case). The inverse of the headphone transfer function can be multiplied by Equation 3:

$$S(\omega, X_A, X_R) = \frac{1}{P(\omega)} H_R(\omega, X_A, X_C, \mathcal{E}) S_A(\omega) \exp(\frac{i\omega D_{AR}}{c}) \qquad (4)$$

producing a directional monophonic signal for the right ear which accounts for the filtering effects of the head and pinna, and cancels the effects of the headphones.

3.6 *Room Transfer Function*

The room transfer function is composed of various effects. Reflections, and attenuation are the key portions of this transfer function. The room being modelled in this research is composed of four walls, a floor, and a ceiling. Again, the orientation of the room is shown in figure 6.

3.6.1 Modelling Reflections. If a single reflecting surface (wall) is assumed to be present, a virtual source placed symmetrically outside the wall represents the imaginary source of the reflected sound off that wall. This is a direct result of the image source model discussed in Chapter 2. The sound signal at the listener's right ear is now composed of two parts:

$$S_A(\omega) \exp(\frac{i\omega D_{AR}}{c}) + S_A(\omega) \exp(\frac{i\omega D_{VR}}{c})$$

where D_{VR} is the distance from the right ear to the virtual source, that is $D_{VR} = dist(X_V, X_R)$.

In the case of a room with four walls, a floor, and a ceiling, an infinite number of reflections are possible. Each of these reflections is represented by a virtual source located in one of an infinite number of virtual rooms located in three dimensional space. An arbitrary sound source location is represented by the position vector $X_{(l,m,n)}$ and is located in the (l, m, n) virtual room $(l, m, n \in \mathbf{Z})$. The original sound source location is denoted by $X_{(0,0,0)}$, and $X_{(1,0,0)}$ is the virtual source located in the virtual room adjacent to the actual room in the positive x direction. Figure 9 depicts how the virtual rooms are arranged around the actual room. Sound generated from a source located at $X_{(0,0,0)}$ produces an infinite number of reflections in a rectangular room. The sound field generated by the actual sound source can be modelled as an infinite number of virtual sources in free space. Let $\mathcal{X}$ denote the collection of the original sound source and all virtual sound source locations. That is, $\mathcal{X} = \{X_{(l,m,n)} | l, m, n \in \mathbf{Z}\}$. Given this notation, the signal at the listener's right ear, accounting for all reflections and their corresponding time delays, can now be represented:

$$S(\omega, \mathcal{X}, X_R) = S_A(\omega) \sum_{l,m,n=-\infty}^{\infty} \exp(\frac{i\omega D_{(l,m,n)R}}{c})$$

where $D_{(l,m,n)R}$ is the distance from the right ear to the $X_{(l,m,n)}$ source, that is $D_{(l,m,n)R} = dist(X_{(l,m,n)}, X_R)$. Given that the azimuth and elevation of each of the virtual sound sources can be determined from the equations in section 3.3, the model can now be extended to account for the filtering effects of the pinna on each reflection.

$$S(\omega, \mathcal{X}, X_R) = S_A(\omega) \sum_{l,m,n=-\infty}^{\infty} \left[\exp(\frac{i\omega D_{(l,m,n)R}}{c})\right] H_R(\omega, X_{(l,m,n)}, X_C, \mathcal{E}) \qquad (5)$$

where $H_R(\omega, X_{(l,m,n)}, X_C, \mathcal{E})$ is the HRTF corresponding to the angle of the $X_{(l,m,n)}$ sound source for the right ear, and X_C is considered a function of X_R. The same calculations could be performed on the left ear. Presenting the resulting two signals simultaneously to the appropriate ears results in a binaural sound accounting for all reflections appropriately delayed. Attenuation due to distance traveled and due to absorption at the walls has yet to be considered.

3.6.2 Modelling Attenuation. Recall from Chapter 2, that as a sound propagates in a room, the sound is attenuated due to absorption by 1) the medium through which it travels (air) and 2) the surfaces from which it is reflected (the walls,

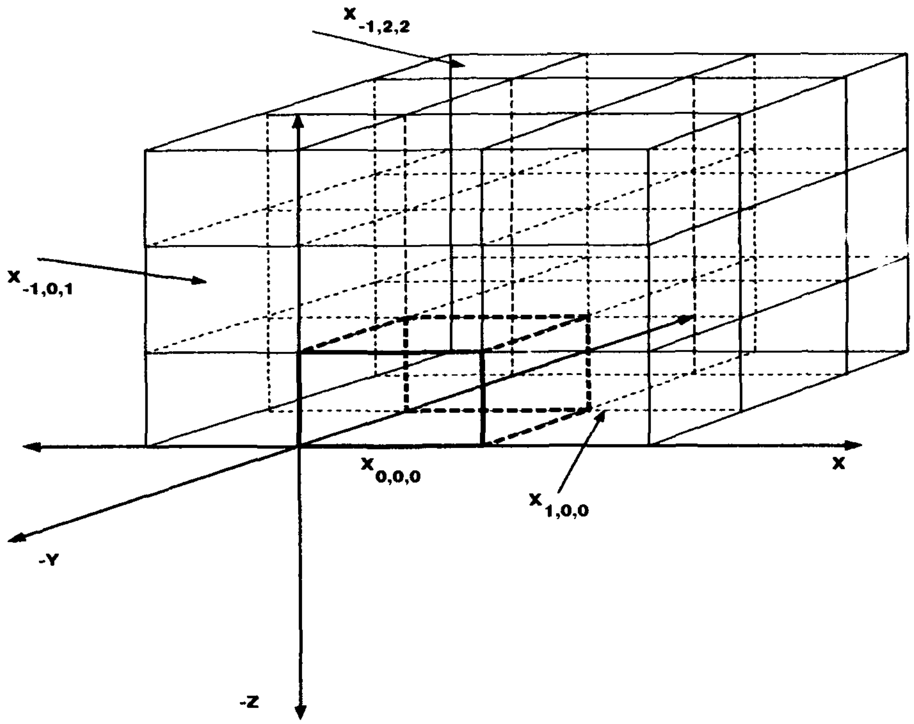

Figure 9. Orientation of the Virtual Rooms Around the Room Containing the Sound Source

ceiling, and floor). The attenuation due to absorption by the medium of a sound traveling from X_A to X_R is a function of the distance the sound travels (D_{AR}), and the properties of the medium (ε_0). The properties of the medium include items such as composition (oxygen, nitrogen, etc.), humidity, and temperature. Including this attenuation as the original signal travels directly to X_R results in the following:

$$S_A(\omega, X_A, X_R) = S_A(\omega)\left(\frac{\varepsilon_0 \exp[i(\omega/c)D_{AR}]}{4\pi D_{AR}}\right)$$

which represents the sound at the right ear prior to filtering by the pinna. In the simulations completed in this thesis, $\varepsilon_0/4\pi$ is assumed to equal 1. Again, this model could be extended to account for all reflections, and the filtering by the appropriate HRTF's; however, first the attenuation by the walls will be included.

Absorption coefficients for hundreds of different materials can be found in the literature (14). The sound absorption coefficient (α) is directly related to the coefficient of reflection (β) by the following equation (2):

$$\alpha = 1 - \beta^2$$

Like α, β is a function of frequency and angle. In this thesis, β is considered to be the same for all reflective surfaces. Given this relation, attenuation caused by absorption by the wall can be accounted for by filtering appropriate fequencies by $\beta^h_{(l,m,n)}(\omega)$, where $\beta^h_{(l,m,n)}(\omega)$ filters the $X_{(l,m,n)}$ source h times ($h = |l| + |m| + |n|$).

3.7 Overall Transfer Function

The final step of this analysis is to put the pieces of the previous sections together. The overall signal to the right ear is described by:

$$S(\omega, \mathcal{X}, X_R) = \frac{\varepsilon_0}{4\pi P(\omega)} S_A(\omega) \qquad (6)$$

$$\times \sum_{l,m,n=-\infty}^{\infty} H_R(\omega, X_{(l,m,n)}, X_R, \mathcal{E})\beta^h_{(l,m,n)}(\omega)\left(\frac{\exp[i(\omega/c)D_{(l,m,n)R}]}{D_{(l,m,n)R}}\right)$$

3.8 Conclusion

Given the overall transfer function for an isotropic sound signal traveling to a receiver through a rectangular room, a separate transfer function can be calculated to each ear as follows:

$$\begin{aligned} \mathcal{H}_L(\omega, \mathcal{X}, X_L, \mathcal{E}) &= \frac{\varepsilon_0}{4\pi P(\omega)} \\ &\times \sum_{l,m,n=-\infty}^{\infty} H_L(\omega, X_{(l,m,n)}, X_L, \mathcal{E})\beta^h_{(l,m,n)}(\omega)\left(\frac{\exp[i(\omega/c)D_{(l,m,n)L}]}{D_{(l,m,n)L}}\right) \end{aligned} \tag{7}$$

and,

$$\begin{aligned} \mathcal{H}_R(\omega, \mathcal{X}, X_R, \mathcal{E}) &= \frac{\varepsilon_0}{4\pi P(\omega)} \\ &\times \sum_{l,m,n=-\infty}^{\infty} H_R(\omega, X_{(l,m,n)}, X_R, \mathcal{E})\beta^h_{(l,m,n)}(\omega)\left(\frac{\exp[i(\omega/c)D_{(l,m,n)R}]}{D_{(l,m,n)R}}\right) \end{aligned} \tag{8}$$

where X_L denotes the location of the left ear. A binaural sound can now be presented to a listener by simultaneously playing each sound to the appropriate ear. This model is the foundation for the binaural room simulation completed in this thesis.

IV. Methodology

4.1 Introduction

The mathematical model developed and presented in Chapter 3 represents the completion of the first research objective of this thesis. The second research objective, as stated in Chapter 1, was to implement portions of the mathematical model electronically in order to generate a binaural signal that can be extracranialized. The process of adding the effects of different portions of the model into a simple binaural room simulation can be broken down into essentially eight separate parts:

1. Setting the dimensions of the room, placing the sound source and the listener at specific locations in the room, and determining the angle of incidence and distance to the listener from the sound source and chosen virtual sources.
2. Determining the angles to be used in the design of the Head Related Transfer Function (HRTF) filters from the angle of incidence data.
3. Designing the FIR filters (from the AAMRL HRTF data) for both the left and right ears for all appropriate angles.
4. Recording a monophonic signal and filtering the recorded sound with each of the HRTF filters to yield a collection of monophonic sound files.
5. Processing the monophonic sound files to account for ITD and time delay to the listener due to the distance the sound travels.
6. Processing of the signals already including time delay information to account for attenuation of the sound due to distance traveled from source to listener.
7. Combining of appropriate monophonic sound files to yield a collection of binaural sound files–each containing different degrees of free-field and room-related effects.

8. Varying the ITD cue to match the size of the listener's head diameter.

This chapter will first review each of the steps that were taken in this binaural room simulation in order to generate the final binaural signals. Finally, the method of determining whether or not a particular sound could be accurately localized and/or extracranialized will be discussed.

4.2 Determining the Angle of Incidence and Distance to the Listener

4.2.1 Room Dimensions. Using the mathematical model developed in Chapter 3, a MathCad template was developed. Using this template, specific room dimensions, sound source locations, and listener location and orientations were selected and inputed into the file. The template allowed for any rectangular shaped room. For the simulation performed in this thesis, the room dimensions were selected to be 15 meters in width (x-axis), 20 meters in length (y-axis), and 5 meters in height (z-axis).

4.2.2 Listener Location and Orientation. The listener could be placed at any location in the room and his head could be oriented in any direction. For each of the seven simulations performed in this thesis the listener was located in the same positon and orientation–($.55L_x$, $.75L_y$, $.15L_z$) always looking in the $\hat{\imath}$ direction at Wall 1 with ears parallel to the xy plane. Given this information, location of the left and right ear were pinpointed. This detailed ear location was not critical to the simulation run in this thesis; however. knowledge of the exact location of each ear may possibly be used in follow-on research which would include the use of a head-tracker and the effects of head motion.

4.2.3 Sound Source Locations. The actual sound source could be placed anywhere inside the room. The sound sources for simulating reflections are placed outside the room–virtual sound sources determined by the image source method. Given a single sound source location. virtual source locations representing the reflections off each of the four walls, the floor, and the ceiling were generated by the template. Seven different original source locations were run in this thesis research:

1. $(.58L_x,.85L_y,.20L_z)$
2. $(.75L_x,.80L_y,.95L_z)$
3. $(.28L_x,.75L_y,.15L_z)$
4. $(.48L_x,.65L_y,.20L_z)$
5. $(.48L_x,.85L_y,.01L_z)$
6. $(.75L_x,.78L_y,.15L_z)$
7. $(.58L_x,.65L_y,.20L_z)$

Six virtual source locations were determined in each of the seven cases above; these virtual sources corresponded to the first reflection off each of the four walls, the floor, and the ceiling. The seventh case in this list included ten reflections. In this case, four additonal virtual sources were calculated; these sources represented the following four double reflections:

- Reflection off the Floor then Wall 1
- Reflection off the Floor then Wall 3
- Reflection off the Floor then the Ceiling
- Reflection off the Ceiling then the Floor

Given the actual sound source location and the virtual sound source locations, the user need not make any more inputs; the computations were done by the MathCad program.

4.2.4 MathCad Template Output. Once room dimensions, listener location and orientation, and sound source location had been input, the program output could be viewed. The template was designed to give the angle of incidence to both the left and the right ear for all sound sources–actual and virtual. This angle was given in two parts; first the azimuth angle was reported. Second, the elevation angle was reported. Finally, program output provided the acoustic path length to left and right ear from each of the sound sources. A sample of the MathCad output can be found in Appendix A.

4.3 Determining the Angles to Use for Filter Design

As was stated in Chapter 2, the HRTF data to be used in this thesis was collected at the Armstrong Aerospace Medical Research Laboratory (AAMRL) located at Wright-Patterson Air Force Base (WPAFB). The data collection was accomplished in an anechoic chamber containing a dome of 272 speakers. A list of the speaker locations by angle of azimuth and elevation can be found in Appendix B. Because of the limited number of speaker locations, a perfect match to the incidence angles obtained from the MathCad template was not likely. For this reason, approximations were necessary. The group of tables located in Appendix C shows the approximate angles of azimuth and elevation given in the MathCad output, and the angles corresponding to the speaker used to design the HRTF filters for each of the seven simulations. Speaker locations were chosen by determining which speaker provided the closest approximation to the azimuth and elevation angles reported in the MathCad output. The speaker locations and corresponding ITDs are shown in Appendix B.

Given that these were the speaker locations chosen, the next step in the simulation was to take the raw data for each of the selected speakers and develop FIR filters which would represent the HRTF's for the corresponding angles of incidence.

4.4 Designing the Filters

The filter design for this room simulation was accomplished using a Sun Sparc 2 workstation and the ESPS software. The "wmse_filt" command was used in the design of all left and right ear filters. The bandedges and responses were determined from the AAMRL raw data and a weighting function of 1 was employed for each band. The manual pages for this ESPS command as well as for all ESPS commands used in this thesis effort can be found in Appendix C. The filters designed using this command were 93 tap FIR filters. Examples for both the left and right ear can be found in Appendix D. The filters were generated using the data taken from measurements on a KEMAR head and bust. The sampling frequency was 40 kHz, and the weighting function was constant at 1. A left and right ear filter were designed for each of the chosen speaker locations.

4.5 Recording and Filtering a Monophonic Signal

This step of the research was also accomplished using the binaural mixing console mentioned in the previous section. In this case, the specific commands used were "s32crecord" and "filter". The manual pages for these commands are found in Appendix C. The first command. "s32crecord", allows for the recording of a monophonic signal using an Ariel A/D converter. For the simulation carried out in this thesis, the recorded sound was male speech, and the sampling frequency of the recording was 40 kHz.

Once a sound data file was created. the monophonic sound file was filtered with each of the previously designed filters using the "filter" command. For six of the seven simulations, the result of this step was 14 sound files-seven left and seven right. The seven files are the result of the direct path, the four wall reflections, and the reflections off the floor and ceiling. In the seventh case. four additional virtual sources representing double reflections were included, so the result was 22 files. Each of these files contained a monophonic speech signal which had been filtered to account for the HRTF, but the signals still needed to be appropriately delayed.

4.6 Accounting for ITD and Time Delay Due to Distance Traveled

4.6.1 ITD. As was stated in Chapter 2, ITD is a critical cue in sound localization–particularly at frequencies below 3000 Hz. Clearly, this cue had to be included in the filtered sound files in order to generate localizable binaural sound files. ITD is a function of angle of incidence and frequency. The frequency dependency of this cue, although genuine, would be very complex to model and research indicates that it is not a critical cue in sound localization (42) (19). In this simulation, the frequency dependence of the ITD cue is ignored, and ITD is assumed to be constant for a given angle. The ITD's used in this thesis were measured at AAMRL using a KEMAR manikin with a head radius of 8.5 cm. The measurements can be found in Appendix A. The ITD for each source was found from the data in Appendix B and then added to the sound corresponding to the ear furthest from the sound source. The delay was added into the sound files using the ESPS command "delay"(See Appendix C).

4.6.2 Distance Time Delay. In addition to the delay added for ITD, both left and right ear signals had to be delayed for the distance traveled by the sound. The distance traveled was calculated by averaging the left and right ear distances given in the MathCad output discussed previously. The calculated distance was divided by the speed of sound ($344m/s^2$ for room temperature with dry air) to obtain the time delay resulting from the distance traveled. This delay was added into all of the sound files using the "delay" command.

4.6.3 Adding the Delayed Sounds to Simulate Reflections. The delayed files now needed to be combined to begin simulating reflections. In this step left and right ear sound files were dealt with separately. Using the ESPS command "addsd" the direct path for the left ear was added to the Wall 1 reflection for the left ear. The result was a sound file that contained information for the direct path and one reflection. Next, this direct plus one reflection file was added to the Wall 2 reflection for the left ear yielding a file containing direct path and two reflections for the left ear. This was done again for the Wall 3 reflection. the Wall 4 reflection, the Floor reflection, and the Ceiling reflection. This same procedure was followed for the right ear as well. The result was again 14 files of interest--seven left and seven right. Each file contained appropriate time delays, accounted for HRTF data, and contained varying degrees of reflection information.

4.7 Accounting for Attenuation Due to Distance Traveled

In Chapter 3, the mathematical model accounting for sound travel through a room included attenuation in the sound resulting both from the distance it traveled through the air and for absorption at walls from which the sound is reflected. If we assume the walls, ceiling, and floor are all made from the same type of material, the attenuation due to absorption at the walls may be factored out of all but the direct path sound. The effect of this attenuation is small when compared to the attenuation in the sound from all sources, including the direct path, resulting from the distance traveled through the air. As a sound moves from a source to a listener it is attenuated proportional to the reciprocal of the distance traveled.

This attenuation can be implemented by using a floating point scaling factor found in the delay command (scale option). Using the scale option the 14 sound files mentioned in section 4.6.2 (before the sounds are added to simulate reflections) were appropriately scaled to yield 14 sound files containing HRTF, time delay, and attenuation information. These files were added to simulate reflections in the same procedure described in section 4.6.3.

4.8 Creating a Binaural Signal

Using the ESPS command "mux" (which is simply the command name) two sound files can be combined in a single sound file containing a left and right ear signal. When played back through an Ariel D/A converter and stereo amplifier using the command "s32cplay" these combined files are binaural - a separate signal is heard at each ear. Using this procedure, the monophonic sound files obtained in section 4.6.3 and section 4.7 were combined to yield 14 binaural sound files- seven containing time delay information and no attenuation, and seven containing both time delay and attenuation information. The files all contained various numbers of reflections, and HRTF data. These 14 sound files (22 in one case) were now available for each of the seven simulations. These sound files were used to make judgements on which factors did and did not affect the extracranialization of the binaural signals.

4.9 Increasing the Head Size

In an effort to increase the extracranialization of the sounds, a final step was added to each simulation. The ITD cue was increased by 25, 50, and 100 percent to simulate an enlarging of the listener's head. The resulting signals contained identical information to those described in section 4.8 with the exception of the increased delay to the ear furthest from the sound source. The increased ITD cue was included in the direct path and all reflections. The results of this adaptation are also found in Chapter 5.

4.10 Conclusion

This chapter has laid out the process for accomplishing the binaural room simulation for generating a signal that can be extracranialized. Using the Sun Sparc 2 workstation and the ESPS software, signals were appropriately filtered, delayed, attenuated, added, and combined to generate a collection of binaural sound files.

V. *Informal Subjective Impressions*

5.1 Introduction

This chapter details the results obtained from the binaural room simulation discussed in Chapter 4. These results were obtained from three sources, 1) an initial investigation through the playing of a comparitive group of sounds, 2) further investigation through the playing of a wide variety of sounds in a doudle blind experiment, and 3) informal tests to determine the affects of changing head size. The factors manipulated in each of these investigations were angle of incidence, number of reflections, attenuation of the reflections, and ITD (in the increasing of the virtual head size).

5.2 Results of the Initial Investigation

When the initial investigation began, the collection of sounds consisted of 42 binaural sound files. The breakdown of the sound files was as follows:

- From 285 degrees of azimuth and 10 degrees of elevation, there were 7 sounds containing HRTF and time delay data. These sounds ranged from direct path to direct path plus all single reflections.
- Also from 285 degrees of azimuth and 10 degrees of elevation, there were 7 sounds containing HRTF, time delay, and attenuation data. These sounds ranged from direct path to direct path plus all single reflections.
- From 240 degrees of azimuth and 10 degrees of elevation, there were 7 sounds containing HRTF and time delay data. These sounds ranged from direct path to direct path plus all single reflections.
- Also from 240 degrees of azimuth and 10 degrees of elevation, there were 7 sounds containing HRTF, time delay, and attenuation data. These sounds ranged from direct path to direct path plus all single reflections.

- From 180 degrees of azimuth and 0 degrees of elevation, there were 7 sounds containing HRTF and time delay data. These sounds ranged from direct path to direct path plus all single reflections.
- Also from 180 degrees of azimuth and 0 degrees of elevation, there were 7 sounds containing HRTF, time delay, and attenuation data. These sounds ranged from direct path to direct path plus all single reflections.

Using these stimuli, listeners were asked to make some forced choice paired comparisons. Thes comparisons were accomplished in order to gain some qualitative information that would be helpful in a more detailed experiment. The information was collected through verbal responses by the subjects. During this initial investigation, responses from the listeners indicated that there was not a significant difference between listening to the sounds through the insert headphones and the Walkman headphones. Furthermore, listener's preferred the comfort of the Walkman type headphones. Therefore, the Walkman headphones were used for all further investigations (the transfer function for these headphones was not flat). The sound files for 285 degrees of azimuth were listened to first. Listeners were made aware that the sounds were all coming from the same direction, and they were asked to verbally state differences noted when listening to a series of sounds.

5.2.1 Comparisons for Addition of Reflections without Attenuation. The first collection of sounds that was played was :

- Direct Path
- Direct Pat' , Reflection off Wall 1
- Direct Path, Reflections off Wall 1 and Wall 2
- Direct Path, Reflections off Wall 1, Wall 2, and Wall 3
- Direct Path, Reflections off Wall 1, Wall 2, Wall 3, and Wall 4
- Direct Path, Reflections off Walls 1, 2, 3, 4, and the Floor

- Direct Path, Reflections off Walls 1, 2, 3, 4, and the Floor and Ceiling

All of the listeners were able to localize the direct path sound signal which included HRTF and ITD data. Listeners tended to lateralize the sound in the proper quadrant; however, most listeners did not extracranialize the sound. The addition of a single reflection into the sound brought a mixture of responses from the listeners. Lateralization in azimuth and elevation was similar to the direct path case, but listeners definitely noted a change. Some subjects reported that the sound was just outside the head, while others reported in-head localization. With the addition of a second reflection, all listeners could tell the sound was coming from their right, but some could not lateralize the source as well as the direct path alone. The addition of a third reflection resulted in similar comments from the listeners. In both cases, more listeners reported localizing the sound slightly outside of the head. The addition of four or more reflections into the sound file, however, resulted in declining accuracy in lateralization of the sound. Subjects began to localize on multiple sound sources, and in most cases the subjects could not localize the source at all. Furthermore, including greater than three reflections caused the listeners to hear echoes or, as some stated, to hear "in stereo". Since the reflections presented to the listeners in this case were not attenuated, they did not sound natural and the brain attempted to localize the source of the reflections as well as the original source.

5.2.2 Comparisons with Attenuated Reflections. After listening to the first collection of sounds, listeners were asked to listen to seven different sounds. These sounds were played in the same order described in the previous section; however, this group of files accounted for the attenuation in the sound due to the distance traveled through air.

Comments from the listeners for the direct path case were similar to those made for the unattenuated direct path. Lateralization in azimuth and elevation was again in the proper quadrant, and listeners reported in-head, on-head, and slightly out-of-head localization. The addition of one reflection caused very little change in the listeners' responses; however, some did report that the sound source seemed farther away. When a third reflection was added, more listeners experienced extracranialization of the sound, and localization in azimuth and elevation was still good. The addition of four, five, and six reflections resulted in similar comments. Nearly all listeners reported out of head localization when four or more reflections were included in the sound signal. Furthermore, none of the listeners reported hearing echoes, and localization of the sounds was still in the proper quadrant.

5.2.3 Comparisons Between Attenuated and Unattenuated Reflections. In the final step of the preliminary investigation, the listeners were asked to compare the quality of different pairs of sounds. The orders of all the comparisons were alternated and listeners received the same pair of sounds multiple times. One of the first comparisons made was direct path (non-attenuated) to direct path (attenuated). Interestingly, many listeners reported that the attenuated direct path signal sounded farther away than the non-attenuated direct path signal. Listeners also reported the perception of increased distance as the number of reflections was increased. Numerous combinations were compared with the most interesting result being that when three or more attenuated reflections were included in the sound signal, all listeners perceived this signal being farther away than the non-attenuated direct path signal.

5.3 Double-Blind Experiment

5.3.1 Purpose and Scope. The emphasis of this thesis was not placed on performing a detailed human factors analysis of the final product of this research; however, a simple experiment was accomplished. The purpose of this experiment was to gain more knowledge about what factors affect extracranialization of sounds and to avoid bias of the experimenter. The experiment and interpretation of its results were used primarily to draw some broad conclusions about the effects of certain aspects of the room simulation on a listener. After the initial investigation, groups of 14 sounds were developed for four additional locations (See Chapter 4) and four double reflections were included in the sound files for 285 degrees of azimuth. This brought the number of binaural sound files to 106. Twenty-five sounds were selected from the 106; the selection of the sounds was purposeful. For example, since the purpose of this experiment was to determine what cues cause the sound to be extracranialized by the listener, many of the sounds containing a large number of non-attenuated reflections were not included. From the initial investigation, these sounds were determined to be difficult to localize and unrealistic. The sounds selected for the test are found in Table 1. These sound files were randomly assigned a code word and then randomly ordered so that the experimenter did not know which sound was being presented to the listener, and therefore could not bias the subjects response. After the experiment, the experimenter matched the code words tothe appropriate sounds and determined the location of each sound that was presented to each subject.

5.3.2 Results. The results of this experiment strongly supported the results of the initial investigation. Fifteen subjects were asked to listen to the collection of sound files listed in Table 1. Five subjects listened to all 25 sounds. The other ten subjects listened to a portion of the collection due to time constaints. Between 7 and 13 subjects listened to each sound. Listener's lateralized the azimuth and elevation of the sound sources well when only the direct path source was presented. Some listeners did experience front/back reversals, but the same listener's continued to experience front/back reversal regardless of the number of attenuated reflections added into the signal. Up/down confusions were less frequent with localization higher than intended being the primary problem. The results in extracranialization of the sounds were also similar. The quantitative results of this experiment are shown in Table 2. The "A" in the table stands for accurate, the "I" for inaccurate, and the "U" for unable to localize.

SOUND	θ (Deg.)	ϕ (Deg.)	NUM. REFL.	ATTENUATION
1	10	0	6	YES
2	10	0	4	YES
3	20	50	2	NO
4	20	50	6	YES
5	20	50	0	YES
6	20	50	0	NO
7	75	10	4	YES
8	75	10	6	YES
9	75	10	0	YES
10	75	10	2	NO
11	120	-25	6	YES
12	120	-25	2	YES
13	120	-25	0	NO
14	180	0	6	YES
15	180	0	4	YES
16	180	0	0	YES
17	180	0	0	NO
18	240	10	6	YES
19	240	10	4	YES
20	285	10	10	YES
21	285	10	6	YES
22	285	10	4	NO
23	285	10	2	NO
24	285	10	2	YES
25	285	10	0	NO

Table 1. Location of the Sounds used in the Double-Blind Experiment

SOUND	AZIMUTH			ELEVATION			DISTANCE		
	A	I	U	A	I	U	IN	ON	OUT
1	4	4	2	8	0	2	3	2	5
2	7	4	1	11	0	1	4	2	6
3	0	2	5	2	0	5	5	0	2
4	4	4	0	1	7	0	2	1	5
5	1	3	5	2	2	5	6	0	3
6	5	3	3	2	9	0	8	3	0
7	7	3	0	8	2	0	0	0	10
8	8	2	0	10	0	0	0	0	10
9	5	2	0	6	1	0	1	0	6
10	5	1	4	8	0	2	5	0	5
11	7	3	0	5	5	0	1	1	8
12	6	2	0	4	4	0	1	2	5
13	8	2	0	6	2	2	3	4	3
14	5	2	3	4	3	3	4	0	6
15	4	1	3	4	1	3	3	1	4
16	6	5	2	5	5	3	1	5	6
17	3	1	3	4	2	2	7	0	1
18	7	3	0	8	2	0	0	2	8
19	7	3	0	8	2	0	0	0	10
20	11	1	0	12	0	0	0	1	11
21	7	3	0	9	1	0	0	1	9
22	3	1	6	2	2	6	8	1	1
23	6	3	3	9	0	3	6	3	3
24	7	2	0	8	1	0	2	0	7
25	9	1	0	9	1	0	3	6	1

Table 2. Responses for the Double-Blind Experiment

When only the non-attenuated direct path was presented to the listeners, only about 20 percent reported that they heard the sound outside their head. When the direct path was attenuated, this percentage rose to 60 percent. Approximately half of the listeners reported hearing echoes, and were unable to localize signals that included only two non-attenuated reflections. This percentage of listeners unable to localize the sound due to echoes or "stereo" rose to nearly 90 percent when four non-attenuated reflections were included. When the reflections were attenuated; however, the number of listener's reporting localization outside their heads increased. With four attenuated reflections, approximately 80 percent of the listeners reported localization extracranially. With six or more attenuated reflections included in the signal, 90 percent localized extracranially.

Although the listeners reported hearing the sound outside their head when attenuated reflections were included in the signal, the perception of the distance to the source varied from 1 to 20 feet. Some subjects indicated that the volume of the signal provided information about the distance. Furthermore, they stated that not all of the signals appeared to be at the same volume.

5.4 Increasing the Virtual Head Size

The ITD cue was increased by 25, 50, and 100 percent in each of the previously generated sound files, and 6 subjects listener to the new sounds. A very interesting result was obtained. Listeners that localized sounds extracranially with the original ITD cues did not receive a recognizable benefit from the increased ITD cue. However, some of the listeners that had trouble extracranializing the original sounds were able to extracranialize the sounds containing increased ITD cues. Not coincidently, these listeners did have slightly larger than average head sizes.

5.5 Summary of Results

The results of this thesis research are encouraging. Analysis of the binaural sound signals generated in the binaural simulation show that a computer generated binaural signals presented over headphones without head coupling can be extracranialized. In particular, the addition of attenuated reflections in a binaural room simulation is key in increasing the listener's extracranialization of synthesized sound. Localization in azimuth and elevation is not noticably improved with the addition of these same cues. Furthermore, the distance between a listener's ears appears to be an important factor in extracranialization; consequently, the ITDs need to be adjusted to account for substantial differences in head size.

VI. Conclusions and Recommendations

6.1 Summary

During the course of this thesis work, two research objectives were focused upon, 1) the development of a mathematical model to describe the transfer function for a sound traveling from a source to a receiver in a rectangular room, and 2) the generation of binaural sound signals that can be extracranialized by a listener wearing headphones without head coupling. In short, this research centered on adding cues described in the mathematical model to a sound file using a Sun Sparc 2 Workstation and the ESPS software. The result was a binaural room simulation which yielded sounds that could be extracranialized by most listeners.

In the first step of the simulation, a monophonic sound file was created. The sound file contained approximately three seconds of male speech. Once the sound file was generated, it was then filtered to account for the HRTF for the direct sound source and various virtual sound sources representing reflections. Next, the existing sound files were appropriately delayed to account for the ITD cue and the travel distance to the listener. When these sound files were presented simultaneously to the appropriate ears of a listener, the resulting binaural signals ranged from direct path to direct path and all single reflections (one case included the first four double reflections). Sounds which accounted for three wall reflections or less could be lateralized intracranially by most listeners. However, when four or more reflections were included in the sound, listeners' ability to localize the sound source deteriorated. In most cases, regardless of the number of reflections added, the listeners failed to extracranialize the sound.

The next step in this research was to include the effects of attenuation resulting from the sound traveling through air. Once the attenuation was added into the sound files, the files were presented simultaneously to listeners. The resulting binaural sound files again ranged from direct path to direct path and all single reflections;

however, the addition of attenuation into these files created a drastically different affect on the listener. Once again, localization was in the proper quadrant. In this case, however, listeners reported extracranialization of the sound. In fact, when only the direct sound source was included, listeners perceived the signal containing attenuation to be farther away than the signal that contained only HRTF and time delay information. Furthermore, the more attenuated reflections (up to 10) that were included in the sound file, the greater the perceived distance.

6.2 Conclusions

The mathematical model developed in Chapter 3 provides an accurate description of the transfer function for a sound traveling from a source to a listener in a rectangular room. Furthermore, the implementation of pieces of this model into a binaural room simulation–in particular the addition of virtual sound sources to account for all single reflections, and the attenuation of all the sound sources due to the distance traveled through air–proved to be an effective means of providing the listener with an extracranialized sound. The implementation of these cues into the sound signal definitely increased the degree of telepresence reported by the listener. Localization in the azimuthal plane was only slightly improved over signals containing no attenuation information HRTF data. Many of the listeners were less accurate in localizing in elevation. A majority of the mistakes came from listeners localizing the sound higher in elevation than intended. Also, listeners encountered difficulty in localizing sounds on or near the median plane. This difficulty persisted regardless of the number of attenuated reflections included. Overall, the goal of creating an extracranialized binaural signal was accomplished by incorporating the effects of attenuation into reflections represented by virtual sound sources.

6.3 *Recommendations for Future Research*

This thesis successfully demonstrated that a binaural room simulation can be used to generate binaural signals that listeners can localize extracranially without head motion. However, the product of this research is far from a final product. There are several research projects that could be performed as follow-on research to this thesis.

One of the biggest problems with the method used in this research was the enormous amount of processing time and memory space required to generate the final binaural signal. This extensive processing of sounds clearly did not allow for any real-time applications. A real-time system that could incorporate head motion and include the necessary attenuated reflections would be an excellent topic. The development of a real-time system that included head motion should be effective in eliminating the problems of localizing on the median plane as well as the front/back reversals.

Several other follow-on projects to this research exist. One project would be to test both left and right ear dominant listeners to determine if ear dominance affects localization of the binaural sounds generated in this thesis. A test for right and left ear dominance was implemented at AFIT in the Fall of 1993 (7). A more extensive study of the effects of head size on localization could also be very informative. The goal of such a study might be to develop a set of "head sizes" that can be selected by a listener for the best fit. The use of IIR filters in place of the FIR filters used in this research for the HRTF's might provide a means of speeding up the pre-processing time. This could be a step in reaching the goal of a real-time virtual audio system.

Appendix A. MathCad Output

The Mathcad output on the following pages was used in the simulation performed at approximately 285 degrees of azimuth. Note that the notation is not identical to that used in the development of the mathematical model, but the concepts are used are the same.

The following Mathcad program calculates the distance traveled from a sound source to a receiver for a direct path and six reflections in a room. The angle of incidence at the receiver is also calculated.

Width of the room(W) in meters $W := 15$

Length of the room(L) in meters $L := 20$

Height of the room(H) in meters $H := 5$

Source Position $X := .58 \cdot W \quad Y := .65 \cdot L$

$$S := \begin{pmatrix} X \\ Y \\ Z \end{pmatrix} \qquad Z := .20 \cdot H$$

Receiver Position $\xi := .55 \cdot W \quad \eta := .75 \cdot L \quad \zeta := .15 \cdot H$

$$R := \begin{bmatrix} \xi \\ \eta \\ \zeta \end{bmatrix}$$

Let eF, eL, and eU be coordinate vectors of unit length where the following combination indicates that the receiver is looking straight ahead at wall 1.

$a_1 := 1 \quad b_1 := 0 \quad c_1 := 0 \quad a_2 := 0 \quad b_2 := 1 \quad c_2 := 0$

$a_3 := 0 \quad b_3 := 0 \quad c_3 := 1$

$$e_F := \begin{bmatrix} a_1 \\ b_1 \\ c_1 \end{bmatrix} \qquad e_L := \begin{bmatrix} a_2 \\ b_2 \\ c_2 \end{bmatrix} \qquad e_U := \begin{bmatrix} a_3 \\ b_3 \\ c_3 \end{bmatrix}$$

Left Ear Position $Ear_L := R + (.085 \cdot e_L)$

Right Ear Position $Ear_R := R - (.085 \cdot e_L)$

Direct Path:

Distance:

Left Ear:

$$Dist_{L0} := \sqrt{(\xi - X)^2 + \left[(\eta + b_2 \cdot .085) - Y\right]^2 + (\zeta - Z)^2}$$

$Dist_{L0} = 2.1476$

Right Ear:

$$Dist_{R0} := \sqrt{(\xi - X)^2 + \left[(\eta - b_2 \cdot .085) - Y\right]^2 + (\zeta - Z)^2}$$

$Dist_{R0} = 1.983$

Incidence Angle:(theta is azimuth and phi is elevation)

Left Ear: Azimuth

$$LS := S - Ear_L \qquad PLS_{eFeL} := (LS\cdot e_F)\cdot e_F + (LS\cdot e_L)\cdot e_L$$

$$f(PLS_{eFeL}) := \mathrm{if}(PLS_{eFeL}\cdot e_L<0,1,-1)$$

$$\theta_{L0} := f(PLS_{eFeL})\cdot \mathrm{acos}\left(\frac{PLS_{eFeL}\cdot e_F}{|PLS_{eFeL}|}\right) \qquad \theta_{L0} = 77.8208\cdot\mathrm{deg}$$

Elevation

$$\phi_{L0} := \frac{\pi}{2} - \mathrm{acos}\left(\frac{LS\cdot e_U}{|LS|}\right) \qquad \phi_{L0} = 6.6849\cdot\mathrm{deg}$$

Right Ear: Azimuth

$$RgtS := S - Ear_R \qquad PRgtS_{eFeL} := (RgtS\cdot e_F)\cdot e_F + (RgtS\cdot e_L)\cdot e_L$$

$$f(PRgtS_{eFeL}) := \mathrm{if}(PRgtS_{eFeL}\cdot e_L<0,1,-1)$$

$$\theta_{R0} := f(PRgtS_{eFeL})\cdot \mathrm{acos}\left(\frac{PRgtS_{eFeL}\cdot e_F}{|PRgtS_{eFeL}|}\right) \qquad \theta_{R0} = 76.7762\cdot\mathrm{deg}$$

Elevation

$$\phi_{R0} := \frac{\pi}{2} - \mathrm{acos}\left(\frac{RgtS\cdot e_U}{|RgtS|}\right) \qquad \phi_{R0} = 7.2427\cdot\mathrm{deg}$$

One Reflection off Wall 1:

$$S_1 := \begin{bmatrix} W + (W - X) \\ Y \\ Z \end{bmatrix}$$

Distance:

Left Ear:

$$Dist_{L1} := \sqrt{(\xi - (2\cdot W - X))^2 + \left[(\eta + b_2\cdot .085) - Y\right]^2 + (\zeta - Z)^2}$$

$$Dist_{L1} = 13.2179$$

Right Ear:

$$Dist_{R1} := \sqrt{(\xi - (2\cdot W - X))^2 + \left[(\eta - b_2\cdot .085) - Y\right]^2 + (\zeta - Z)^2}$$

$$Dist_{R1} = 13.1921$$

Incidence Angle:(theta is azimuth and phi is elevation)

Left Ear: Azimuth

$$LS1 := S_1 - Ear_L \qquad PLS1_{eFeL} := (LS1\cdot e_F)\cdot e_F + (LS1\cdot e_L)\cdot e_L$$

$$f(PLS1_{eFeL}) := if(PLS1_{eFeL}\cdot e_L<0, 1, -1)$$

$$\theta_{L1} := f(PLS1_{eFeL})\cdot acos\left(\frac{PLS1_{eFeL}\cdot e_F}{|PLS1_{eFeL}|}\right) \qquad \theta_{L1} = 9.0774\cdot deg$$

Elevation

$$\phi_{L1} := \frac{\pi}{2} - acos\left(\frac{LS1\cdot e_U}{|LS1|}\right) \qquad \phi_{L1} = 1.0837\cdot deg$$

Right Ear: Azimuth

$$RgtS1 := S_1 - Ear_R \qquad PRgtS1_{eFeL} := (RgtS1\cdot e_F)\cdot e_F + (RgtS1\cdot e_L)\cdot e_L$$

$$f(PRgtS1_{eFeL}) := if(PRgtS1_{eFeL}\cdot e_L<0, 1, -1)$$

$$\theta_{R1} := f(PRgtS1_{eFeL})\cdot acos\left(\frac{PRgtS1_{eFeL}\cdot e_F}{|PRgtS1_{eFeL}|}\right) \qquad \theta_{R1} = 8.3482\cdot deg$$

Elevation

$$\phi_{R1} := \frac{\pi}{2} - acos\left(\frac{RgtS1\cdot e_U}{|RgtS1|}\right) \qquad \phi_{R1} = 1.0859\cdot deg$$

One Reflection off Wall 2:

$$S_2 := \begin{pmatrix} X \\ -Y \\ Z \end{pmatrix}$$

Distance:

Left Ear:

$$Dist_{L2} := \sqrt{(\xi - X)^2 + \left[(\eta + b_2\cdot .085) + Y\right]^2 + (\zeta - Z)^2}$$

$$Dist_{L2} = 28.0897$$

Right Ear:

$$Dist_{R2} := \sqrt{(\xi - X)^2 + \left[(\eta - b_2\cdot .085) + Y\right]^2 + (\zeta - Z)^2}$$

$$Dist_{R2} = 27.9197$$

Incidence Angle:(theta is azimuth and phi is elevation)

Left Ear: Azimuth

$$LS2 := S_2 - Ear_L \qquad PLS2_{eFeL} := (LS2 \cdot e_F) \cdot e_F + (LS2 \cdot e_L) \cdot e_L$$

$$f(PLS2_{eFeL}) := if(PLS2_{eFeL} \cdot e_L < 0, 1, -1)$$

$$\theta_{L2} := f(PLS2_{eFeL}) \cdot acos\left(\frac{PLS2_{eFeL} \cdot e_F}{|PLS2_{eFeL}|}\right) \qquad \theta_{L2} = 89.082 \cdot deg$$

Elevation

$$\phi_{L2} := \frac{\pi}{2} - acos\left(\frac{LS2 \cdot e_U}{|LS2|}\right) \qquad \phi_{L2} = 0.5099 \cdot deg$$

Right Ear: Azimuth

$$RgtS2 := S_2 - Ear_R \qquad PRgtS2_{eFeL} := (RgtS2 \cdot e_F) \cdot e_F + (RgtS2 \cdot e_L) \cdot e_L$$

$$f(PRgtS2_{eFeL}) := if(PRgtS2_{eFeL} \cdot e_L < 0, 1, -1)$$

$$\theta_{R2} := f(PRgtS2_{eFeL}) \cdot acos\left(\frac{PRgtS2_{eFeL} \cdot e_F}{|PRgtS2_{eFeL}|}\right) \qquad \theta_{R2} = 89.0765 \cdot deg$$

Elevation

$$\phi_{R2} := \frac{\pi}{2} - acos\left(\frac{RgtS2 \cdot e_U}{|RgtS2|}\right) \qquad \phi_{R2} = 0.513 \cdot deg$$

One Reflection off Wall 3:

$$S_3 := \begin{pmatrix} -X \\ Y \\ Z \end{pmatrix}$$

Distance:

Left Ear:

$$Dist_{L3} := \sqrt{(\xi + X)^2 + \left[(\eta + b_2 \cdot .085) - Y\right]^2 + (\zeta - Z)^2}$$

$$Dist_{L3} = 17.0796$$

Right Ear:

$$Dist_{R3} := \sqrt{(\xi + X)^2 + \left[(\eta - b_2 \cdot .085) - Y\right]^2 + (\zeta - Z)^2}$$

$$Dist_{R3} = 17.0597$$

Incidence Angle:(theta is azimuth and phi is elevation)

Left Ear: Azimuth

$$LS3 \quad S_3 \quad Ear_L \qquad PLS3_{eFeL} \quad (LS3 \cdot e_F) \cdot e_F + (LS3 \cdot e_L) \cdot e_L$$

$$f(PLS3_{eFeL}) \quad if(PLS3_{eFeL} \cdot e_L < 0, 1, \ 1)$$

$$\theta_{L3} \quad f(PLS3_{eFeL}) \cdot acos\left(\frac{PLS3_{eFeL} \cdot e_F}{|PLS3_{eFeL}|}\right) \qquad \theta_{L3} = 172.9873 \cdot deg$$

Elevation

$$\phi_{L3} \quad \frac{\pi}{2} \quad acos\left(\frac{LS3 \cdot e_U}{|LS3|}\right) \qquad \phi_{L3} = 0.8387 \cdot deg$$

Right Ear: Azimuth

$$RgtS3 \quad S_3 \quad Ear_R \qquad PRgtS3_{eFeL} \quad (RgtS3 \cdot e_F) \cdot e_F + (RgtS3 \cdot e_L) \cdot e_L$$

$$f(PRgtS3_{eFeL}) \quad if(PRgtS3_{eFeL} \cdot e_L < 0, 1, \ 1)$$

$$\theta_{R3} \quad f(PRgtS3_{eFeL}) \cdot acos\left(\frac{PRgtS3_{eFeL} \cdot e_F}{|PRgtS3_{eFeL}|}\right) \qquad \theta_{R3} = 173.5541 \cdot deg$$

Elevation

$$\phi_{R3} \quad \frac{\pi}{2} \quad acos\left(\frac{RgtS3 \cdot e_U}{|RgtS3|}\right) \qquad \phi_{R3} = 0.8397 \cdot deg$$

One Reflection off Wall 4:

$$S_4 \quad \begin{vmatrix} X \\ L + (L \quad Y) \\ Z \end{vmatrix}$$

Distance:

Left Ear:

$$Dist_{L4} \quad \sqrt{(\xi \quad X)^2 + \left[(\eta + b_2 \cdot .085) \quad (2 \cdot L \quad Y)\right]^2 + (\zeta \quad Z)^2}$$

$$Dist_{L4} = 11.9261$$

Right Ear:

$$Dist_{R4} \quad \sqrt{(\xi \quad X)^2 + \left[(\eta \quad b_2 \cdot .085) \quad (2 \cdot L \quad Y)\right]^2 + (\zeta \quad Z)^2}$$

$$Dist_{R4} = 12.096$$

Incidence Angle:(theta is azimuth and phi is elevation)

Left Ear: Azimuth

$$LS4 := S_4 - Ear_L \qquad PLS4_{eFeL} := (LS4 \cdot e_F) \cdot e_F + (LS4 \cdot e_L) \cdot e_L$$

$$f(PLS4_{eFeL}) := \mathrm{if}(PLS4_{eFeL} \cdot e_L < 0, 1, -1)$$

$$\theta_{L4} := f(PLS4_{eFeL}) \cdot \mathrm{acos}\left(\frac{PLS4_{eFeL} \cdot e_F}{|PLS4_{eFeL}|}\right) \qquad \theta_{L4} = -87.8371 \cdot \mathrm{deg}$$

Elevation

$$\phi_{L4} := \frac{\pi}{2} - \mathrm{acos}\left(\frac{LS4 \cdot e_U}{|LS4|}\right) \qquad \phi_{L4} = 1.2011 \cdot \mathrm{deg}$$

Right Ear: Azimuth

$$RgtS4 := S_4 - Ear_R \qquad PRgtS4_{eFeL} := (RgtS4 \cdot e_F) \cdot e_F + (RgtS4 \cdot e_L) \cdot e_L$$

$$f(PRgtS4_{eFeL}) := \mathrm{if}(PRgtS4_{eFeL} \cdot e_L < 0, 1, -1)$$

$$\theta_{R4} := f(PRgtS4_{eFeL}) \cdot \mathrm{acos}\left(\frac{PRgtS4_{eFeL} \cdot e_F}{|PRgtS4_{eFeL}|}\right) \qquad \theta_{R4} = -87.8675 \cdot \mathrm{deg}$$

Elevation

$$\phi_{R4} := \frac{\pi}{2} - \mathrm{acos}\left(\frac{RgtS4 \cdot e_U}{|RgtS4|}\right) \qquad \phi_{R4} = 1.1843 \cdot \mathrm{deg}$$

One Reflection off Floor:

$$S_F := \begin{pmatrix} X \\ Y \\ -Z \end{pmatrix}$$

Distance:

Left Ear:

$$Dist_{LF} := \sqrt{(\xi - X)^2 + \left[(\eta + b_2 \cdot .085) - Y\right]^2 + (\zeta + Z)^2}$$

$$Dist_{LF} = 2.759$$

Right Ear:

$$Dist_{RF} := \sqrt{(\xi - X)^2 + \left[(\eta - b_2 \cdot .085) - Y\right]^2 + (\zeta + Z)^2}$$

$$Dist_{RF} = 2.6329$$

Incidence Angle:(theta is azimuth and phi is elevation)

Left Ear: Azimuth

$$LSF := S_F - Ear_L \qquad PLSF_{eFeL} := (LSF \cdot e_F) \cdot e_F + (LSF \cdot e_L) \cdot e_L$$

$$f(PLSF_{eFeL}) := if(PLSF_{eFeL} \cdot e_L < 0, 1, -1)$$

$$\theta_{LF} := f(PLSF_{eFeL}) \cdot acos\left(\frac{PLSF_{eFeL} \cdot e_F}{|PLSF_{eFeL}|}\right) \qquad \theta_{LF} = 77.8208 \cdot deg$$

Elevation

$$\phi_{LF} := \frac{\pi}{2} - acos\left(\frac{LSF \cdot e_U}{|LSF|}\right) \qquad \phi_{LF} = -39.3667 \cdot deg$$

Right Ear: Azimuth

$$RgtSF := S_F - Ear_R \qquad PRgtSF_{eFeL} := (RgtSF \cdot e_F) \cdot e_F + (RgtSF \cdot e_L) \cdot e_L$$

$$f(PRgtSF_{eFeL}) := if(PRgtSF_{eFeL} \cdot e_L < 0, 1, -1)$$

$$\theta_{RF} := f(PRgtSF_{eFeL}) \cdot acos\left(\frac{PRgtSF_{eFeL} \cdot e_F}{|PRgtSF_{eFeL}|}\right) \qquad \theta_{RF} = 76.7762 \cdot deg$$

Elevation

$$\phi_{RF} := \frac{\pi}{2} - acos\left(\frac{RgtSF \cdot e_U}{|RgtSF|}\right) \qquad \phi_{RF} = -41.6565 \cdot deg$$

One Reflection off Ceiling:

$$S_C := \begin{bmatrix} X \\ Y \\ H + (H - Z) \end{bmatrix}$$

Distance:

Left Ear:

$$Dist_{LC} := \sqrt{(\xi - X)^2 + \left[(\eta + b_2 \cdot .085) - Y\right]^2 + (\zeta - (2 \cdot H - Z))^2}$$

$$Dist_{LC} = 8.5213$$

Right Ear:

$$Dist_{RC} := \sqrt{(\xi - X)^2 + \left[(\eta - b_2 \cdot .085) - Y\right]^2 + (\zeta - (2 \cdot H - Z))^2}$$

$$Dist_{RC} = 8.4813$$

Incidence Angle:(theta is azimuth and phi is elevation)

Left Ear: Azimuth

$$LSC := S_C - Ear_L \qquad PLSC_{eFeL} := (LSC \cdot e_F) \cdot e_F + (LSC \cdot e_L) \cdot e_L$$

$$f(PLSC_{eFeL}) := if(PLSC_{eFeL} \cdot e_L < 0, 1, -1)$$

$$\theta_{LC} := f(PLSC_{eFeL}) \cdot acos\left(\frac{PLSC_{eFeL} \cdot e_F}{|PLSC_{eFeL}|}\right) \qquad \theta_{LC} = 77.8208 \cdot deg$$

Elevation

$$\phi_{LC} := \frac{\pi}{2} - acos\left(\frac{LSC \cdot e_U}{|LSC|}\right) \qquad \phi_{LC} = 75.5038 \cdot deg$$

Right Ear: Azimuth

$$RgtSC := S_C - Ear_R \qquad PRgtSC_{eFeL} := (RgtSC \cdot e_F) \cdot e_F + (RgtSC \cdot e_L) \cdot e_L$$

$$f(PRgtSC_{eFeL}) := if(PRgtSC_{eFeL} \cdot e_L < 0, 1, -1)$$

$$\theta_{RC} := f(PRgtSC_{eFeL}) \cdot acos\left(\frac{PRgtSC_{eFeL} \cdot e_F}{|PRgtSC_{eFeL}|}\right) \qquad \theta_{RC} = 76.7762 \cdot deg$$

Elevation

$$\phi_{RC} := \frac{\pi}{2} - acos\left(\frac{RgtSC \cdot e_U}{|RgtSC|}\right) \qquad \phi_{RC} = 76.5886 \cdot deg$$

Two Reflections off Floor and Wall1:

$$S_{F1} := \begin{bmatrix} W + (W - X) \\ Y \\ Z \end{bmatrix}$$

Distance:

Left Ear:

$$Dist_{LF1} := \sqrt{(\xi - (2 \cdot W - X))^2 + \left[(\eta + b_2 \cdot .085) - Y\right]^2 + (\zeta + Z)^2}$$

$$Dist_{LF1} = 13.3309$$

Right Ear:

$$Dist_{RF1} := \sqrt{(\xi - (2 \cdot W - X))^2 + \left[(\eta - b_2 \cdot .085) - Y\right]^2 + (\zeta + Z)^2}$$

$$Dist_{RF1} = 13.3053$$

Incidence Angle:(theta is azimuth and phi is elevation)

Left Ear: Azimuth

$$LSF1 := S_{F1} - Ear_L \quad PLSF1_{eFeL} := (LSF1 \cdot e_F) \cdot e_F + (LSF1 \cdot e_L) \cdot e_L$$

$$f(PLSF1_{eFeL}) := if(PLSF1_{eFeL} \cdot e_L < 0, 1, -1)$$

$$\theta_{LF1} := f(PLSF1_{eFeL}) \cdot acos\left(\frac{PLSF1_{eFeL} \cdot e_F}{|PLSF1_{eFeL}|}\right) \quad \theta_{LF1} = 9.0774 \cdot deg$$

Elevation

$$\phi_{LF1} := \frac{\pi}{2} - acos\left(\frac{LSF1 \cdot e_U}{|LSF1|}\right) \quad \phi_{LF1} = 7.5432 \cdot deg$$

Right Ear: Azimuth

$$RgtSF1 := S_{F1} - Ear_R \quad PRgtSF1_{eFeL} := (RgtSF1 \cdot e_F) \cdot e_F + (RgtSF1 \cdot e_L) \cdot e_L$$

$$f(PRgtSF1_{eFeL}) := if(PRgtSF1_{eFeL} \cdot e_L < 0, 1, -1)$$

$$\theta_{RF1} := f(PRgtSF1_{eFeL}) \cdot acos\left(\frac{PRgtSF1_{eFeL} \cdot e_F}{|PRgtSF1_{eFeL}|}\right) \quad \theta_{RF1} = 8.3482 \cdot deg$$

Elevation

$$\phi_{RF1} := \frac{\pi}{2} - acos\left(\frac{RgtSF1 \cdot e_U}{|RgtSF1|}\right) \quad \phi_{RF1} = -7.5578 \cdot deg$$

Two Reflections off Floor and Wall3:

$$S_{F3} := \begin{pmatrix} -X \\ Y \\ -Z \end{pmatrix}$$

Distance:

Left Ear:

$$Dist_{LF3} := \sqrt{(\xi + X)^2 + \left[(\eta + b_2 \cdot .085) - Y\right]^2 + (\zeta + Z)^2}$$

$$Dist_{LF3} = 17.1672$$

Right Ear:

$$Dist_{RF3} := \sqrt{(\xi + X)^2 + \left[(\eta - b_2 \cdot .085) - Y\right]^2 + (\zeta + Z)^2}$$

$$Dist_{RF3} = 17.1474$$

Incidence Angle:(theta is azimuth and phi is elevation)

Left Ear: Azimuth

$$LSF3 := S_{F3} - Ear_L \quad PLSF3_{eFeL} := (LSF3 \cdot e_F) \cdot e_F + (LSF3 \cdot e_L) \cdot e_L$$

$$f(PLSF3_{eFeL}) := if(PLSF3_{eFeL} \cdot e_L < 0, 1, -1)$$

$$\theta_{LF3} := f(PLSF3_{eFeL}) \cdot acos\left(\frac{PLSF3_{eFeL} \cdot e_F}{|PLSF3_{eFeL}|}\right) \quad \theta_{LF3} = 172.9873 \cdot deg$$

Elevation

$$\phi_{LF3} := \frac{\pi}{2} - acos\left(\frac{LSF3 \cdot e_U}{|LSF3|}\right) \quad \phi_{LF3} = -5.8508 \cdot deg$$

Right Ear: Azimuth

$$RgtSF3 := S_{F3} - Ear_R \quad PRgtSF3_{eFeL} := (RgtSF3 \cdot e_F) \cdot e_F + (RgtSF3 \cdot e_L) \cdot e_L$$

$$f(PRgtSF3_{eFeL}) := if(PRgtSF3_{eFeL} \cdot e_L < 0, 1, -1)$$

$$\theta_{RF3} := f(PRgtSF3_{eFeL}) \cdot acos\left(\frac{PRgtSF3_{eFeL} \cdot e_F}{|PRgtSF3_{eFeL}|}\right) \quad \theta_{RF3} = 173.5541 \cdot deg$$

Elevation

$$\phi_{RF3} := \frac{\pi}{2} - acos\left(\frac{RgtSF3 \cdot e_U}{|RgtSF3|}\right) \quad \phi_{RF3} = -5.8576 \cdot deg$$

Two Reflections off Floor and Ceiling:

$$S_{FC} := \begin{bmatrix} X \\ Y \\ H + (H + Z) \end{bmatrix}$$

Distance:

Left Ear:

$$Dist_{LFC} := \sqrt{(\xi - X)^2 + \left[(\eta + b_2 \cdot .085) - Y\right]^2 + (\zeta - (2 \cdot H + Z))^2}$$

$$Dist_{LFC} = 10.4696$$

Right Ear:

$$Dist_{RFC} := \sqrt{(\xi - X)^2 + \left[(\eta - b_2 \cdot .085) - Y\right]^2 + (\zeta - (2 \cdot H + Z))^2}$$

$$Dist_{RFC} = 10.4371$$

Incidence Angle:(theta is azimuth and phi is elevation)

Left Ear: Azimuth

$$LSFC := S_{FC} - Ear_L \quad PLSFC_{eFeL} := (LSFC\cdot e_F)\cdot e_F + (LSFC\cdot e_L)\cdot e_L$$

$$f(PLSFC_{eFeL}) := if(PLSFC_{eFeL}\cdot e_L<0, 1, -1)$$

$$\theta_{LFC} := f(PLSFC_{eFeL})\cdot acos\left(\frac{PLSFC_{eFeL}\cdot e_F}{|PLSFC_{eFeL}|}\right) \quad \theta_{LFC} = 77.8208\cdot deg$$

Elevation

$$\phi_{LFC} := \frac{\pi}{2} - acos\left(\frac{LSFC\cdot e_U}{|LSFC|}\right) \quad \phi_{LFC} = 78.2446\cdot deg$$

Right Ear: Azimuth

$$RgtSFC := S_{FC} - Ear_R \, PRgtSFC_{eFeL} := (RgtSFC\cdot e_F)\cdot e_F + (RgtSFC\cdot e_L)\cdot e_L$$

$$f(PRgtSFC_{eFeL}) := if(PRgtSFC_{eFeL}\cdot e_L<0, 1, -1)$$

$$\theta_{RFC} := f(PRgtSFC_{eFeL})\cdot acos\left(\frac{PRgtSFC_{eFeL}\cdot e_F}{|PRgtSFC_{eFeL}|}\right) \quad \theta_{RFC} = 76.7762\cdot deg$$

Elevation

$$\phi_{RFC} := \frac{\pi}{2} - acos\left(\frac{RgtSFC\cdot e_U}{|RgtSFC|}\right) \quad \phi_{RFC} = 79.136\cdot deg$$

Two Reflections off Ceiling and Floor:

$$S_{CF} := \begin{bmatrix} X \\ Y \\ -H - (H + Z) \end{bmatrix}$$

Distance:

Left Ear:

$$Dist_{LCF} := \sqrt{(\xi - X)^2 + \left[(\eta + b_2\cdot .085) - Y\right]^2 + (\zeta + (2\cdot H + Z))^2}$$

$$Dist_{LCF} = 11.942$$

Right Ear:

$$Dist_{RCF} := \sqrt{(\xi - X)^2 + \left[(\eta - b_2\cdot .085) - Y\right]^2 + (\zeta + (2\cdot H + Z))^2}$$

$$Dist_{RCF} = 11.9135$$

Incidence Angle:(theta is azimuth and phi is elevation)

Left Ear: Azimuth

$$LSCF := S_{CF} - Ear_L \qquad PLSCF_{eFeL} := (LSCF \cdot e_F) \cdot e_F + (LSCF \cdot e_L) \cdot e_L$$

$$f(PLSCF_{eFeL}) := if(PLSCF_{eFeL} \cdot e_L < 0, 1, 1)$$

$$\theta_{LCF} := f(PLSCF_{eFeL}) \cdot acos\left(\frac{PLSCF_{eFeL} \cdot e_F}{|PLSCF_{eFeL}|}\right) \qquad \theta_{LCF} = 77.8208 \cdot deg$$

Elevation

$$\phi_{LCF} := \frac{\pi}{2} - acos\left(\frac{LSCF \cdot e_U}{|LSCF|}\right) \qquad \phi_{LCF} = 79.711 \cdot deg$$

Right Ear: Azimuth

$$RgtSCF := S_{CF} - Ear_R \qquad PRgtSCF_{eFeL} := (RgtSCF \cdot e_F) \cdot e_F + (RgtSCF \cdot e_L) \cdot e_L$$

$$f(PRgtSCF_{eFeL}) := if(PRgtSCF_{eFeL} \cdot e_L < 0, 1, 1)$$

$$\theta_{RCF} := f(PRgtSCF_{eFeL}) \cdot acos\left(\frac{PRgtSCF_{eFeL} \cdot e_F}{|PRgtSCF_{eFeL}|}\right) \qquad \theta_{RCF} = 76.7762 \cdot deg$$

Elevation

$$\phi_{RCF} := \frac{\pi}{2} - acos\left(\frac{RgtSCF \cdot e_U}{|RgtSCF|}\right) \qquad \phi_{RCF} = -80.4958 \cdot deg$$

Appendix B. Speaker Locations for HRTF Measurements

This appendix contains the data for speaker location used in measuring the HRTF's. The following list provides the location and ITD for each of the 272 speakers on the dome.

Speaker Locations and Interaural Time Delays
(angles in degrees, and time delays in microseconds)

Speaker	Azimuth	Elev.	ITD	Speaker	Azimuth	Elev.	ITD
1	10.57	0	90	51	90	57.25	320
2	31.72	0	260	52	90	44.71	450
3	44.71	0	372	53	90	31.72	573
4	57.98	0	470	54	90	10.57	690
5	69.09	0	570	55	90	-10.57	648
6	82.41	0	658	56	90	-31.72	510
7	97.59	0	730	57	90	-44.71	428
8	111.09	0	770	58	90	-57.25	334
9	123.02	0	480	59	90	-69.09	210
10	135.29	0	388	60	90	-82.41	66
11	148.28	0	280	61	270	-82.41	112
12	169.43	0	100	62	270	-69.09	264
13	190.57	0	85	63	270	-57.25	396
14	211.72	0	274	64	270	-44.71	460
15	224.71	0	390	65	270	-31.72	556
16	236.98	0	480	66	270	-10.57	684
17	249.09	0	784	67	270	10.57	680
18	262.41	0	744	68	270	31.72	564
19	277.59	0	675	69	270	44.71	423
20	290.91	0	586	70	270	57.25	320
21	303.02	0	484	71	270	69.09	180
22	315.25	0	380	72	270	82.41	50

23	328.28	0	265	73	20.3	67.6	80
24	349.43	0	90	74	20.5	52.4	120
25	0	79.43	0	75	16.0	39.8	125
26	0	58.28	0	76	12.9	27.1	114
27	0	45.29	0	77	10.6	14.7	100
28	0	32.77	0	78	55.0	72.0	150
29	0	20.91	0	79	40.9	58.4	282
30	0	7.59	0	80	34.9	44.2	212
31	0	-7.59	0	81	29.3	32.4	271
32	0	-20.91	0	82	24.9	20.1	215
33	0	-32.77	0	83	21.1	7.6	206
34	0	-45.29	0	84	66.3	60.2	344
35	0	-58.28	0	85	51.2	47.4	414
36	0	-79.43	0	86	45.0	35.3	449
37	180	-79.43	0	87	40.1	24.2	354
38	180	-58.28	0	88	35.8	12.3	324
39	180	-45.29	0	89	71.7	47.6	498
40	180	-32.77	0	90	59.5	36.0	532
41	180	-20.91	0	91	53.9	24.4	400
42	180	-7.59	0	92	49.1	12.2	411
43	180	7.59	0	93	74.9	34.8	558
44	180	20.91	0	94	68.2	23.3	648
45	180	32.77	0	95	62.4	11.5	510
46	180	45.29	0	96	81.9	20.9	630
47	180	58.28	0	97	75.1	10.2	620
48	180	79.43	0	98	159.7	67.6	90
49	90	82.41	54	99	159.5	52.4	130
50	90	69.09	186	100	164.0	39.8	130

Speaker	Azimuth	Elev.	ITD	Speaker	Azimuth	Elev.	ITD
101	167.1	27.1	120	156	330.7	32.4	210
102	169.4	14.7	105	157	335.1	20.1	194
103	125.0	72.0	160	158	338.9	7.6	170
104	139.1	58.4	210	159	293.7	60.2	235
105	145.1	44.2	250	160	308.8	47.4	303
106	150.7	32.4	245	161	315.0	35.3	300
107	155.1	20.1	230	162	319.9	24.2	303
108	158.9	7.6	200	163	324.2	12.3	292
109	113.7	60.2	280	164	288.3	47.6	360

110	128.8	47.4	327	165	300.5	36.0	390
111	135.0	35.3	350	166	306.1	24.4	404
112	139.9	24.2	340	167	310.9	12.2	407
113	144.2	12.3	310	168	285.1	34.8	470
114	108.3	47.6	400	169	291.8	23.3	510
115	120.5	36.0	430	170	297.6	11.5	518
116	126.1	24.4	440	171	278.1	20.9	615
117	130.9	12.2	430	172	284.9	10.2	628
118	105.1	34.8	520	173	10.6	-14.7	80
119	111.8	23.3	543	174	12.9	-27.1	90
120	117.6	11.5	530	175	16.0	-39.8	80
121	98.1	20.9	630	176	20.5	-52.4	90
122	104.9	10.2	730	177	20.3	-67.6	60
123	200.3	67.6	56	178	21.1	-7.6	172
124	200.5	52.4	110	179	24.9	-20.1	190
125	196.0	39.8	106	180	29.3	-32.4	205
126	192.9	27.1	90	181	34.9	-44.2	180
127	190.6	14.7	80	182	40.9	-58.4	158
128	235.0	72.0	124	183	55.0	-72.0	370
129	220.9	58.4	180	184	35.8	-12.3	280
130	214.9	44.2	220	185	40.1	-24.2	290
131	209.3	32.4	220	186	45.0	-35.3	270
132	204.9	20.1	205	187	51.2	-47.4	250
133	201.1	7.6	180	188	66.6	-60.2	266
134	246.3	60.2	250	189	49.1	-12.2	386
135	231.2	47.4	299	190	53.9	-24.2	394
136	225.0	35.3	320	191	59.5	-36.0	340
137	220.1	24.2	323	192	71.7	-47.6	540
138	215.8	12.2	310	193	62.4	-11.5	490
139	251.7	47.6	370	194	68.2	-23.2	470
140	239.5	36.0	431	195	74.9	-34.8	458
141	233.9	24.4	430	196	75.1	-10.2	590
142	229.1	12.2	420	197	81.9	-20.9	556
143	254.9	34.8	493	198	169.4	-14.7	90
144	248.2	23.3	530	199	167.1	-27.1	92
145	242.4	11.5	523	200	164.0	-39.8	100
146	261.9	20.9	610	201	159.5	-52.4	102
147	255.1	10.2	750	202	159.7	-67.6	60
148	339.7	67.6	50	203	158.9	-7.6	190
149	339.5	52.4	100	204	155.1	-20.1	200
150	344.0	39.8	113	205	150.7	-32.4	200
151	347.1	27.1	90	206	145.1	-44.2	200

152	349.4	14.7	80	207	139.1	-58.4	170
153	305.0	72.0	120	208	125.0	-72.0	300
154	319.1	58.4	170	209	144.2	-12.3	370
155	325.1	44.2	210	210	139.9	-24.2	290

Speaker	Azimuth	Elev.	ITD	Speaker	Azimuth	Elev.	ITD
211	135.0	-35.3	280	242	251.7	-47.6	379
212	128.8	-47.4	270	243	242.4	-11.5	490
213	113.7	-60.2	298	244	248.2	-23.2	470
214	130.9	-12.2	400	245	254.9	-34.8	470
215	126.1	-24.4	380	246	255.1	-10.2	790
216	120.5	-36.0	360	247	261.9	-20.9	576
217	108.3	-47.6	368	248	349.4	-14.7	90
218	117.6	-11.5	480	249	347.1	-27.1	100
219	111.8	-23.3	462	250	344.0	-39.8	95
220	105.1	-34.8	454	251	339.5	-52.4	100
221	104.9	-10.2	780	252	339.7	-67.6	72
222	98.1	-20.9	550	253	338.9	-7.6	180
223	190.6	-14.7	83	254	335.1	-20.1	196
224	192.9	-27.1	90	255	330.7	-32.4	208
225	196.0	-39.8	90	256	325.1	-44.2	190
226	200.5	-52.4	110	257	319.1	-58.4	170
227	200.3	-67.6	80	258	305.0	-72.0	320
228	201.1	-7.6	180	259	324.2	-12.3	296
229	204.9	-20.1	190	260	319.9	-24.2	306
230	209.3	-32.4	207	261	315.0	-35.3	288
231	214.9	-44.2	212	262	308.8	-47.4	270
232	220.9	-58.4	190	263	293.7	-60.2	418
233	235.0	-72.0	348	264	310.9	-12.2	400
234	215.8	-12.3	300	265	306.1	-24.4	390
235	220.1	-24.2	300	266	300.5	-36.0	360
236	225.0	-35.3	290	267	288.3	-47.6	560
237	231.2	-47.4	286	268	297.6	-11.5	510
238	246.6	-60.2	307	269	291.8	-23.3	510
239	229.1	-12.2	402	270	285.1	-34.8	500
240	233.9	-24.4	390	271	284.9	-10.2	610
241	239.5	-36.0	372	272	278.1	-20.9	590

Appendix C. Tables Comparing the HRTF Angles to the MathCad Output

The following seven tables provide a comparison of the angles of azimuth and elevation generated by the MathCad template to the angles of azimuth and elevation of the nearest speaker location. Within each table, the actual source and all virtual sources used in the individual simulation are compared. Recall that the orientation used to measure azimuth was different for the AAMRL HRTF measurements (speaker locations) and the MathCad template. Figure 1 shows the orientation used in determining the azimuth angle for speaker location, while figure 8 shows the orientation used to measure azimuth in the MathCad template. The number in parenthesis in the MathCad Azimuth column is the azimuth angle given by the orientation used in the HRTF measurements.

SOURCE	SPEAKER		MATHCAD	
	AZIMUTH	ELEVATION	AZIMUTH	ELEVATION
DIRECT PATH	10.57	0	-11(11)	0
REF. OFF WALL1	10.57	0	-3(3)	0
REF. OFF WALL2	277.59	0	84(276)	0
REF. OFF WALL3	169.43	0	-178(178)	0
REF. OFF WALL4	69.09	0	-72(72)	0
REF. OFF FLOOR	12.9	-27.1	-11(11)	-26
REF. OFF CEIL.	20.3	67.6	-11(11)	70

Table 3. Comparison of Mathcad Output to Speaker Location used in the Simulation for a Source at -10 Degrees Azimuth

SOURCE	SPEAKER		MATHCAD	
	AZIMUTH	ELEVATION	AZIMUTH	ELEVATION
DIRECT PATH	169.43	0	180(180)	0
REF. OFF WALL1	10.57	0	0(0)	0
REF. OFF WALL2	262.41	0	98(262)	0
REF. OFF WALL3	169.43	0	180(180)	0
REF. OFF WALL4	111.09	0	-112(112)	0
REF. OFF FLOOR	180	-20.91	180(180)	-20
REF. OFF CEIL.	180	58.28	180(180)	64.5

Table 4. Comparison of Mathcad Output to Speaker Location used in the Simulation for a Source at 180 Degrees Azimuth

SOURCE	SPEAKER		MATHCAD	
	AZIMUTH	ELEVATION	AZIMUTH	ELEVATION
DIRECT PATH	20.5	52.4	-18(18)	51
REF. OFF WALL1	0	20.91	-5(5)	20
REF. OFF WALL2	270	10.57	84(276)	7
REF. OFF WALL3	169.4	14.7	-177(177)	11
REF. OFF WALL4	68.2	23.3	-71(71)	23
REF. OFF FLOOR	20.3	-67.6	-18(18)	-60
REF. OFF CEIL.	20.5	52.4	-18(18)	55

Table 5. Comparison of Mathcad Output to Speaker Location used in the Simulation for a Source at 20 Degrees Azimuth 50 Degrees Elevation

SOURCE	SPEAKER		MATHCAD	
	AZIMUTH	ELEVATION	AZIMUTH	ELEVATION
DIRECT PATH	242.4	11.5	117(243)	6
REF. OFF WALL1	349.43	0	8(352)	1
REF. OFF WALL2	262.41	0	92(268)	.5
REF. OFF WALL3	190.57	0	172(188)	1
REF. OFF WALL4	97.59	0	-95(95)	1
REF. OFF FLOOR	239.5	-36	117(243)	-37
REF. OFF CEIL.	235.0	72	117(243)	74

Table 6. Comparison of Mathcad Output to Speaker Location used in the Simulation for a Source at 120 Degrees Azimuth

SOURCE	SPEAKER		MATHCAD	
	AZIMUTH	ELEVATION	AZIMUTH	ELEVATION
DIRECT PATH	126.1	-24.4	-117(117)	-17
REF. OFF WALL1	10.57	0	-8(8)	-3
REF. OFF WALL2	277.59	0	92(268)	-1
REF. OFF WALL3	169.43	0	-173(173)	-2.5
REF. OFF WALL4	104.9	-10.2	-97(97)	-5
REF. OFF FLOOR	120.5	-36.0	-117(117)	-19
REF. OFF CEIL.	125.0	72.0	-117(117)	76

Table 7. Comparison of Mathcad Output to Speaker Location used in the Simulation for a Source at -120 Degrees Azimuth

SOURCE	SPEAKER		MATHCAD	
	AZIMUTH	ELEVATION	AZIMUTH	ELEVATION
DIRECT PATH	75.1	10.2	-77(77)	7
REF. OFF WALL1	10.57	0	-9(9)	1
REF. OFF WALL2	277.59	0	89(271)	.5
REF. OFF WALL3	169.43	0	-173(173)	1
REF. OFF WALL4	82.41	0	-87(87)	2
REF. OFF FLOOR	74.9	-34.8	-77(77)	-40
REF. OFF CEIL.	90	82.41	-77(77)	76

Table 8. Comparison of Mathcad Output to Speaker Location used in the Simulation for a Source at -75 Degrees Azimuth

SOURCE	SPEAKER		MATHCAD	
	AZIMUTH	ELEVATION	AZIMUTH	ELEVATION
DIRECT PATH	284.9	10.2	77(283)	7
REF. OFF WALL1	349.43	0	9(351)	1
REF. OFF WALL2	277.59	0	89(271)	1
REF. OFF WALL3	190.57	0	173(187)	1
REF. OFF WALL4	82.41	0	-88(88)	1
REF. OFF FLOOR	285.1	-34.8	77(283)	-40
REF. OFF CEIL.	293.7	60.2	77(283)	76
FLOOR, WALL1	338.9	-7.6	9(351)	-8
FLOOR, WALL3	180	-7.59	173(187)	-6
FLOOR, CEILING	270	82.41	77(283)	79
CEILING, FLOOR	270	-82.41	77(283)	-80

Table 9. Comparison of Mathcad Output to Speaker Location used in the Simulation for a Source at 75 Degrees Azimuth

Appendix D. ESPS Manual Pages and C Code

This appendix contains the manual pages for each of the commands used in developing the binaural signals produced in this thesis. The C Code for the delay command is also included.

D.1 Addsd

```
Manual Pages for the ESPS Command "addsd":

     NAME
          addsd - add ESPS sampled data files with optional scaling
          multsd - multiply ESPS sampled data files with optional
          scaling
     SYNOPSIS
          addsd [ -x debug_level ] [ -r range ] [ -p range ] [ -g
          scale ] [ -z ] [ -t ] file1 file2 file3
          multsd [ -x debug_level ] [ -r range ] [ -p range ] [ -g
          scale ] [ -z ] [ -t ] file1 file2 file3
     DESCRIPTION
          Addsd (1-ESPS) and multsd (1-ESPS) are the same binary file.
          The function the program does (either adding or multiplying)
          depends on the name that is used to call it. The options and
          syntax are the same for both adding and multiplying. Because
          the calling name is used in the program logic, both addsd
          (1-ESPS) and multsd (1-ESPS) cannot be linked or copied to
          new names.  Below only addsd (1-ESPS) is described, but the
          description of multsd (1-ESPS) is completely analogous.

          Addsd takes sampled data from file1, adds it sample-by-
          sample to the sampled data in file2, possibly scaling the
          data in file2 first, and outputs the results as an ESPS
          FEA_SD file file3. If there are not enough records in file2,
          and if the -t option is not used, addsd reuses file2,
          starting with its first record.

          Both file1 and file2 must be ESPS FEA_SD files, and they
          must have the same sampling frequency (record_freq);
          otherwise addsd exits with an error message. The output file
```

data type is selected to "cover" the two input data types. That is, all values of the input types can be stored in the output type. For example, if one file is type SHORT_CPLX and the other is type FLOAT, the output type is FLOAT_CPLX. If "-" is supplied in place of file1, then standard input is used. If "-" is supplied in place of file3, standard output is used.

OPTIONS

The following options are supported:

-x debug_level

If debug_level is positive, addsd prints debugging messages and other information on the standard error output. The messages proliferate as the debug_level increases. If debug_level is 0, no messages are printed. The default is 0. Levels up through 2 are supported currently.

-p range

Selects a subrange of records from file1 using the format start-end or start:end. Either start or end may be omitted, in which case the omitted parameter defaults respectively to the start or end of file1. The first record in file1 is considered to be frame 1, regardless of its position relative to any original source file. The default range is the entire input file file1. The selected subrange from file1 is then added to the (possibly scaled) data from the file2, starting with the first record of file2. If the subrange does not exist in file1, addsd exits with an

-r range

-r is a synonym for -p.

-g scale

Causes addsd to multiply the data in file2 by scale before adding it to the data in file1. The format for scale is either integer or floating point.

-z Supresses warning messages that normally are generated if the contents of file2 are used more than once.

-t Truncates the processing if there are not enough records in file2. In this case, file3 will contain as many records as there are in file2 or, if the -p option is used, as many records as in the intersection of range and the full range of file2.

ESPS PARAMETERS

The ESPS parameter file is not read by addsd.

ESPS COMMON

If Common processing is enabled, the following items are read from the ESPS Common File provided that file1 is not standard input, and provided that the Common item filename matches the input file name file1:

start - integer

This is the starting point in file1.

nan - integer

This is the number of points to add from file1.

If start and/or nan are not given in the common file, or if the common file can't be opened for reading, then start defaults to the beginning of the file and nan defaults to the number of points in the file. In all cases, values of start and nan are ignored if the -p is used.

If Common processing is enabled, the following items are written into the ESPS Common file provided that the output file is not <stdout>:

start - integer

The starting point (1) in the output file file3.

nan - integer

The number of points in the output file file3

prog - string

This is the name of the program (addsd in this case).

filename - string

The name of the output file file3.

ESPS Common processing may be disabled by setting the environment variable USE_ESPS_COMMON to off. The default ESPS Common file is .espscom in the user's home directory. This may be overidden by setting the environment variable ESPSCOM to the desired path. User feedback of Common processing is determined by the environment variable ESPS_VERBOSE, with 0 causing no feedback and increasing levels causing increasingly detailed feedback. If ESPS_VERBOSE is not defined, a default value of 3 is assumed.

ESPS HEADERS

The following items are copied from the header of file1 to the header of file3:

variable.comment
variable.refer
record_freq

If the -g option is used, a generic header item scale is added to the output file header that contains the -g specified value. Max_value in file3 is not set.

The generic header item start_time is written in the output file. The value written is computed by taking the start_time value from the header of the first input file (or zero, if such a header item doesn't exist) and adding to it the relative time from the first record in the file to the first record processed.

SEE ALSO

ESPS (5-ESPS), FEA_SD (5-ESPS), record (1-ESPS), range (1-ESPS), copysd (1-ESPS)

WARNINGS

If there are not enough records in file2 - i.e., if addsd has to start over at the beginning of file2, - a warning message is printed. This warning is inhibited by the -z option, and does not apply if -t is used.

```
BUGS
     None known.
AUTHOR
     Ajaipal S. Virdy.
     Modified for ESPS 3.0 by David Burton and John Shore.
     Modified for FEA_SD by David Burton.
     Modified to support multiplication by David Burton.
```

D.2 Delay

The following is the C code written by Capt. John Columbi for the "delay" command used during this thesis work.

```
/************************************************************************\
 *
\************************************************************************/

#include <stdio.h>

/* ESPS includes */
#include "esps/esps.h"
#include "esps/feasd.h"
#include "esps/fea.h"
#include "esps/unix.h"
#define MYROUND(a) ( a - (int)a >=.5 ? ((int)a+1) : ((int)a) )


#define SYNTAX USAGE("delay [-x debug_level] [-h help] -d millisec
-s scale -j jog#samples  infile outfile" );



char           *get_cmd_line();
long            debug_level = 0;

extern          optind;
extern char    *optarg;

main(argc, argv)
```

```
short           argc;
char          **argv;
{
struct SPECTRUM *spectrum, *spec_temp;
FILE            *in_fptr = stdin, *out_fptr = stdout;
struct header   *ih, *oh;
struct feasd    *in_feasd_rec,*temp_rec;
struct feasd    *out_feasd_rec;

short           *in_data, *delay_wave;
float            record_freq = 40000.;
float          **spec_data;
int              num_in_pts, num_frames;
int              jogsample = 0;

short            c, i, num_of_files;
char            *in_fname = "stdin", *out_fname = "stdout";
char            *ProgName = "delay";
char            *Version = "1.3";
char            *Date = "Today";
char            *cmd_line;
float delay = 0.0;
float gain = 1.0;
/* Check the command line options. */
cmd_line = get_cmd_line(argc, argv); /* store copy of command line */
while ((c = getopt(argc, argv, "x:d:j:s:h:")) != EOF) {
switch (c) {

case 'x':
debug_level = atoi(optarg);
break;
case 'd':
delay = atof(optarg);
fprintf(stderr, "delay signal by %f millisecs\.n",delay);
break;
case 's':
gain = atof(optarg);
fprintf(stderr, "scale signal by %f.\n",gain);

break;
case 'j':
jogsample = atoi(optarg);
```

```
fprintf(stderr, "jog signal by %i.\n",jogsample);
break;
case 'h':
printhelp();
break;

default:
SYNTAX;
break;
}
}


/* Get the filenames */
if ((num_of_files = argc - optind) > 2 || num_of_files == 0) {
fprintf(stderr, "Improper filename specification.\n");
SYNTAX;
}
/* open the files, reading the headers of the input files. */
if (optind < argc) {
in_fname = eopen(ProgName, argv[optind++], "r", FT_FEA, FEA_SD,
 &ih, &in_fptr);
}
out_fname = eopen(ProgName, argv[optind], "w", NONE, NONE,
  (struct header **) NULL, &out_fptr);
if (debug_level) {
fprintf(stderr, "input filename = %s   ", in_fname);
fprintf(stderr, "output filename = %s   ", out_fname);
}

num_in_pts = ih->common.ndrec;

in_feasd_rec = allo_feasd_recs(ih, SHORT, num_in_pts, (char *) NULL, NO);
temp_rec = allo_feasd_recs(ih, SHORT, num_in_pts, (char *) NULL, NO);

get_feasd_recs(in_feasd_rec, 0, num_in_pts, ih, in_fptr);
in_data = (short *) in_feasd_rec->data;
delay_wave = (short *) temp_rec->data;
delay_waveform(in_data, delay_wave, num_in_pts,delay, gain,jogsample);

/*-- set up the output file. --*/
oh = copy_header(ih);
```

```
add_source_file(oh, in_fname, ih);
(void) strcpy(oh->common.prog, ProgName);
(void) strcpy(oh->common.vers, Version);
(void) strcpy(oh->common.progdate, Date);

write_header(oh, out_fptr);
in_feasd_rec->data = (short *)delay_wave;
put_feasd_recs(in_feasd_rec, 0, num_in_pts,oh, out_fptr);



fclose(in_fptr);
fclose(out_fptr);

if (debug_level)
fprintf(stderr, "Program Completed.\n");
}
/*------------------------------------------------------------*/
delay_waveform(waveform, delayed, num_pts,delay,gain,jogsample)
short          *waveform,*delayed;
int             num_pts;
float delay,gain;
{
int             pt,cpptr, samples = 0;
double mean=0;
samples = ROUND(delay*40000.0/1000.0);
if(jogsample) samples = jogsample;
if (debug_level)
fprintf(stderr, "Delaying by %d samples\n", samples);

for (cpptr=samples,pt = 0; pt < num_pts-samples; pt++,cpptr++) {
delayed[cpptr] = waveform[pt]*gain;
}
for (cpptr=0; cpptr < samples; cpptr++) {
delayed[cpptr] = waveform[0]*gain;
}

}
/*------------------------------------------------------------*/
printhelp()
{
```

```
fprintf(stderr,"\nUSAGE: delay [-x debug_level] [-h help] -d millisec
-s scale -j jog#samples  infile outfile\n" );
fprintf(stderr,"\nDelay an ESPS sampled data file\n\n");
fprintf(stderr,"Flags include\n");
fprintf(stderr,"\t -d delay : delay by so many millisecs, fills
with first sample \n\t\t (since often background signal is not
all 0's\n");
fprintf(stderr,"\t -s scale : scale the samples by this float
value\n");
fprintf(stderr,"\t -j samples : instead of delay in millisecs, delay
by jogging this number of samples\n");
}
```

D.3 Filter

Manual Pages for the ESPS Command "filter":

NAME

filter - Performs digital filtering on a sampled data file.

SYNOPSIS

filter [-P param_file] [-p range] [-r range] [-d data_type] [-f filtername] [-F filt_file] [-x debug_level] [-i up/down] [-z] in_file out_file

DESCRIPTION

The program filter takes the input sampled data file, in_file, and produces an output sampled data file, out_file, after performing a digital filtering operation on it. The output sampled data file is of type FEA_SD (5-ESPS). Filter allows the user to change the data type of the output file by using the -d option; see below for more details. The program accepts "-" for either the input file or the output file to use the standard input and standard output, respectively.

The program may implement either finite impulse response (FIR) or infinite impulse response (IIR) filters. A set of numerator coefficients and (optionally) a set of denominator coefficients are specified either in the parameter file or in the FEAFILT file. Currently, only real filters may be used; if a filter is complex, it's imaginary part is ignored. The numerator coefficients then become the $\{a_i\}$ and the denominator coefficients become the $\{b_i\}$ in the following z-domain transfer function:

$$H(z) = \frac{a_0 + a_1 z^{-1} + a_2 z^{-2} + \ldots + a_{m-1} z^{-m+1}}{b_0 + b_1 z^{-1} + b_2 z^{-2} + \ldots + b_{n-1} z^{-n+1}}$$

An FIR filter corresponds to the case where, in the equation above, $b_0 = n = 1$. An FIR filter may also be specified by choosing the order of the denominator to be zero, and entering no denominator coefficients at all.

The program uses a different initialization procedure for FIR filters than it uses for IIR filters. For FIR filtering, the first output will be computed from data samples occurring before the starting point in the input file (as defined by the parameter start , for example), if they exist. Data samples which would occur before the first sample in the input file are assumed to be zero. For IIR filtering, all inputs and outputs occurring before the starting point are assumed to be zero.

OPTIONS

The following options are supported:

-P param_file

uses the parameter file param_fil rather than the default, which is params.

-r range

Perform the filtering operation on the specified range of points. range is a character string which is interpreted in the format understood by range_switch (3-ESPSu). r and p are synonyms.

-p range
Perform the filtering operation on the specified range of points. range is a character string which is interpreted in the format understood by range_switch (3-ESPSu). -r is a synonym for -p.

-d data_type
The argument data_type is a character representing the desired output data type in out_file: b for byte, s for short, l for long, f for float, and d for double. This data type conversion is often useful when the input data type is short and the filtering operation produces sample values greater in magnitude than 215 . The output type is real or complex in agreement with the input type; for example if d is specified, the output type is DOUBLE if the input is real and DOUBLE_CPLX if the input is complex.

-f filtername
If the coefficients are being read from the parameter file, then filtername is the body of the name of the variable that contains the number of coefficients and the actual coefficients. This means that the coefficients will be found in the arrays filtername_num and filtername_den and the size of those arrays will be specified by filtername_nsiz and filtername_dsiz, respectively. The default name in this case is filter.

If the coefficients are being read from a FEAFILT file, then filtername is the number of the filter record to use. The default name in this case is 1, the first record in the file.

-F filt_file
Read the coefficients from the FEAFILT file filt_file rather than from the parameter file. In this case the header of filt_file is added to the header of the program output as a source file.

-x debug_level
: A value of zero (the default value) will cause filter to do its work silently, unless there is an error. A nonzero value will cause various parameters to be printed out during program initialization.

-i up/down
: Perform interpolation filtering such that the output sampling rate is equal to (src_sf)*(up/down). Both up and down are integers less than 10. Effectively, the program increases the sampling rate to up*(src_sf), filters this signal with the specified filter, and then downsamples the resulting signal by a factor of down.

-z
: By specifying -z, the start_time generic value is reduced by the value of the delay_samples generic header value, if it exists. This often helps time align a filtered signal with the input signal. Note that if delay_samples is not defined in the input file header, a value of 0 is assumed.

ESPS PARAMETERS

The values of parameters obtained from the parameter file are printed if the environment variable ESPS_VERBOSE is 3 or greater. The default value is 3.

The following parameters are read from the parameter file:

start - integer

The first point in the input sampled data file that is processed. The samples are assumed to be numbered starting with one so that setting start = 1 will cause processing to begin with the first sample.

nan - integer

The number of points in the sampled data file to process.

filtername_nsiz - integer

The number of numerator coefficients in the transfer function for the filter filtername.

filtername_dsiz - integer

The number of denominator coefficients in the transfer function for the filter filtername. A value of zero means that a denominator coefficient array need not be entered.

filtername_num - float array

The numerator coefficients. They are specified in order starting with a_0.

filtername_den - float array

The denominator coefficients. They are specified in order starting with b_0.

ESPS COMMON

If the input is standard input, COMMON is not read. The following items may be read from COMMON:

filename - string

This is the name of the input file. If the command line specifies only one filename, it is assumed to be the output filename and COMMON is read to get the input filename. If the input filename is specified on the command line, it must match filename in COMMON or the other items (below) are not read.

start - integer

This is the starting point in the input file.

nan - integer

This is the number of points to process.

If the output is standard output, COMMON is not written. Otherwise the following items are written to COMMON.

filename - string

This is the name of the output file.

start - integer

This is the starting point in the output file and is always equal to one.

nan - integer

This is the number of points in the output file.

ESPS Common processing may be disabled by setting the environment variable USE_ESPS_COMMON to "off". The default ESPS Common file is .espscom in the user's home directory. This may be overidden by setting the environment variable ESPSCOM to the desired path. User feedback of Common processing is determined by the environment variable ESPS_VERBOSE, with 0 causing no feedback and increasing levels causing increasingly detailed feedback. If ESPS_VERBOSE is not defined, a default value of 3 is assumed.

ESPS HEADER

The file header of out_file will contain mostly the same information as is contained in that of in_file, except where they are altered by the parameters in the parameter file. The -i option changes the sf header item. The filter coefficients will be stored in the output header as the filter zfunc.

A generic header item start_time (type DOUBLE), is added to the output file header. It contains the starting time in seconds of the first point in the output file. This start time is relative to the original sampled data file. This means that if the input file has a start_time generic in it, the output file's start_time value is computed relative to the input file's start_time. Also see the -z option.

For example, if the input file has a start_time = 1.0 seconds, the input file's sampling frequency = 8000 samples/second, and the starting point in the input file = 2000, the output file's start_time = 1.0 + 2000/8000 = 1.25 seconds.

SEE ALSO

notch_filt(1-ESPS), FEAFILT(5-ESPS), atofilt(1-ESPS), wmse_filt(1-ESPS), iir_filt(1-ESPS)

None known.

AUTHOR

Brian Sublett; ESPS 3.0 modifications by David Burton. FEA_SD modifications by David Burton; multichannel and complex FEA_SD modifications by Bill Byrne.

D.4 Mux

Manual Pages for the ESPS Command "mux":

NAME

mux - multiplex sampled-data files into a single multichannel or complex file

SYNOPSIS

mux [-{prs} range] . . . [-x debug_level] [-J] [-P param_file] input1.fsd [input2.fsd . . .] output.fsd

DESCRIPTION

The mux (''multiplex'') program combines its input sampled-data (FEA_SD) files, or equal-length portions of them, into a single multichannel output sampled-data file, possibly also combining real channels in pairs to form complex channels.

Normally the number of output channels is the total number of channels in all the input files. Each output record contains the data from the samples field in one record of each input file. This is organized as a single vector containing one sample value from each channel of each input file. Within the vector, the data from the first input file comes first, followed by the data from the second, and so on, in the order of the file names on the command line. The channels of any one input file keep the same relative ordering in the output file that they had in the input file.

When the -J option is used, the number of output channels is only half the normal number; input channels of a real data type are combined in pairs into single output channels of the corresponding complex type.

The input files must be consistent in data type and sampling frequency; the output file has the same data type (unless -J is used) and the same sampling frequency as the input files. Any fields other than samples in the input are ignored.

By default, the first output record contains data from the first record in each input file, and in general the nth output record contains data from the nth record in each input file; but a later starting point in each input file can be chosen with the -p, -r, or -s option. By default, records are processed until an input file runs out of data, but a shorter range of data can be chosen with -p, -r, or -s.

If ''-'' is written for an input file, the standard input is used. At most one input file may be standard input. Names of disk files, however, may be repeated (duplicating channels). Since different -p, -r, and -s options may apply to each instance of a repeated input file name, it is possible to align and juxtapose different portions of a single input file. If ''-'' is written for the output, the standard output is used.

OPTIONS

The following options are supported:

-p range

For this program -p and -r are synonymous. See -r for the interpretation and the format of the argument.

-r first:last

-r first:+incr

Determines the range of points (records) to be taken from an input file. In the first form, a pair of unsigned integers gives the numbers of the first and last records of the range. (Counting starts with 1 for the first record in the file.) If first is omitted, 1 is used. If last is omitted, the last record in the file is used. The second form is equivalent to the first with last = first + incr .

This option and the -p and -s options may be repeated up to a maximum total number, for all three kinds, of the number of input files. The first -p, -r, or -s option applies to the first input file, the second to the second, and so on. If there are fewer -p, -r, and -s options than input files, the last such option applies to all the remaining input files. In particular, if there is only one -p, -r, or -s option, it applies to all the input files.

If two options disagree as to the number of records to be processed, the smaller number applies. In fact mux stops processing as soon as it encounters either the end of a specified range or the actual end of file in any input file. Certain inconsistencies in these various stopping criteria will evoke warning messages; see the Diagnostics section for details.

-s start:end

-s start:+incr

Determines the range in seconds of the data to be taken from an input file. In the first form, a pair of floating-point numbers give the beginning time and ending time of the range. The second form is equivalent to the first with last = first + incr. Each sample has a time given by s + (r-1)/f, where s is the value of the generic header item "start_time", r is the record number, and f is the sampling frequency, given by the generic header item "record_freq". This time may depend on the channel number, since the "start_time" item may be a vector with a component per channel; for present purposes the value for the first channel (number 0) is used. The range selected by the

-s option consists of the records for which the time is less than end but not less than start.
This option and the -p and -r options may be repeated to supply different ranges for different input files.
See the -r option for details.

-x debug_level
If debug_level is positive, the program prints debugging messages as it progresses---the higher the number, the more messages. The default level is 0, for no debugging output.

-J Join pairs of input channels to form single complex output channels. The total number of channels in the input files must be even, and the output file has half that number of channels. The input channels are taken in the usual order and grouped in pairs to form the real and imaginary parts of the output channels. The pairing is without regard to whether two input channels come from the same input file or consecutive files. The last channel of a file, if not paired with the previous input channel, is paired with the first channel of the next input file.

The input files must all have the same real data type: DOUBLE, FLOAT, LONG, SHORT, or BYTE. (See FEA(5-ESPS) for an explanation of these type codes.) The output file has the corresponding complex data type: DOUBLE_CPLX, FLOAT_CPLX, LONG_CPLX, SHORT_CPLX, or BYTE_CPLX.

If two channels with with different time alignments are combined into one complex channel, time-alignment information may be lost. A warning message is printed in that case. See the discussion of the "start_time" generic header item in the section on ESPS Headers.

-P param_file
The name of the parameter file. The default name is ''params''.

ESPS PARAMETERS

The parameter file is not required to be present, as there are default values for all parameters. If the file exists, the following parameters may be read if they are not determined by command-line options.

start - integer array

The starting record number in each input file. The array elements are matched with input files in order. If there are more input files, the last array element applies to the unmatched file. If there are more array elements, the unmatched ones are ignored. This parameter is not read if the -p, -r, or -s option is specified. The default is all 1's, meaning the beginning of each input file.

nan - integer

The number of records to process in each input file. A value of 0 (the default) means continue processing until the end of an input file is reached. This parameter is not read if the -p, -r, or -s option is specified.

make_complex - string

A value of "YES" or "yes" means join pairs of real channels to form complex channels as if the -J option is in force. A value of "NO" or "no" means make a separate output channel for each input channel as usual. No other values are allowed. This parameter is not read if the -J option is specified. The default value is "NO".

ESPS COMMON

The ESPS Common file is not read.

If Common processing is enabled, and the output file is not standard output, the program writes the Common parameters prog, filename, start, and nan to record the program's name, the name of the output file, the starting record number of the output file (always 1), and the number of points in the output file.

ESPS Common processing may be disabled by setting the environment variable USE_ESPS_COMMON to off. The default ESPS Common file is espscom in the user's home directory.

This may be overidden by setting the environment variable ESPSCOM to the desired path. User feedback of Common processing is determined by the environment variable ESPS_VERBOSE, with 0 causing no feedback and increasing levels causing increasingly detailed feedback. If ESPS_VERBOSE is not defined, a default value of 3 is assumed.

ESPS HEADERS

The output header is a new FEA_SD file header, with appropriate items copied from the input headers.

The generic header item "record_freq", which must have the same value in all input files, is copied into the output header.

The generic header item "start_time" records the starting time for each output channel. It is a single number if all output channels have the same starting time; otherwise it is a vector with one element per channel. The starting time for a channel is its starting time in the input file plus an offset in case the data taken from the input file do not start with the first record. The offset is given by (r-1)/f where r is the starting record number in the input file and f is the sampling frequency given by the "record_freq" header item. The -J option can create complex channels whose real and imaginary parts have inconsistent starting times. When that happens, a warning message is printed, and the starting time for the real part is recorded in the "start_time" header item.

If every input file has a "max_value" header item, then the output file has a "max_value" header item containing the same information.

EXAMPLES

Multiplex data from three input files to produce an output file xxx. Processing begins with the sampled data in the first record in each input file. The output file has the same length as the shortest input file.

```
mux aaa bbb ccc xxx
```

Start at time 0.5 in each input file and process 0.5 seconds of data from each. (Suppose the sampling frequency is 8000 Hz, and the start times in the three input files are 0.0, 0.0, and 0.5. Then the starting record numbers are 4001, 4001, and 1, respectively. The start time in the output file is 0.5 for all channels.)

```
mux -s0.5:1.0 aaa bbb ccc xxx
```

Start at time 0.5 in file aaa and with the first record in the other two input files. (With the assumptions of the previous example, the starting record numbers in the three input files are 4001, 1, and 1, respectively. The start times in the output file header are 0.5 for data from files aaa and ccc and 0.0 for data from file bbb.)

```
mux -s0.5: -p1: aaa bbb ccc xxx
```

Juxtapose data from an input file with a test signal and pass the result to another program.

```
testsd - | mux aaa - - | more_processing -
```

If aaa has two channels of real data, this will convert it to a single-channel file zzz of complex data.

```
mux -J aaa zzz
```

If aaa and bbb are single-channel files of real data, this will join them into a single-channel file of complex data.

```
mux -J aaa bbb zzz
```

Multiplex a portion of a file with a later portion of the same file.

```
mux -p1001:2000 -p2501: aaa aaa xxx
```

SEE ALSO

demux(1-ESPS), copysps(1-ESPS), addgen(1-ESPS), FEA_SD(5-ESPS), FEA(5-ESPS)

DIAGNOSTICS

The program exits with an error message if any of the following occur.

The command line cannot be parsed.

Fewer than two file names are specified (one in, one out).

Fewer input file names are specified than -p, -r, and -s options.

More than one input file name is ''-''.

An input file cannot be opened or is not an ESPS sampled-data file.

The input files do not all have the same sampling frequency.

The input files do not all have the same data type.

The -J option is specified with input files of a complex data type.

The -J option is specified, and the total number of input channels is odd.

A starting record specified with a -p, -r, or -s option does not exist in all the files that the option applies to.

The program issues a warning message if the end of a range specified by a -p, -r, or -s option is not reached, and the option argument (see the Options section) ends with an explicit last, end, or +incr. (This doesn't apply to option arguments that default to end-of-file by omitting what follows the colon.) The end of the range may fail to be reached either because the end of an input file is reached first or because another -p, -r, or -s option causes an earlier stop.

The program issues a warning message if processing for the -J option joins two channels that are not properly time-aligned (so that they would require conflicting entries in the output "start_time" header item).

BUGS

The -s option is not implemented in this version of the program.

AUTHOR

Manual page by Rodney Johnson. Program by Alan Parker.

D.5 S32cplay

Manual Pages for the ESPS Command "s32cplay":

NAME

s32cplay - send sampled data (PCM) to an Ariel dsp32c S-bus/ProPort D/A converter.

SYNOPSIS

s32cplay [-r range] [-s start time] [-e end time] [-f sample rate] [-c channel] [-x debug-level] [-H] [[-i] file] [more-files]

DESCRIPTION

S32cplay sends all or a portion of one or more ESPS, SIGnal, NIST or headerless sampled data files to a S-32C/ProPort digital-to-analog converter. A subrange of data within the files may be chosen; this subrange may be specified in seconds or sample points. Dual-channel (stereo) or single-channel (monaural) data may be converted. Single-channel input data may be directed to either or both output channels.

Playback may be stopped by sending the terminal's interrupt character (normally control-C) after playback has started.

If "-" is given for a filename, then the input is taken from standard input and must be an ESPS file or a headerless file (i.e., SIGnal or NIST/Sphere files cannot be used with standard input).

OPTIONS

The following options are supported:

-r range

Select a subrange of points to be played, using the format start-end , start:end or start:+count. Either the start or the end may be omitted; the beginning or the end of the file are used if no alternative is specified.

If multiple files were specified, the same range from each file is played.

-s start time

Specify the start time in seconds. Play will continue to the end of file or the end time specified with -e. -s may not be used with -r.

-e end time

Specify the playback end time in seconds. Play will start at the beginning of file or the time specified by -s. -e may not be used with -r.

-f frequency

Specifies the sampling frequency. The closest frequency to that requested will be selected from those available and the user will be notified if the selected value differs from that requested. If -f is not specified, the sampling frequency in the header is used, else the default value for headerless files is 16kHz (assuming the standard Ariel crystal).

-c channel

Select the output channel configuration. For files with headers, the default is to play stereo if the file is stereo and to provide identical output on both channels if the file is single-channel. If the file has a header and is single channel, acceptable arguments to -c are 0 or 1 to select channel A or B respectively for output. If the file has no header, the default is to assume single-channel data and provide identical output to both channels. For headerless files, acceptable arguments to -c are 0 (output on channel A), 1 (output on channel B) or 2 (stereo data, stereo output).

-H Force s32cplay to treat the input as a headerless file. This is probably unwise to use unless the gain on your loudspeaker or earphones is way down, since a file that really does have a header, or a file composed of data types other than shorts (of the correct byte order!) will cause a terrible sound.

-i input file

Specify a file to be D/A converted. Use of -i before the file designation is optional if the filename is the last command-line component. If no input file is specified, or if "-" is specified, input is taken from stdin.

-x debug_level

Setting debug_level nonzero causes several messages to be printed as internal processing proceeds. The default is level 0, which causes no debug output.

INTERACTION WITH XWAVES

S32cplay is designed to optionally use the server mode of xwaves (1-ESPS). This is especially handy when s32cplay is used as an xwaves external play command (e.g. by setting the xwaves global play_prog). When the latter is the case, play commands initiated via xwaves' menu operations may be interrupted by pressing the left mouse button in the data view. Xwaves will send a signal 30 (SIGUSR1) to the play program. S32cplay responds to this by sending back to xwaves a command "set da_location xx", where xx is the sample that was being output when play was interrupted. This setting, in conjunction with xwaves' built-in callback procedure for handling child-process exits, causes the xwaves signal display to center itself on the sample where play was halted.

The SIGUSR1 signal to terminate s32cplay may come from any source. If it comes from sources other than xwaves, the environment variables WAVES_PORT and WAVES_HOST must be correctly defined (see espsenv (1-ESPS)), for correct functioning of the xwaves view positioning. (Of course, xwaves must actually be displaying the signal in question at the time and xwaves must have initiated the play.)

S32cplay may also be interrupted with kill -2 (SIGINT) or kill -3 (SIGQUIT). These signals are caught gracefully and s32cplay halts immediately, but no message is sent to xwaves. No message is sent if the play operation finishes without interruption.

ESPS PARAMETERS

The parameter file is not read.

ESPS COMMON

ESPS Common is not read or written.

DIAGNOSTICS

S32cplay informs the user if the input file does not exist, if inconsistent options are used, or if an unsupported sample rate is requested. Also see WARNINGS below.

If the starting point requested is greater than the last point in the file, then a message is printed. If the ending point requested is greater than the last point in the file, it is reset to the last point and processing continues.

WARNINGS

S32cplay supports only the "native" Ariel ProPort sampling rates. These are tabulated in the Ariel User's Manual. Note that optional crystals are available from Ariel that will provide a different selection of frequencies. Should the crystal(s) be changed, it will be necessary to edit ESPS_BASE/s32cbin/s32c0.srtable to reflect the new selection. If you play a file that is sampled at an unsupported rate, s32cplay plays the data at the closest supported rate and issues a warning.

S32cplay provides stereo D/A conversion at rates up to at least 48kHz when playing from local disk. Sampling from network disks is often feasible as well. The maximum rate over the network is unpredictable in general, but we routinely achieve 16kHz stereo at Entropic Research Laboratory. Of course rate limitations due to network speed will be less severe for single-channel playing. Obviously, processes supplying input to s32cplay on a pipe must be able to keep up with the average aggregate sampling frequency.

Note that s32cplay requires that the dsp32c program called "ar_atod" be located in ESPS_BASE/s32cbin. If it is located anywhere else, the environment variable ARIELS32C_BIN_PATH must be set accordingly.

FILES

ESPS_BASE/s32cbin/ar_atod

the dsp32c program that runs on the Ariel S-32C card.

ESPS_BASE/s32cbin/s32c0.srtable
the table of available sampling rates (dependent on crystal frequencies).

BUGS

If readable header IS present, but -H is specified, the header is treated like sampled data -- usually resulting in very unpleasant sounds.

If a s32cplay is in progress when another S32C card operation is requested on the same machine, unpredictable results will occur.

There is currently no reliable notification in the event of loss of realtime.

EXPECTED CHANGES

Implement a locking mechanism to prevent collision of multiple simultaneous attempts to use the Ariel board.

Implement a robust check for loss of real-time operation.

SEE ALSO

SD (5-ESPS), testsd (1-ESPS), copysd (1-ESPS), s32crecord (1-ESPS), sfconvert (1-ESPS), s32csgram (1-ESPS), send_xwaves2 (3-ESPS)

AUTHORS

David Talkin at Entropic Research Laboratory and Michael McCandless at MIT Laboratory of Computer Science.

D.6 S32crecord

Manual Pages for the ESPS Command "s32crecord":

NAME

s32crecord - mono or stereo record to disk or pipe for Ariel dsp32c S-bus/ProPort A/D converter.

SYNOPSIS

s32crecord [-s duration] [-f sample rate] [-c channel] [-S] [-W xwaves display args.] [-P] [-p prompt string] [-x debug-level] [-H] [[-o] file]

DESCRIPTION

S32crecord provides mono or stereo sampling at selected rates up to at least 48kHz when recording to local disk. Direct recording onto network disks is often feasible as well. Output files have ESPS FEA_SD headers, or, optionally, no headers. Output may optionally be directed to stdout. S32crecord has special adaptations that permit tight coupling with xwaves (see INTERACTION WITH XWAVES below).

When no output file is specified, s32crecord will, by default, write a FEA_SD file to standard output. When stereo recording is selected (see the -c option) channels 0 and 1 alternate in the file, with channel 0 being first. Note that processes consuming the output of s32crecord on a pipe must be able to keep up with the average aggregate sampling frequency. Options are available to control the sampling rate, recording duration, prompting, header suppression, and immediate display by xwaves.

OPTIONS

The following options are supported:

-s duration[10]

Specifies the maximum duration of the recording session in seconds. Recording may be interrupted before this time expires with SIGINT, SIGQUIT or SIGUSR1 (see below). The default duration is 10 seconds. The upper limit is set only by disk space.

-f frequency[16000]

Specifies the sampling frequency. The closest frequency to that requested will be selected from those available and the user will be notified if the selected value differs from that requested. If -f is not specified, the default sampling frequency of 16kHz is used (assuming the standard Ariel crystal).

-c channel[1]
By default s32crecord samples and stores single-channel data from channel B. Alternate-channel recording is selected by setting channel to 0 for channel A, to 1 for channel B, or to 2 for two-channel (stereo) recording.

-S Enable the xwaves (1-ESPS) "make" command via send_xwaves2 (3-ESPS) when the requested recording time has elapsed or when recording is interrupted. This permits immediate examination of the recorded passage using xwaves. See INTERACTION WITH XWAVES below.

-W The argument to this option will be appended to the send_xwaves "make" command to permit display customization (e.g. via window location and size specifications). See INTERACTION WITH XWAVES below.

-P Enable a prompt message when A/D has actually commenced. The default message is a "bell ring" and the text "Start recording now ...." This prompt may be changed with the -p option.

-p prompt string
Prompt string will be used as the alert that recording is commencing. Specifying -p forces -P.

-H Suppresses header creation. A "bare" sample stream will result. The default is to produce an ESPS FEA_SD file with one or two channels depending on the -c option.

-x debug_level
Setting debug_level nonzero causes several messages to be printed as internal processing proceeds. The default is level 0, which causes no debug output.

-o output
Specifies a file for output. Use of -o before the file designation is optional if the filename is the last

command-line component. If no output file is specified or if "-" is specified, output will go to stdout.

INTERACTION WITH XWAVES

S32crecord is designed to optionally use the server mode of xwaves (1-ESPS) for display of its output file on completion of the record operation. This is implemented using send_xwaves2 (3u-ESPS). The following conditions must be met for this feature to work. (1) Xwaves must be running in the server mode. (2) The environment variables WAVES_PORT and WAVES_HOST must be correctly defined (see espsenv (1-ESPS)). (3) The record operation must be interrupted with a SIGUSR1 signal (e.g. via "kill -30 pid," where pid is the process ID of the s32crecord process), or if s32crecord is not thus interrupted, the -S flag must have been set. (4) Output must be to a file.
An example mbuttons (1-ESPS) script to implement a primitive record control panel follows:

```
"RECORD Chan0"  exec s32crecord -c0 -P -s60 -S -W"name $$
                loc_y 150" \ xx$$& echo $! > foo
"RECORD Chan1"  exec s32crecord -c1 -P -s60 -S -W"name $$
                loc_y 150" \ xx$$& echo $! > foo
"RECORD Stereo" exec s32crecord -P -s60 -S -W"name $$
                loc_y 150" \ xx$$& echo $! > foo
"STOP"          kill -30 `cat foo`
"ERASE"         f=`cat foo` ; k=`echo $f 1 - p q | dc` ; \
        kill -2 $f ; rm -f xx$k ; send_xwaves kill name $k
```

Note how the -W option is used to name the display ensemble and to fix the vertical location of the waveform at the same place on consecutive invocations. In general, the -W option can be used to augment the display generation as described under the "make" command in the xwaves manual. Note that the "STOP" function is implemented with a "kill -30" (SIGUSR1). This causes s32crecord to send the "make" command to xwaves. If either kill -2 (SIGINT) or kill -3 (SIGQUIT) is sent to s32crecord, it will terminate gracefully, but will not send any messages to xwaves. The -S option causes the xwaves display operation to occur even in the non-interrupted case (i.e. after 60 sec of recording). The above script is not robust, but may serve as a useful starting point for more serious attempts.

ESPS PARAMETERS

The parameter file is not read.

ESPS COMMON

ESPS Common is not read or written.

WARNINGS

Note that s32crecord requires that the dsp32c program called "ar_atod" be located in ESPS_BASE/s32cbin. If it is located anywhere else, the environment variable ARIELS32C_BIN_PATH must be set accordingly.

When output is to a file, the ESPS header, if it is present, will correctly reflect the absolute maximum sample value encountered during recording and the number of samples recorded. If output is to a pipe, these values are not recorded in the header.

If the output of s32crecord goes to the terminal, bad things will happen to the terminal configuration and you may not be able to regain control of the terminal. In this case, kill the window or kill the record process remotely from another window.

S32crecord supports only the "native" Ariel ProPort sampling rates. These are tabulated in the Ariel User's Manual. Note that optional crystals are available from Ariel that will provide a different selection of frequencies. Should the crystal(s) be changed, it will be necessary to edit ESPS_BASE/s32cbin/s32c0.srtable to reflect the new selection. If an unsupported rate is requested, s32crecord records at the closest supported rate and issues a warning.

The maximum rate over the network is unpredictable in general, but we routinely achieve 16kHz stereo at Entropic Research Laboratory. Of course rate limitations due to network speed will be less severe for single-channel recording.

FILES

ESPS_BASE/s32cbin/ar_atod
the dsp32c program that runs on the Ariel S-32C card.

ESPS_BASE/s32cbin/s32c0.srtable
the table of available sampling rates (dependent on crystal frequencies).

BUGS

If a s32cplay or another s32crecord operation is started while one is in progress, unpredictable results will occur.

There is currently no reliable notification in the event of loss of realtime.

EXPECTED CHANGES

Implement a locking mechanism to prevent collision of multiple simultaneous attempts to use the Ariel board.

Implement a robust check for loss of real-time operation.

SEE ALSO

SD (5-ESPS), testsd (1-ESPS), copysd (1-ESPS), s32cplay (1-ESPS), sfconvert (1-ESPS), s32csgram (1-ESPS)

AUTHORS

David Talkin at Entropic Research Laboratory and Michael McCandless at MIT Laboratory of Computer Science.

Appendix E. Plots of Sample HRTF Filters

The following filters are examples of those used in each simulation completed in this thesis. The angles are given by the orientation shown in figure 1 and figure 2.

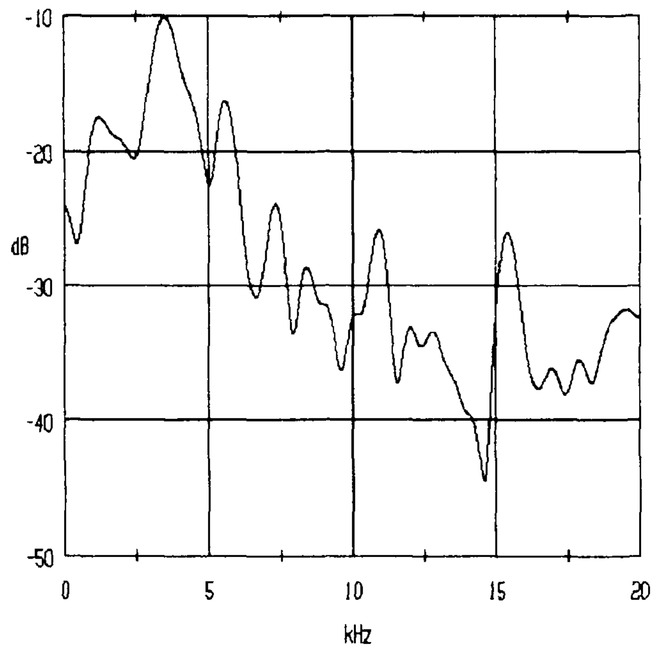

Figure 10. Filter Designed for the Left Ear at Approximately 180 degrees Azimuth

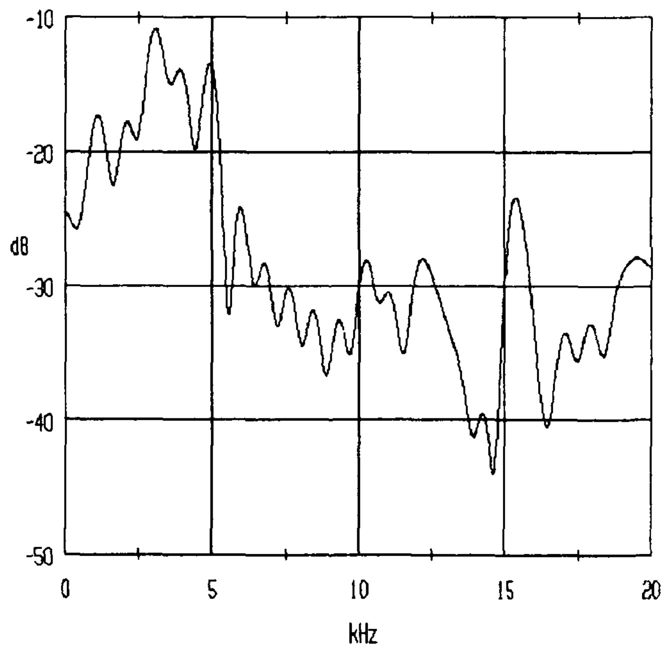

Figure 11. Filter Designed for the Right Ear at Approximately 180 degrees Azimuth

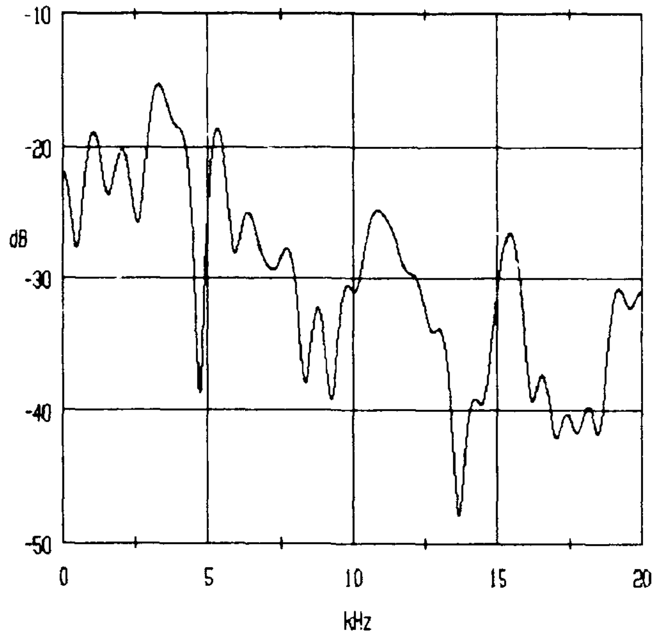

Figure 12. Filter Designed for the Left Ear at Approximately 278 degrees Azimuth

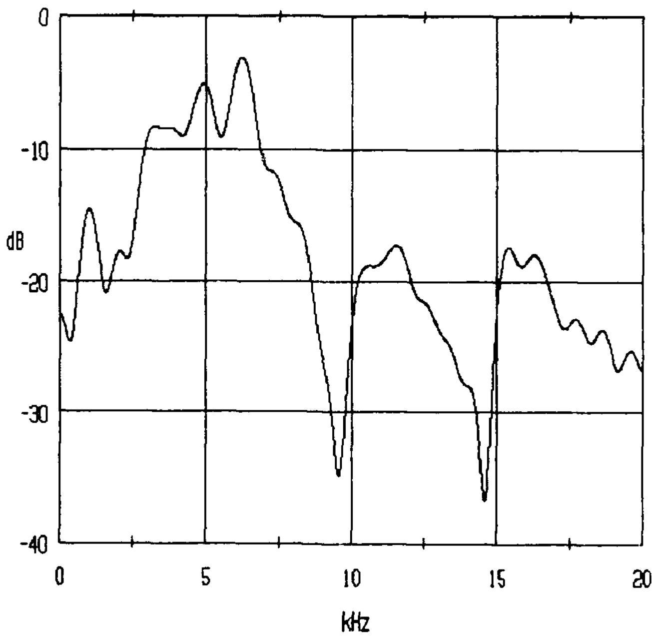

Figure 13. Filter Designed for the Right Ear at Approximately 278 degrees Azimuth

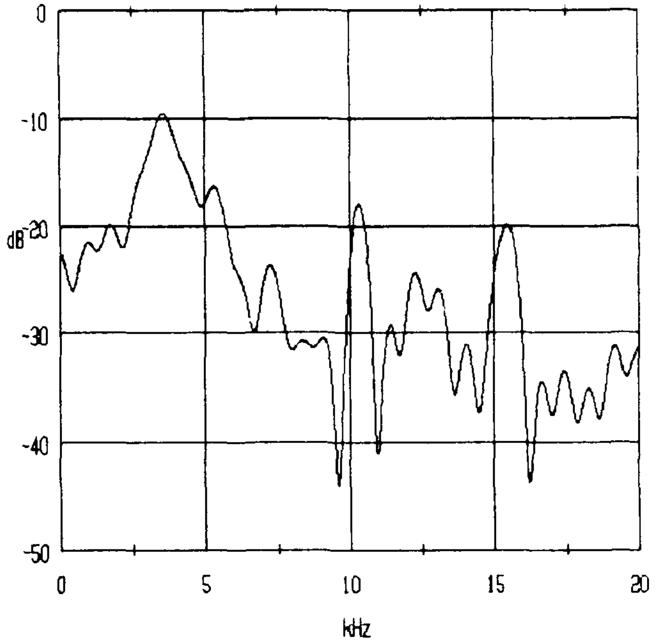

Figure 14. Filter Designed for the Left Ear at Approximately 350 degrees Azimuth

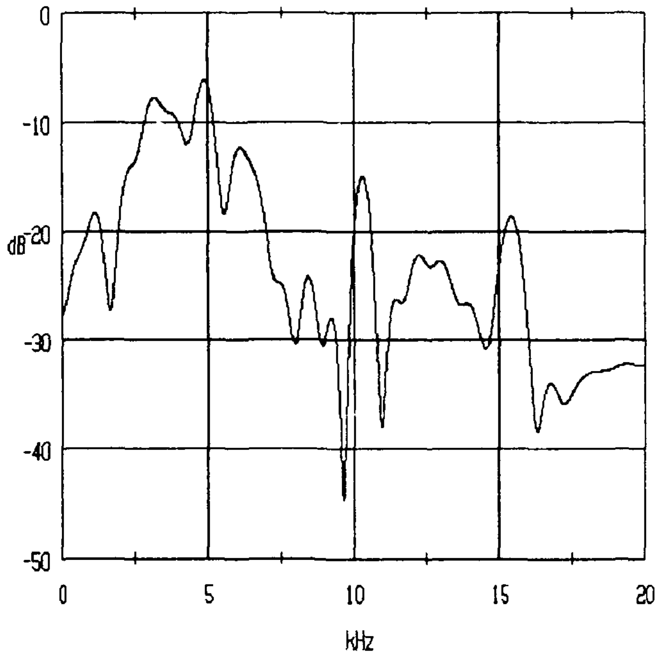

Figure 15. Filter Designed for the Right Ear at Approximately 350 degrees Azimuth

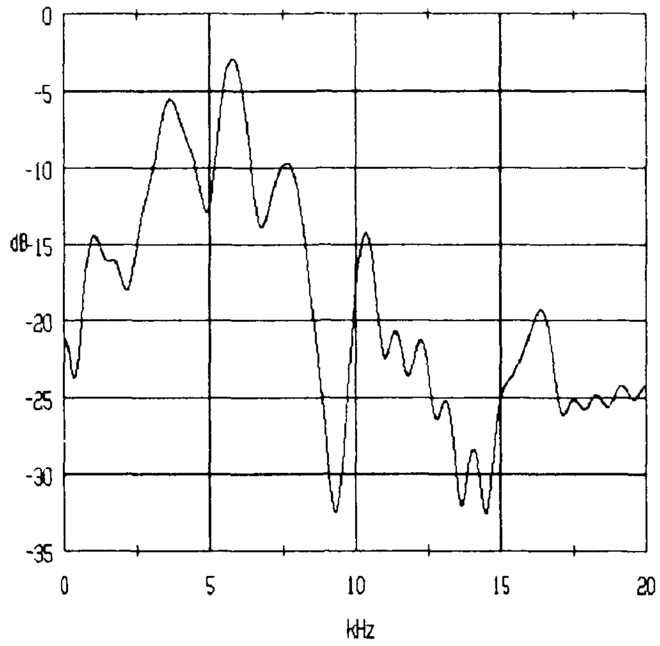

Figure 16. Filter Designed for the Left Ear at Approximately 82 degrees Azimuth

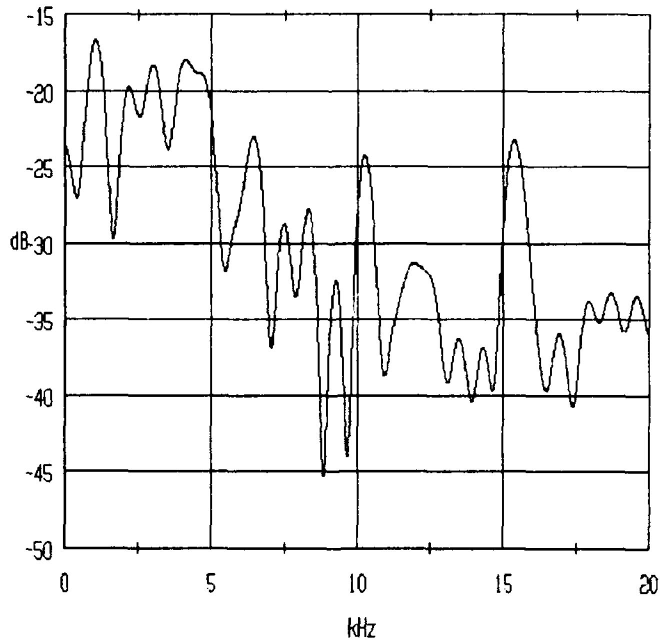

Figure 17. Filter Designed for the Right Ear at Approximately 82 degrees Azimuth

Bibliography

1. Abbagnaro, Louis A., et al. "Measurements of Diffraction and Interaural Delay of a Progressive Sound Wave Caused by the Human Head," *Journal of the Acoustical Society of America*, *58*:693–700 (September 1975).
2. Allen, Jont B. and David A. Berkley. "Image Method for Efficiently Simulating Small-Room Acoustics," *Journal of the Acoustical Society of America*, *65*:943–950 (April 1979).
3. Blauert, Jens. *Spatial Hearing*. MIT Press, 1983.
4. Blauert, Jens. "An Introduction to Binaural Technology." *Conference on Binaural and Spatial Hearing*. 1993.
5. Borish, Jeffrey. "Extension of the Image Model to Arbitrary Polyhedra," *Journal of the Acoustical Society of America*, *85*:787–796 (February 1984).
6. Clifton, Rachel. "Context and Learning Affect Echo Perception." *Conference on Binaural and Spatial Hearing*. 1993.
7. Colombi, John. "Dichotic Listening," *Unpublished* (1993).
8. Durlach, Nathaniel and Barbara Shinn-Cunningham. "Auditory Displays and Localization." *Conference on Binaural and Spatial Hearing*. 1993.
9. Ericson, Marc (Electronics Engineer, AAMRL). "Personal Interview." *Unpublished*. August 1993.
10. Genuit, Klaus and Mahlon Burkhard. "Artificial Head Measurement Systems for Subjective Evaluation of Sound Quality," *Journal of Sound and Vibration*, 18–23 (March 1992).
11. Gierlich, H.W. "The Application of Binaural Technology," *Applied Acoustics*, *36*:219–243 (March 1992).
12. Hafter, Ervin. "Binaural Precedence and Other Onset Effects." *Conference on Binaural and Spatial Hearing*. 1993.
13. Hartmann, William Morris. "Listening in a Room and the Precedence Effect." *Conference on Binaural and Spatial Hearing*. 1993.
14. Knudsen, Vern O. and Cyril M. Harris. *Acoustical Designing in Architecture*. American Institute of Physics, 1978.
15. Konishi, Masakazu. "Listening With Two Ears," *Scientific American*, 66–73 (April 1993).
16. Krokstad, A., et al. "Calculating the Acoustical Room Response by the use of the Ray Tracing Technique," *Journal of Sound and Vibration*, *8*:118–125 (1968).

17. Kuhn, George F. "Model for the Interaural Time Differences in the Azimuthal Plane," *Journal of the Acoustical Society of America*, *62*:157–167 (July 1977).

18. Lehnert, Hilmar and Jens Blauert. "Principles of Binaural Room Simulation," *Applied Acoustics*, *36*:259–291 (March 1992).

19. McKinley, Richard L. *Concept and Design of an Auditory Localization Cue Synthesizer*. MS thesis, Air Force Institute of Technology (AU), 1988.

20. McKinley, Richard L. "Flight Demonstration of a 3-D Audio Display." *Conference on Binaural and Spatial Hearing*. 1993.

21. Middlebrooks, John. "A Computational Model of Narrowband Sound Localization." *Conference on Binaural and Spatial Hearing*. 1993.

22. Middlebrooks, John C. "Directional Sensitivity of Sound-Pressure Levels in the Human Ear Canal," *Journal of the Acoustical Society of America*, *86*:89–108 (July 1989).

23. Moller, Henrik. "Fundamentals of Binaural Technology," *Applied Acoustics*, *36*:171–218 (March 1992).

24. Morse, Philip M. *Vibration and Sound*. McGraw-Hill Book Company, Inc., 1948.

25. Morse, Philip M. and K. Uno Ingard. *Theoretical Acoustics*. McGraw-Hill Book Company, Inc., 1968.

26. Ondet, A. M. and J. L. Barby. "Modeling of Sound Propagation in Fitted Workshops Using Ray-tracing," *Journal of the Acoustical Society of America*, *85*:787–796 (February 1989).

27. Pierce, Allan D. *Acoustics: An Introduction to Its Physical Principles and Applications*. McGraw-Hill Book Company, Inc., 1981.

28. Rayleigh, John William Strutt (Third Baron). *Theory of Sound, I*. Dover Publications, 1907.

29. Rayleigh, John William Strutt (Third Baron). *Theory of Sound, II*. Dover Publications, 1907.

30. Richarson, E.G. *Sound*. Edward Arnold and Co., 1927.

31. Sayers, Bruce and Colin E. Cherry. "Mechanism of Binaural Fusion in the Hearing of Speech," *Journal of the Acoustical Society of America*, *24* (1976).

32. Scarborough, Captain Eric L. *Enhancement of Audio Localization Cue Synthesis by Adding Environmental and Visual Cues*. MS thesis, Air Force Institute of Technology (AU), 1992.

33. Shaw, Edgar A.G. "Acoustical Characteristics of the Human External Ear." *Conference on Binaural and Spatial Hearing*. 1993.

34. Stephens, R. W. B. and A. E. Bate. *Acoustics and Vibrational Physics*. Edward Arnold Publishers, Ltd., 1950.

35. Stewart, George Walter. *Introductory Acoustics*. D. Van Nostrand Company, Inc., 1933.

36. Sytronics, Inc. *Virtual Acoustics for Pilot Displays*. Technical Report, Sytronics, Inc., September 1992.

37. Vorlander, Michael. "Simulation of the Transient and Steady-State Sound Propagation in Rooms Using a New Combined Ray-tracing/Image-source Algorithm," *Journal of the Acoustical Society of America*, *86*:172–178 (July 1989).

38. Wenzel, Elizabeth. "Research on Virtual Acoustic Environments at NASA." *Conference on Binaural and Spatial Hearing*. 1993.

39. Wenzel, Elizabeth M. "Localization in Virtual Acoustic Displays," *Presence*, *1* (1992).

40. Wightman, Frederic L. and Doris J. Kistler. "Headphone Simulation of Free-Field Listening.II:Psychophysical Validation," *Journal of the Acoustical Society of America*, *85* (1989).

41. Wightman, Frederic L. and Doris J. Kistler. "Headphone Simulation of Free-Field Listening.I:Stimulus Synthesis," *Journal of the Acoustical Society of America*, *85* (1989).

42. Wightman, Frederic L. and Doris J. Kistler. "Factors Affecting Relative Importance of Sound Localization Cues." *Conference on Binaural and Spatial Hearing*. 1993.

43. Winstanley, J.W. *Textbook on Sound*. Longmans, Green and Co., 1952.

Vita

Brian A. Smith was born on 15 November 1967 in Beech Grove, Indiana. In 1986, he graduated from Greenfield-Central High School and accepted an appointment to the United States Air Force Academy. In 1990, he was commissioned out of the Academy and received a Bachelor of Science in Chemistry. After graduation, Lt. Smith attended Undergraduate Pilot Training (UPT) at Reese AFB, TX. He graduated from UPT in November of 1991 and received a "banked" assignment to Wright-Patterson AFB. Stationed at WPAFB to earn a masters degree, Lt. Smith also took part in several experiments at Armstrong Laboratories. He was a test subject in high G experiments as well as for motion sickness experiments. Following the completion of the AFIT masters program, Lt. Smith returned to flying status.

Permanent address: 1901 Roop Ct
Kettering, OH 45420

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*

2. REPORT DATE
December 1993

3. REPORT TYPE AND DATES COVERED
Master's Thesis

4. TITLE AND SUBTITLE
BINAURAL ROOM SIMULATION

5. FUNDING NUMBERS

6. AUTHOR(S)
Brian A. Smith

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Air Force Institute of Technology, WPAFB OH 45433-6583

8. PERFORMING ORGANIZATION REPORT NUMBER
AFIT/GAM/ENC/93D-1

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
AL/CFBA
Attn: Mr. Mark A. Ericson
2610 Seventh St.
Bldg. 441
WPAFB, OH 45433-7901

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Distribution Unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

Abstract

Research in binaural and spatial hearing is of particular interest to the Air Force. Applications in cockpit communication, target recognition, and aircraft navigation are being explored. This thesis examines human auditory localization cues and develops a mathematical model for the transfer function of a sound signal traveling from an isotropic point source through a rectangular room to both ears of a listener. Using this model as a guide, non-head coupled binaural sound signals are generated in a binaural room simulation. Reflection and attenuation cues included in the computer generated signals are varied in order to determine which cues enhance the listener's degree of extracranialization. Results of this research indicate that the addition of three or more attenuated reflections into a non-head coupled binaural signal provide the listener with a binaural sound that is localized extracranially.

14. SUBJECT TERMS
Binaural Sound, Binaural Room Simulation, Auditory Localization Cues, Reflections, Attenuation, Extracranialized Sound, Head Coupling

15. NUMBER OF PAGES
127

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UL