Development of a Virtual Acoustic Showroom for Simulating Listening Environments and Audio Speakers

Christopher M. Collins

Thesis submitted to the Faculty of the
Virginia Polytechnic Institute and State University
in partial fulfillment of the requirements for the degree of

Masters of Science
in
Mechanical Engineering

Dr. Marty E. Johnson, Chair
Dr. James Carneal
Dr. Doug Bowman

May 24, 2004
Blacksburg, Virginia

Keywords: Virtual Acoustics, Image Source, Inverse Filter

Development of a Virtual Acoustic Showroom for Simulating Listening Environments and Audio Speakers

Christopher M. Collins

(ABSTRACT)

Virtual acoustic techniques can be used to create virtual listening environments for multiple purposes. Using multi-speaker reproduction, a physical environment can take on the acoustical appearance of another environment. Implementation of this environment auralization could change the way customers evaluate speakers in a retail store.

The objective of this research is to develop a virtual acoustic showroom using a multi-speaker system. The two main components to the virtual acoustic showroom are simulating living environments using the image source method, and simulating speaker responses using inverse filtering. The image source method is used to simulate realistic living environments by filtering the environment impulse response by frequency-dependant absorption coefficients of typical building materials. Psychoacoustic tests show that listeners can match virtual acoustic cues with appropriate virtual visual cues. Inverse filtering is used to "replace" the frequency response function of one speaker with another, allowing a single set of speakers to represent any number of other speakers. Psychoacoustic tests show that listeners, could not distinguish the difference between the original speaker and the reference speaker that was mimicking the original. The two components of this system are shown to be accurate both empirically and psychologically.

Acknowledgments

I would like to thank my advisor, Marty Johnson, who has given me this opportunity and has taught me so much about acoustics and virtual acoustics. Thank you for pushing me to always dig deeper and to extract all the possibilities as to why something was happening. Much of the credit for this work goes to you.

I would also like to thank Jamie Carneal. I appreciate the many hours you have willingly given to me, offering insight on test results, and helping me solve problems as they came up. Thank you also, for agreeing to serve on my committee.

I would also like to thank Doug Bowman. Your help in providing insight into virtual reality testing with the head mounted display has been invaluable. Thank you for your willingness to serve on my committee.

There were several others who assisted in the work, and to whom thanks are due: Steve Booth for assisting in the construction of the virtual showroom prototype, Jan Troge for programming the virtual acoustic simulation software, and John Lucas for helping me create and implement the virtual 3-D rooms for the head mounted display. Thanks, also, to Lynda, Dawn, and the rest of VAL for all of your support, encouragement, and friendship.

In addition, I would like to thank my mom and dad. Your never-ending prayers and support have helped keep me going throughout this entire process.

To my wife, Adrianne, I dedicate this thesis. Without your love, your support, and your sacrifices, none of this would have been possible. Thank you.

And to God, thank you for always being my guide and my support.

Contents

List of Figures

List of Tables

Chapter 1

Introduction to Virtual Acoustics

The objective of this thesis is to communicate the author's research in virtual acoustics, specifically in the development of a virtual acoustic showroom. In order to understand the motivation and the scope of this research, as well as the current literature on this topic, one must have an understanding of virtual acoustics, which will be described first.

1.1 Introduction to Virtual Acoustics

Virtual acoustics and auralization are two terms used interchangeably for describing the creation and reproduction of virtual sound fields. Kleiner et. al [1] gave the definition for auralization as: "Auralization is the process of rendering audible, by physical or mathematical modeling, the sound field of a source in a space, in such a way as to simulate the binaural listening experience at a given position in the modeled space." There are a number of different aspects that are considered when developing virtual acoustic simulations. Modeling of the source, the room, and the listener are required before reproducing the sound fields.

The source is often modeled as an omni-directional point source, because it is the simplest approach. There are a number of ways to model the room, including; ray-tracing, the image source method, and hybrids between them. The listener is generally modeled by a set of head-related transfer functions, which model the diffraction and filtering of the head, shoulders and pinnae of the listener [1]. After completing the modeling, the reproduction of the virtual sound fields can be implemented.

As mentioned in the definition of auralization, binaural simulation is one method to reproduce virtual sound fields. This method acts to reproduce the sound pressure levels at the listener's eardrums that would have been created by the real source. To achieve the appropriate levels at the listener's eardrums, one of two methods are generally applied. One is the use of two speakers, located close to each other, making use of crosstalk cancellation filters. This requires that the listener stay positioned in a specific location and the auralization is lost by moving outside the specified location. The other method is to use headphones. The listener is able to move around without loosing the auralization, but has to wear headphones, which wouldn't be appropriate in some situations.

Another method of reproducing virtual sound fields is to use multiple loudspeakers. Using a number of loudspeakers, the sound pressure levels surrounding the listener are reproduced for the desired auralization. This method is the one focused on in this thesis for two main reasons; one, it does not require the listener to remain stationary for the auralization, and two, the listener is not required to wear headphones.

1.2 Motivation for Work

Currently, customers are able to listen to a large number of speakers on display in an audio equipment retail store. This requires a substantial amount of retail space and a significant number of speakers. The purpose of having such a display is to allow the customer to hear what the speaker sounds like prior to making a purchase. But with having a large display, there are major drawbacks for both the customer and the vendor. Drawbacks for the vendor include large retail space and a large amount of inventory. For the customer, the drawbacks include a limited selection of speakers to listen to and the lack of a consistent room interaction; the customer makes a purchase decision on how the speaker sounds in the store, not on what it will sound like in their house.

It may be possible to use virtual acoustic techniques to develop a virtual acoustic showroom that simulates both the listening environment and the audio speakers. Using an appropriate room acoustic modeling technique, acoustic environments can be simulated based on the construction materials, the size, and the shape of typical living environments. The customer would then be able to hear what the simulated acoustic environments would sound like, by reproducing those environments through the use of auralization techniques. The customer would select, from a list, a virtual acoustic environment that closely matches the construction, geometry, and size of their room, thereby presenting them with an accurate representation of what a set of speakers will sound like in their room.

If a set of every speakers carried by the retail store were physically present in the showroom, at a minimum, the problem of having a large amount of display inventory has not been addressed. A more severe problem arises in that the placement for each set of speakers is not consistent throughout the showroom. In other words, one set of speakers may be placed in the corners of the showroom, while another set may be placed in the center of the walls. This makes for comparing the two sets of speakers nearly impossible for two critical reasons. First, the modal response of the showroom will interact and modify a speaker's response, depending on the location of the speaker within the showroom. This interaction is consistent between different speakers, provided the locations of the speakers is the same. Secondly, Olive et. al [2] reported that a speakers location, referenced from the room and the listener's position, affects the listener preference ratings among speakers. In fact, the subjective effects of speaker placements are much larger than the differences between types of speakers. Therefore, if the speakers are not in the same location in the showroom, artificial differences between them may solicit the purchase of one speaker over another.

To overcome the issues of speaker placement, as just described, an inverse filtering method could be implemented in such a way as to allow the response of one speaker to mimic another. Using one set of speakers (a reference set) in the showroom, to mimic all the other speakers that are carried by the retail store, the problems of speaker placement would be eliminated. Another benefit of using one set of speakers is the reduction, and near elimination, of display inventory held by the store.

Combining the auralization of virtual acoustic environments with speaker mimicking into one showroom facility will allow customers to customize the listening parameters to represent their living environment and the speakers of interest before a purchase is ever made. They could potentially have an infinite number of speakers to choose from and an infinite number of environments to listen to them in. Benefits of this showroom facility, for the retail stores, would include a substantial reduction of display inventory, a more efficient use of retail space, and a new-technology based marketing tool.

1.3 Review of Literature

The research described by this thesis combines both the image source method for simulating virtual acoustic environments and inverse filtering for simulating speaker responses for developing a virtual acoustic showroom system. The image source method has been widely researched for its ability to accurately represent acoustic properties of a room in predictive modeling. However, there has not been as much research of its implementation into a virtual acoustic reality simulation. Simulation schemes of virtual acoustic environments can be divided into binaural (headphones), crosstalk canceled binaural (loudspeaker), and multichannel reproduction [1].

The simulation of room acoustics has found its place in determining the design of acoustical environments. Along with the calculation of different parameters given by the simulation, the auralization of the acoustical environment is of great importance. Several researchers have taken room acoustical models calculated by computers and have applied different techniques in order to auralize the generated models. Kleiner et. al [1] provides an overview into auralization methods used by researchers. One motivation for this auralization is to be able to accurately recreate, or simulate, the aural impression of the acoustic characteristics in a planned hall, prior to construction.

Using the auralization of a given acoustical environment within another physical environment could allow the physical environment to take on the acoustical characteristics contained within the auralization. Having the ability to change a rooms acoustical properties to appropriately match the event taking place at that moment, and then change it for a different event later, would allow multipurpose halls to cater to each event by providing optimal acoustic conditions. This then lends itself to use in virtual reality systems, where audiovisual stimulus is provided in an interactive capacity. The virtual reality community has placed an emphasis on the visual aspect of virtual environments, with audio being more of a side note. Incorporating the virtual audio, correlating to the visual scene, would greatly enhance the overall impression within the virtual environments.

A number of modeling techniques exist for simulating acoustic environments, including, waveguide mesh [3, 4], ray or beam tracing [5, 6], image source method [7, 8, 9, 10], and finite-element and boundary-element modeling [1]. For the auralization of acoustic environments, the image source method is generally used over the other methods. The other methods require a substantial amount of computational power and are prohibitive, at present, in real-time implementations. To increase the efficiency of calculating the acoustic environments for the auralization process, researchers often use a hybrid of the image source method where only the early reflections are calculated using the image source method, then recursive digital filters are used for the late (diffuse) reverberation [11, 12].

A number of room acoustic software packages use a hybrid of the image source method for calculating acoustic properties of environments. Rindel and Christensen [13] describe the use of the ODEON computer model for the purpose of creating and verifying auralizations of different acoustical environments. The first test in the verification process was to compare the calculation results from the simulation with the measurement of the simulated impulse response using a monaural auralization filter. The second test consisted of subjective comparisons by listeners between the computer simulation, a binaural recording in the real room by a dummy head, and a convolution of the measured binaural impulse response with an anechoic recording.

Also using BRIR or Head Related Transfer Functions (HRTF) for the implementation of auralized acoustic simulations is Savioja et. al [11, 14]. In both cases headphones were used for the auralization of the simulated acoustic environments. The listeners were also able to move throughout the virtual environment that accompanied the auralization. In [11] the authors create an interactive virtual acoustic environment, which allows listeners to interact with the sound source. The validation of the auralization for this environment was limited to informal listening. The model was then modified to reflect the listeners' opinions until they were satisfied. Major drawbacks exist for using binaural auralization for both headphone and crosstalk canceled loudspeakers. With the crosstalk canceled loudspeakers, the listener is confined to a location designated by the crosstalk filters. The auralization effect is lost when the listener moves from that position. For headphones, the position is not confined to a particular location, however the listener is "tethered" to the system.

Multichannel reproduction has also been implemented in auralizing virtual acoustic environments. Pulkki and Lokki [15] use a method termed Vector Based Amplitude Panning (VBAP), where a total of sixteen speakers were used for the auralization process. As with the previous research discussed, the authors used a hybrid image source method for simulating acoustic environments. The VBAP requires less computational capacity than filtering using HRTFs and allows some movement within the area of speakers without affecting the experience of the auralization. The authors report that informal tests found that the system produced a natural sounding virtual auditory environment.

Another multichannel system was implemented by Gardner [12], which used only four loudspeakers. This system was designed to allow the acoustical properties of a room to be modified for optimal acoustic conditions of various events in that room. Therefore, the auralization of different acoustical environments was for stationary sources, where as the other simulations (VBAP and BRIR) allow for non-stationary sources. Verification of this system consisted of informal listening tests. Gardner reported that there were noticeable changes in auralization between different simulated environments, and that it was enjoyable to listen to the system for long periods of time without fatigue. Gardner then expanded the four speaker system into a six speaker system [16]. The results of the expanded system were reported as having improved spatial coverage of the listening environment over the previous system. The validation of this system was, again, informal listening.

As previously mentioned, a portion of this thesis uses an inverse filtering method for speaker compensation. Mäkivirta et. al [17] also use inverse filtering for speaker compensation. They use the inverse filter to reduce the mode decay rate of a loudspeaker-room system at low frequencies. The results show that by modifying the input signal into the loudspeaker with an inverse filter, the speaker-room mode decay below $200Hz$ is reduced considerably.

While much research has already been conducted on the use of the image source method, or a hybrid, for simulating acoustic environments for auralization, most auralization methods revolve around binaural implementation. Some research has also been conducted on auralization implementation using multiple loudspeakers. For all of this research, very little validation of the systems have been analyzed. The binaural implementations have been validated more so than the multispeaker implementations. Also, the use of inverse filters has been investigated for modifying the low frequency response of a loudspeaker, but using the inverse technique to modify the entire frequency response of a loudspeaker has not.

1.4 Scope of Work

As previously mentioned, the goal of this research is to develop a virtual acoustic showroom where both acoustic environments and speakers can be simulated. Only a minimal number of speakers are to be used in order to achieve accurate auralizations. To ensure proper auralization, comparisons will be made between the Sabine predicted reverberation time and both the calculated environment filters (pre-auralization) and measurements of the environment filters (auralization). Psychological testing will also be conducted to determine the human perception of the virtual acoustic environments. To simulate speakers, inverse filtering will be used so that the frequency response function of one speaker can be mimicked by another.

Psychological testing will be conducted to determine if listeners can distinguish a difference between the real speaker and the speaker mimicking it.

It is also desired to expand the virtual acoustic showroom beyond bookshelf speaker systems, so as to include 5.1 system and 6.1 system DVD surround sound formats and a vehicle simulations based on classes of vehicles. Increasing the number of speakers for both additional systems will be addressed. All together, these three virtual showrooms would dramatically reduce the amount of inventory held in the store for display listening. They would also give a much better representation of how the speakers will sound in the chosen environment.

The next chapter of this thesis will explain the theory behind the different components used in creating a virtual acoustic showroom. Chapter 3 discusses the development, implementation, and testing of the virtual acoustic environments. This is followed by the development, implementation, and testing of the speaker mimicking process discussed in Chapter 4. Chapter 5 discusses the human interaction with the two main components of the showroom, speaker mimicking and auralization of the virtual acoustic environments. Detailed in the chapter are psychoacoustic tests evaluating the perception of the system. Finally, Chapter 6 will give the conclusions gathered from the implementation and testing of this virtual acoustic showroom system. Recommendations and future work will also be presented in the conclusions.

Chapter 2

Theory

This chapter will discuss the theory behind the concepts used throughout this thesis. Section 2.1 discusses different types of speakers and the modeling of a speaker. Section 2.2 discusses room acoustics, including reverberation time and mode excitations. Section 2.3 explains the filtering used for speaker mimicking. Finally Section 2.4 discusses psychoacoustic issues.

2.1 Speaker Response

Over the years, speakers have changed and evolved into what they are today, primarily dynamic permanent magnet loudspeakers for home and car stereos and piezo electric speakers for laptop computers and cell phones. The dynamic permanent magnet loudspeaker (Figure 2.1 [18]) has a light coil of wire (voice coil), to which electric signals representing sound are applied. These change the magnetic field within the coil and cause movement by interacting with the magnetic field of the magnet. When the voice coil moves, it causes the cone to vibrate, which creates audible sound waves. The frequency at which the voice coil and cone vibrate depends on the frequency of the signals, while the distance that they move in and out depends on the amplitude [19].

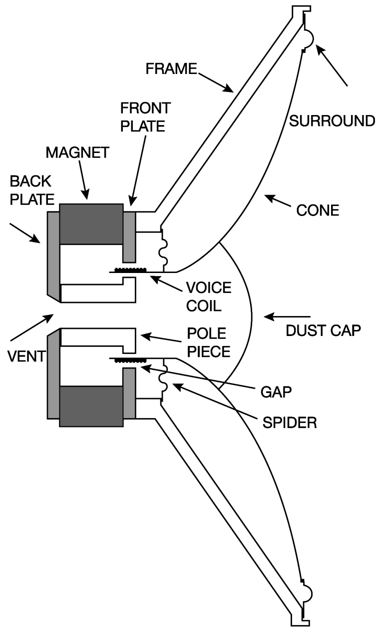

Figure 2.1: Cutaway view of a permanent magnet speaker.

The sound field that is emitted from a speaker relies on several different factors, including: the material the cone is made out of, the size of magnet, the size of the cone, responses of crossovers or filters used in multiple speaker cabinets, and a host of other factors. One important factor is the cabinet in which the speaker is housed. This cabinet acts as a baffle, forcing sound in certain directions. Without a baffle, the lower frequencies do not radiate efficiently because the fluid is pulled around to the back of the speaker by the vacuum created by the movement of the cone [19].

2.1.1 Types of Speakers

The majority of the work described in this thesis deals with bookshelf speakers, with some attention given to floor standing speakers, surround-sound speakers, sub-woofers, and car speakers. Each of these speaker types will be described in some detail within this section.

The major differences between these five types of speakers are: the size of the enclosure, the size of the speaker, the number of speakers, and the overall frequency range. Most home audio systems use bookshelf speakers that are either two-way or three-way. As indicated by the name, a two-way configuration has two speakers, a dome tweeter and a low-mid range driver, while a three-way configuration has a dome tweeter, a mid-range driver, and a low-range driver. Floor standing speakers are larger than bookshelf speakers and usually contain a sub woofer, or other device to enhance the low frequency response, in addition to the two/three-way configuration of the bookshelf. The sub-woofer adds in the very low frequencies that a bookshelf speaker is unable to reproduce.

Surround-sound speakers are small in size so that they can be placed inconspicuously around a room in 5.1 and 6.1 systems. A 5.1 system utilizes four surround-sound speakers placed around the listener, with one center speaker generally placed on the television, and a sub woofer. The 6.1 system has the same setup as the 5.1 system, but with the addition of a rear center speaker placed directly behind the listener. These systems give the listener a feeling of sound and motion all around them, due to the speaker directivity. The frequency range of surround-sound speakers is mostly in the mid to high range, due to their small size, but when used in conjunction with a separate sub-woofer, obtains the full frequency range desired. The small speakers provide the directivity of the sound, while the sub-woofer is omni-directional.

All of these speakers are contained within enclosures that are critical to the speaker's performance. The enclosure must be of appropriate dimensions for the frequency response of the speakers. It can have air ports in it, which allows air to move in and out of the enclosure with the movement of the speaker cone. It can be sealed so that the backing volume does not change as the speaker cone moves. Or it could have a combination of both of these. In this case, the sub-woofer would most likely be ported while the mid-range driver would be sealed. A ported enclosure will have an increase in bass response over a non-ported enclosure.

Another important part of what influences the response of a speaker is the use of crossovers, which is a filter that shapes the response of the roll-off from a speaker to match up with the ramp-up of another speaker in the same enclosure. This helps to reduce, or eliminate, major drops (a sudden decrease in amplitude greater than $10dB$ over a narrow band of frequencies) in the overall frequency response.

Unlike the speakers mentioned already, car speakers generally do not have their own enclosure. Instead, their backing volume comes from the door panel or the trunk space. They also come in many different sizes and shapes to fit various models of cars.

It is important to understand that differences in enclosures, crossovers, two-way versus three-way, size of speaker, etc. make a large difference in the performance of a speaker. However, for this thesis, it is not imperative to understand the intricacies in how each factor changes the response.

2.1.2 Modeling Radiation From a Speaker

A pulsating sphere is the simplest acoustic source to analyze and is denoted as a monopole. The radius of the monopole is considered to be very small in comparison to the wavelength of the radiated sound, therefore qualifying it as a simple source [20]. According to Kinsler et. al [21], the sound pressure p, of a monopole of average radius a, is given in equation 2.1.

$$p(r,t) = \frac{1}{2} j \rho_0 c \left(\frac{Q}{\lambda r} \right) e^{j(\omega t - kr)} \tag{2.1}$$

Where r is the distance from the source, t is time, ρ_0 is the density of air, c is the speed of sound in air, λ is the sound wavelength, ω is the angular frequency, k is the acoustic wavenumber ω/c, and Q is the complex source strength defined as

$$Q = 4\pi a^2 U_0 \tag{2.2}$$

Where $4\pi a^2$ is the surface area of the monopole and U_0 is the velocity of the monopole's surface.

Equation 2.1 is for a simple source in a free space. If however, the simple source is mounted on, or very close to, a rigid plane boundary, and the dimensions of the boundary are much greater than a wavelength of sound, the source would be classified as being baffled. The plane boundary, or baffle, forces all the sound energy to radiate in a hemisphere instead of a sphere. Therefore, the pressure field for the baffled source will be double that of the same source in free space. This is shown in the equations for the pressure amplitude of a simple source in free space (equation 2.3) and a simple source baffled (equation 2.4).

$$P = \frac{1}{2}\rho_0 c\, Q/\lambda r \tag{2.3}$$

$$P = \rho_0 c\, Q/\lambda r \tag{2.4}$$

Another acoustic source that is often used for modeling is the plane circular piston. This acoustic source is beneficial in this context, as it can be used to model a loudspeaker. According to Kinsler et. al [21] by dividing the surface of the piston into infinitesimal elements, ds, each of which acts as a baffled simple source of strength $dQ = U_0 ds$, the pressure at any field point can be obtained. Integrating over the surface of the piston gives the total pressure generated (equation 2.5). Figure 2.2 [21] defines the geometry of the piston and field point locations for deriving the acoustic field.

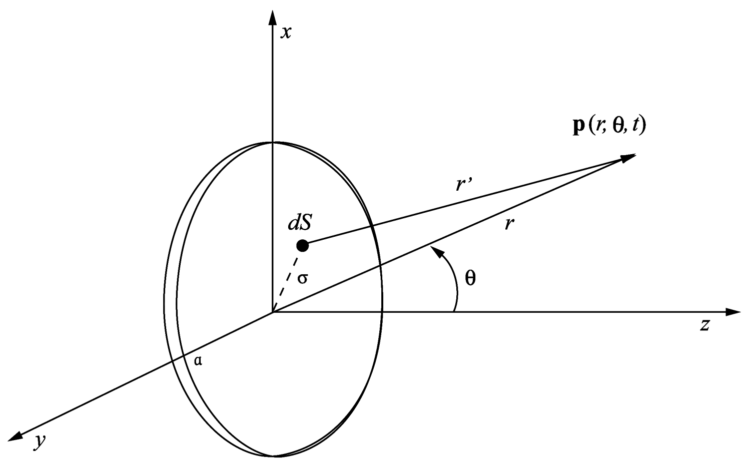

Figure 2.2: Geometry used in deriving the acoustic field of a baffled circular plane piston for $p(r, \theta, t)$.

$$p(r,\theta,t) = j\rho_0 c\,\frac{U_0}{\lambda}\int_s \frac{1}{r'} e^{j(\omega t - kr')} ds \tag{2.5}$$

Equation 2.5 is difficult to solve for explicitly, however, there exists two closed form solutions, one for the acoustic axis (along the z axis) and the other for the far field.

Along the acoustic axis, the pressure can be found using equation 2.6 and the pressure amplitude using equation 2.7.

$$p(r,0,t) = \rho_0 c\, U_0 \left\{1 - \exp\left[-jk(\sqrt{r^2+a^2} - r)\right]\right\} e^{j(\omega t - kr)} \tag{2.6}$$

$$P(r,0) = 2\rho_0 c\, U_0 \left|\sin\left\{\frac{1}{2}kr\left[(\sqrt{1+(a/r)^2} - 1)\right]\right\}\right| \tag{2.7}$$

Studying equation 2.7 reveals that there are fluctuations between 0 and $2\rho_0 c\, U_0$ as r ranges between 0 and ∞. These fluctuations, resulting in local minima and maxima, can be found for values of r using equation 2.8. Figure 2.3 [21] shows an example of the local minima and maxima as a function of r/a.

$$r_m/a = a/m\lambda - m\lambda/4a \qquad m = 0, 1, 2, \ldots \tag{2.8}$$

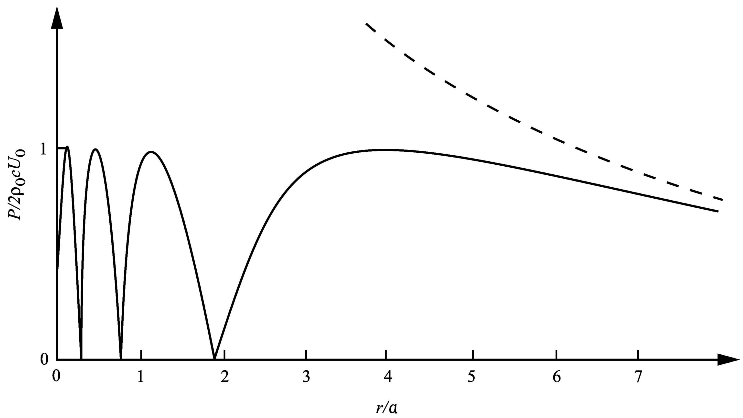

Figure 2.3: Axial pressure amplitude for a baffled circular plane piston of radius a radiating sound of wavenumber k with $ka = 8\pi$. Solid line is calculated from the exact theory. Dashed line is the far field approximation extrapolated into the near field.

When r_1 ($m = 1$) is greater than r, the pressure amplitude decreases with a $1/r$ dependance. For distances of $r < r_1$, the fluctuations between the minima and maxima suggest that the acoustic field is complicated close to the piston. Therefore the point of r_1 is used to distinguish between the near and far field.

The far field solution is restricted to distances where $r \gg a$. Figure 2.4 [21] shows the geometry used in deriving the pressure equation (equation 2.9) for this case.

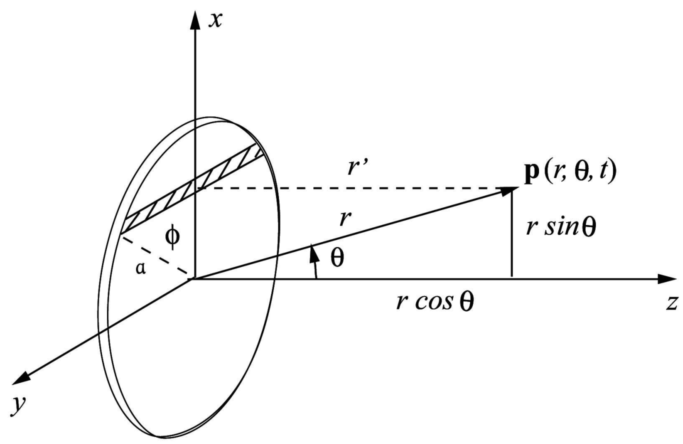

Figure 2.4: Geometry used in deriving the acoustic field of a baffled circular plane piston in the far field.

$$p(r,\theta,t) = \frac{j}{2}\rho_0 c\, U_0 \frac{a}{r} ka \left[\frac{2J_1(ka\sin\theta)}{ka\sin\theta}\right] e^{j(\omega t - kr)} \tag{2.9}$$

Where J_1 is the Bessel function of the first order. Figure 2.5 [21] shows a beam pattern for a circular plane piston with a radius of a and radiating sound with $ka = 10$. It is clear from the figure that at $\theta = 0°$ is the greatest amplitude of sound pressure. The maximum sound pressure in the first side lobe is approximately $17.5dB$ lower than the maximum sound pressure of the major lobe. Between each lobe of the beam pattern, there are pressure nodes. At these locations the sound pressure level becomes negligible. It is important to understand the beam pattern so that accurate measurements of a speaker's response can be obtained without placing the microphone between lobes.

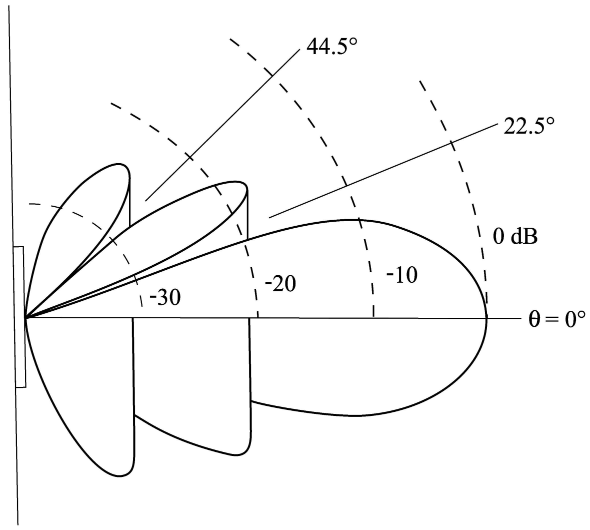

Figure 2.5: Beam pattern $b(\theta)$ for a circular plane piston of radius a radiating sound with $ka = 10$.

2.1.3 Determining Near Field Versus Far Field Radiation

As mentioned in the previous section, sound fields can be divided into distinguishable regions. Two of which are the geometric near field and the far field. Equation 2.7 revealed the fluctuations of minima and maxima that are characteristic of the near field. In the far field, the sound pressure levels decrease monotonically at a rate of $6dB$ for each doubling of the distance from the source and are characterized by the criteria given in equation 2.10 [20].

$$r \gg \lambda/(2\pi), \quad r \gg a, \quad r \gg \pi a^2/(2\lambda) \tag{2.10}$$

Where the inequality represents a factor of three or greater, r is the distance to the source, a is the characteristic source dimension, and λ is the wavelength of radiated sound.

2.2 Room Acoustics

The listening experience a person has in a room is regulated by the acoustical properties of that room. The size, shape, construction material, and furnishings are the acoustical makeup of the room's properties. Changing different combinations of these properties can give a room a very lively sound quality, such as in a church or a gymnasium, or give the room a dead sound quality, such as in a recording studio. In both of these cases, the amount of reverberation in the room is the main contributor to the overall sound quality of that room.

Reverberation is the non-direct sound energy that is reflected off surfaces in the room. When a source is present in a room, the sound waves produced by the source expand radially outward and impinge on surfaces in the room. The sound wave is then either reflected, absorbed, or diffracted by the surfaces, according to the frequency dependant material absorption characteristics, and the shape of the surfaces. The sound wave that reaches the listener's ear first by a linear path, without having been reflected off a surface, is called the direct sound. Measuring the time delay between the direct sound and later reflections gives two classifications for reverberation: early reverberation and late reverberation. Early reverberation refers to reflections that reach the listener up to $80ms$ after the direct sound, while late reverberation refers to reflections whose delays are greater than $80ms$ [22]. In practice, early reverberation is generally only reflected once, while late reverberation is comprised as reflections of reflections. This makes late reverberation diffuse, therefore causing individual reflections to be lost in the overall energy field.

The sound power in a room that is very reverberant will be considerably greater than that of a room that is only slightly reverberant, due to the large number of reflections. The sound field in the area near the source is dominated by the radiation from that source. As you move away from the source, more and more reflections will be audible, and, at some point, the sound field will be dominated by the reverberation in the room. Equation 2.11 is used to calculate the distance, r_d, from the source where the energy from both the source and the reverberation are equal.

$$r_d = \frac{1}{4}\sqrt{A/\pi} \tag{2.11}$$

Where A is the total sound absorption in the room, which will be discussed in Section 2.2.1 [21]. Any closer to the source, and the sound would be dominated by the source. Any further away, and the sound would be dominated by the reverberation.

When measuring the frequency response of a speaker in a room, the reverberation in that room will prohibit the true response of the speaker from being accurately measured. In order to measure the true response of a speaker, it is necessary to conduct the measurements in a free-field environment. A free-field environment can be simulated using an anechoic chamber. An anechoic chamber is constructed such that all sound waves that contact the chamber surfaces are absorbed, down to the lower frequency limit of the chamber. The lower limit for this true response is governed by the cutoff frequency of the anechoic chamber, which is determined by the depth of the anechoic wedges used in the chamber. The lowest frequency in which the anechoic chamber is considered to still provide a free-field environment is where the thickness, or depth, of each wedge is equal to 1/4 of a wavelength [23].

2.2.1 Reverberation Time

Knowing that the amount of reverberation in a room is controlled by size, shape, construction materials, and furnishings, it could be possible to classify the room's acoustical properties in terms of reverberation. Wallace Sabine developed an equation that takes into consideration the characteristics of a room and its reverberation, yielding one number to classify the room's acoustical properties, known as Reverberation Time (T_{60}). By definition the T_{60} is the length of time (sec) it takes for sound energy in a room to decay $60dB$ from a steady state once the source is turned off. Sabine's reverberation equation is given in equation 2.12 [24].

$$T = 0.163\frac{V}{S\bar{\alpha} + 4mV} \tag{2.12}$$

Where V is the volume (m^3) of the room, S is the total surface area (m^2) of the room, $\bar{\alpha}$ is the average absorption coefficient of the room and is defined in equation 2.13, and m is the attenuation constant of air. For small rooms, the term $4mV$ can be dropped since the attenuation due to air is negligible.

$$\bar{\alpha} = \frac{1}{S}\sum S_i\alpha_i \tag{2.13}$$

Where S is the total surface area and α_i is the absorption coefficient for surface S_i.

Sabine's T_{60} equation assumes a uniform distribution of absorption throughout a room, creating a relatively diffuse sound field. However, if one surface in the room is highly absorptive and the others are not, this equation could stray from the true values by a significant amount [25]. Other analytical methods for calculating T_{60}'s have been developed to overcome this limitation faced by Sabine's formula. Fitzroy developed an equation to be used when opposing walls in a room have similar absorption coefficients, equation 2.14.

$$T = 0.16\frac{V}{S^2}\left[\frac{-S_x}{\ln(1-\alpha_x)} + \frac{-S_y}{\ln(1-\alpha_y)} + \frac{-S_z}{\ln(1-\alpha_z)}\right] \tag{2.14}$$

Where $S_x,\ S_y,\ S_z$ are the areas of the walls normal to the coordinate system $x,\ y,\ z$ and $\alpha_x,\ \alpha_y,$ and α_z are the average absorption coefficients for those walls. A modification to equation 2.14 was proposed by Neubauer which takes into consideration opposing walls with different absorption coefficients, equation 2.15.

$$T = 0.16\frac{V}{S^2}\left[\frac{S_x}{\alpha_x^*} + \frac{S_y + S_z}{\alpha_{y,z}^*}\right] \tag{2.15}$$

Where α_x^* corresponds to the absorption of the walls parallel to the x-axis and $\alpha_{y,z}^*$ that of walls parallel to the y and z axes. See [26] for the derivation and more details.

The importance of having an analytical model that takes into account different amounts of absorption from the surfaces, rather than a total average absorption, can be seen when, for example, the floor and ceiling of a room are highly reflective and the walls are very absorbant. In this situation, it is possible to have standing waves between the floor and ceiling which would dominate the reverberation decay and not be recognizable by the classical reverberation equations like Sabine's.

To make T_{60} predictions easier and more accurate, absorption coefficients (α) for an extensive number of materials have been measured and tabulated. An absorption coefficient is the measured ratio of the absorbed energy and the incident energy of that material. These measurements are generally taken in an impedance tube or in a reverberant test chamber (see [20, 24] for more details). Since absorption is frequency dependant, these tables list values for multiple frequencies, generally octave bands of $125, 250, 500, 1000, 2000$, and $4000Hz$. Values for α range between 0, being not absorbant, to 1, being completely absorbant. For example, at $500Hz$ concrete has an α of 0.01 while $50mm$ thick fiber-glass has an α of 1.00 [20].

In order to determine a room's reverberation time experimentally, white noise is played through a speaker in the enclosure until the sound energy has reached a steady state. The source is then abruptly shut off, and a microphone is used to measure the pressure in the room as it decays from steady state. In most cases, due to ambient background noise, it is rare that the signal will actually decay by $60dB$. This is easily overcome by interpolating down to $-60dB$, relative to the steady state level. A decay of $30dB$, T_{30}, is sometimes used instead of $60dB$; however, most texts use T_{60} as their measure of reverberation time.

2.2.2 Mode Excitation

Room modes, or standing waves, occur between surfaces when there is a relationship of $\lambda = 2d/n$, where λ is the wavelength of the sound wave, d is the distance between surfaces, and n is the mode number. From this, it is clear that the lowest frequency that could create a standing wave ($n = 1$) is where the distance between surfaces is equal to half a wavelength. This frequency is known as the fundamental frequency.

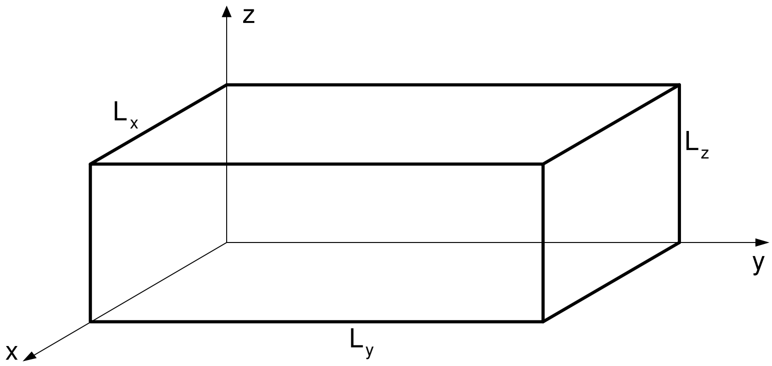

Figure 2.6: Dimensions of a rectangular room

Finding frequencies for different room modes, requires solving the wave equation, equation 2.16, using separation of variables and the boundary conditions of the room. This is a well-defined process and the details can be found in many texts, including [20, 21, 24].

$$\frac{\partial^2 p}{\partial x^2} + \frac{\partial^2 p}{\partial y^2} + \frac{\partial^2 p}{\partial z^2} = \frac{1}{c^2}\frac{\partial^2 p}{\partial t^2} \tag{2.16}$$

The solution to the wave equation, 2.16, in a loss-less, rigid-walled, rectangular cavity of dimensions L_x, L_y, and L_z, results in a three-dimensional standing wave equation 2.17.

$$P_{n_x n_y n_z}(x, y, z) = A \cos\frac{n_x \pi x}{L_x} \cos\frac{n_y \pi y}{L_y} \cos\frac{n_z \pi z}{L_z} e^{i\omega t} \tag{2.17}$$

Where A is an arbitrary constant representing the amplitude of the three-dimensional standing wave equation. The eigenfunctions $P_{n_x n_y n_z}$ obtained in equation 2.17 can be used to define the useful relationship between frequency and room modes, given in equation 2.18.

$$f_n = \frac{c}{2}\sqrt{\left(\frac{n_x}{L_x}\right)^2 + \left(\frac{n_y}{L_y}\right)^2 + \left(\frac{n_z}{L_z}\right)^2} \qquad n = 0, 1, 2, \ldots \tag{2.18}$$

The subscript n on the variable f indicates that the particular frequency is a function of the mode numbers n_x, n_y, and n_z and c is the speed of sound $343 m/s$. There are three types of modes that can occur in the rectangular enclosure: axial, tangential, and oblique. Axial modes are those where the sound wave is parallel to an axis in the enclosure, for instance $(n_x, 0, 0)$ mode, and only reflects between two opposing surfaces. Tangential modes are those where the sound wave is parallel to two opposing surfaces, for instance $(n_x, n_y, 0)$ mode, and reflecting between the other four. Oblique modes are sound waves that are oblique to all room surfaces (n_x, n_y, n_z), reflecting off all six surfaces [20]. Having $n = 0$ for all three directions, for instance (0,0,0) mode, is the zeroeth order mode that corresponds to the static compression of the air in the room, which is strongly affected by "leaks" in the room, such as opening a window.

As can be shown from equation 2.17, moving from one location to another within an enclosure can yield significantly different pressure levels, depending on whether that location is at a pressure maximum (anti-node) or a pressure minimum (node). It is very important to know the effect of modes in an enclosure when selecting the placement for a sound source used for measurements. If the sound source is located at a pressure anti-node of a particular mode, that mode can be excited. In a rectangular enclosure, all normal modes have a pressure anti-node in the corners. Therefore, to excite all normal modes in a room, the source should be placed in the corner where all three surfaces intersect. The distinction of individual modes is only realistic for relatively low frequencies. After a certain point, there are too many modes in a given frequency band, and the modes begin to overlap. The frequency at which this crossover occurs is called the Schroeder frequency and is denoted by equation 2.19 [21].

$$f \geq 2000\sqrt{T/V} \tag{2.19}$$

Where T is the reverberation time (sec) and V is the room volume (m^3). Above this frequency single room modes are generally not a concern.

2.2.3 Image Source

There are many different methods used to model a room's acoustics, including: an exact solution, ray tracing, and image source. All of these methods have the same underlying assumption: that the acoustic field satisfies the wave equation [27].

An exact solution for a room's acoustical properties is determined by solving the differential wave equation (equation 2.16). It is solved in a similar manner as discussed in Section 2.2.2, with the exception of boundary conditions. The solution earlier assumed to have completely rigid walls and, therefore, having the particle velocity of the normal components as zero. However, this is an unrealistic assumption for most room surfaces. Modifying the boundary conditions to include a surface impedance term will take into account non-rigid surfaces and include complex terms in the analysis of the wave equation. This method is very complicated and, in most cases, the equation can only be solved numerically. However, the benefit of this method is that it is an exact low-frequency solution. It is also the most reliable method and can help keep results from simplified models in perspective [24].

The ray tracing method is based on tracing sound paths in an environment where a sound source radiates a finite number of sound "rays" in all directions. As each sound ray impinges on a surface, it is filtered in correlation with the angle of incidence and wall absorption. The ray is traced until it either collides with the listener, or until its energy level falls below a given threshold. A benefit of this method is its ability to accurately model the effects of diffusion. However, this accuracy comes at a computationally intensive price [22].

Figure 2.7 shows a simple two-dimensional ray tracing model with diffusion at each reflection point. An omni-directional source is placed toward the left side of the rectangular enclosure, with sound "rays" emitting in every direction, and the receiver's position near the right side. Each of the sound "rays" travel some distance before being reflected off a surface of the enclosure and directed to the listener's position. At each reflection point, the ray is slightly diffused due to the texture and roughness of the surface finish in a frequency-dependant manner.

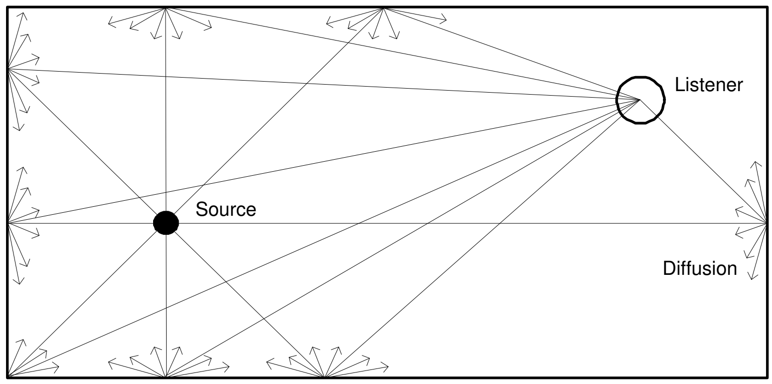

Figure 2.7: Simple two-dimensional ray tracing model with an omni-directional source on the left, the receiver on the right and diffusion at each reflection point.

The image source method is a variant of the ray tracing technique, where the laws of optics are used and a sound wave can be modeled as a beam of light in a "room of mirrors." It is a very robust method, and guarantees that all specular paths, up to a given order or reverberation time, will be found [5]. It is not, however, well suited for complex room shapes because the image sources are not arranged in a pattern, and a condition of visibility has to be checked for each source.

The room environments considered in this thesis are, largely, simple rectangular shapes. Therefore, virtual sources can be arranged in a pattern on the outside of the room representing reflections, making the virtual sources very easy to calculate using the image source method. Figure 2.8 shows an example of a source and two of its virtual sources. From this, it is clear that the direct path is the first sound ray to reach the listener's position. This is followed by the first order reflection, which is modeled as virtual source 1. Virtual source 2 provides the second order reflection to the listener. The order of a reflection is represented by the number of reflections the ray incurs prior to reaching the listener. In other words, reflections with order n are reflected off the enclosure's walls n times. The number of first order reflections is determined by the number of walls in the room.

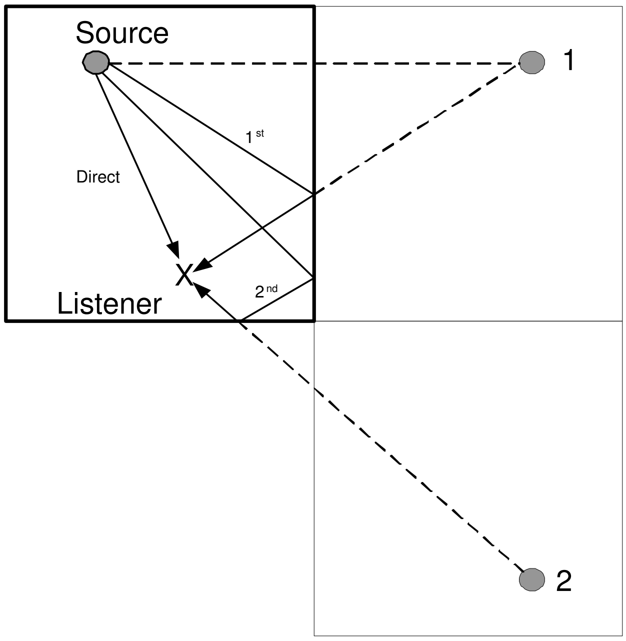

Figure 2.8: Simulation of direct, 1^{st} order, and 2^{nd} order reflections from virtual sources.

To model a room with this method, a pressure wave equation needs to be solved. An initial assumption about the room is that it has rigid walls as boundary conditions. According to Allen and Berkley [7] a point source at a single frequency can be modeled using equation 2.20.

$$p(\omega, X, X') = \frac{e^{j\omega(R/c-t)}}{4\pi R} \tag{2.20}$$

Where X is the source locations (x, y, z), X' is the listener location (x', y', z') and $R = |X - X'|$. Having rigid walls allows for an image to be placed symmetrically on the opposite side of the wall, giving equation 2.21 for a single reflection.

$$p(\omega, X, X') = \left[\frac{e^{j(\omega/c)R_+}}{4\pi R_+} + \frac{e^{j(\omega/c)R_-}}{4\pi R_-}\right] e^{-j\omega t} \tag{2.21}$$

with the distances from the microphone (or listener) to the source R_- and to the image R_+ being

$$R_-^2 = (x - x')^2 + (y - y')^2 + (z - z')^2 \tag{2.22}$$

$$R_+^2 = (x + x')^2 + (y - y')^2 + (z - z')^2 \tag{2.23}$$

where $x = 0$ is at the wall. For all six walls of an enclosure, equation 2.21 can be written in the frequency domain as

$$p(\omega, X, X') = \sum_{p=1}^{8} \sum_{r=-\infty}^{\infty} \frac{e^{j(\omega/c)|R_p + R_r|}}{4\pi |R_p + R_r|} e^{-j\omega t} \tag{2.24}$$

where $p = 8$ given the permutations over $\pm$ of $R_p = (x \pm x', y \pm y', z \pm z')$, r is the integer vector (n, l, m) and $R_r = 2(nL_x, lL_y, mL_z)$.

Taking the time domain representation of equation 2.24 yields the room impulse response, equation 2.25.

$$p(t, X, X') = \sum_{p=1}^{8} \sum_{r=-\infty}^{\infty} \frac{\delta\left[t - (|R_p + R_r|/c)\right]}{4\pi |R_p + R_r|} \tag{2.25}$$

Figure 2.9 shows a two dimensional view of the original source (0) and the image source placements, where their number represents their reflection order. Equation 2.25 is representative of the three dimensional construction of image sources.

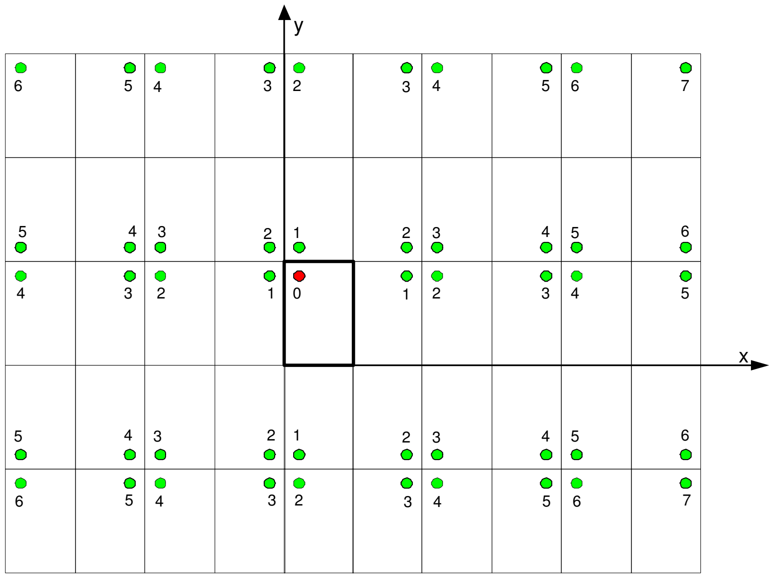

Figure 2.9: Two dimensional view of image source placements.

Equation 2.25 can be modified so that the effect of the walls can be taken into consideration. Assigning absorption coefficients to the walls and relaxing the rigid wall constraint imposed earlier, will allow each reflection to be attenuated according to the wall(s) that it "passes through." The absorption coefficient (α) and reflection coefficient (β) are related by $\alpha = 1 - \beta^2$ and, when included in equation 2.25, result in the room impulse response of equation 2.26.

$$p(t, X, X') = \sum_{p=0}^{1} \sum_{r=-\infty}^{\infty} \beta_{x1}^{|n-q|} \beta_{x2}^{|n|} \beta_{y1}^{|l-j|} \beta_{y2}^{|l|} \beta_{z1}^{|m-k|} \beta_{z2}^{|m|} \left[\frac{\delta\left[t - (|R_p + R_r|/c)\right]}{4\pi |R_p + R_r|} \right] \tag{2.26}$$

Where R_p is expressed in terms of $p = (q, j, k)$ as

$$R_p = (x - x' + 2qx', y - y' + 2jy', z - z' + 2kz') \tag{2.27}$$

and the reflection coefficients (β) are indexed with 1 or 2 representing the walls adjacent to the coordinate origin and the opposing wall, respectively. Here, the sum with vector index p is used to indicate three sums, one for each of the three components of $p = (q, j, k)$ and

$r = (n, l, m)$ is a similar sum.

The summation of pressure at the listener's location due to each reflection makes up the overall room impulse response. If the listener's location moves, the summation at the new location would have to be calculated, while the image sources would remain the same. By using absorption coefficients, which are frequency dependent, these calculations can be performed for multiple frequencies to obtain a more accurate room response. Figure 2.10 gives an example an impulse response that was measured in a small church [28].

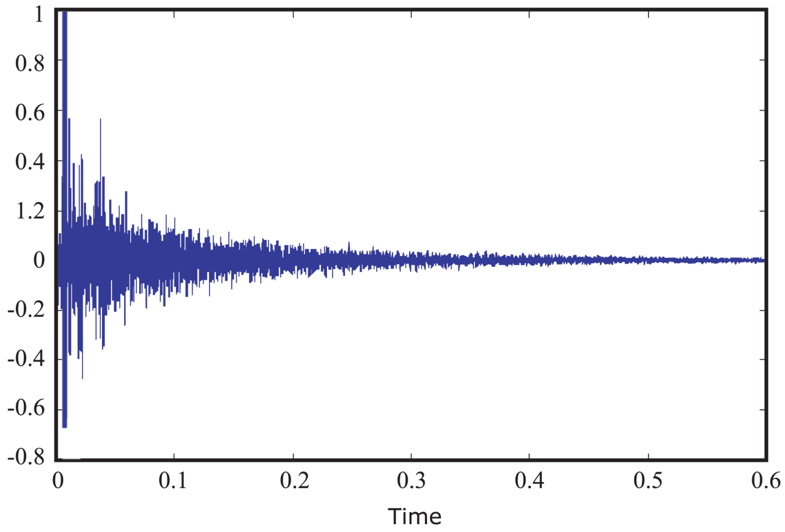

Figure 2.10: Impulse response of a small church.

2.3 Filtering

To make one speaker sound like another speaker requires a method of pre-filtering the signal being sent to the speaker. One such method of filtering is called inverse filtering. An inverse filter for a speaker transfer function H, in the frequency domain, is denoted as

$$H_{inv} = 1/H \tag{2.28}$$

By including the response of the second speaker with the inverse filter, the dynamics of the second speaker will result when played through the first speaker. In other words if H_o is the transfer function in the frequency domain of the original speaker to some point $L_{(x,y,z)}$ and $1/H_r$ is the inverse of the transfer function for the reference speaker to the same point, then the filter is given as equation 2.29.

$$H_{mimic} = H_o/H_r \tag{2.29}$$

The mimic filter, H_{mimic} (equation 2.29), is never performed directly in the time domain, but rather, it is calculated in the frequency domain [29]. A signal that is sent to a speaker is in the time domain, and in order to convolve the mimic filter with that signal, the filter must also be in the time domain. An Inverse Fast Fourier Transform (IFFT) algorithm is used to transform the filter from the frequency domain to the time domain. If the length of the filter is a power of two (2^n) then it is possible to use this efficient algorithm [30].

Once it is converted into the time domain, the filter can be convolved with the input signal for the reference speaker, which is also in the time domain. Convolution in the time domain is equivalent to multiplication in the frequency domain, so that

$$y(t) = x(t) * h(t) \Leftrightarrow Y(\omega) = X(\omega)H(\omega) \tag{2.30}$$

where $*$ is the convolution operator, $y(t)$ is the time domain output, $x(t)$ is the time domain input, $h(t)$ is the impulse response of the speaker. $Y(\omega)$, $X(\omega)$, and $H(\omega)$ are the output, input and frequency response function respectively, with ω specifying the frequency domain.

Multiplying the mimic filter (equation 2.29) with an input signal, X_o, results in a modified input signal X_m. The new input signal is then sent to the reference speaker resulting in

$$\begin{aligned} y(t) &= IFFT\Big[X_m H_r\Big] \\ &= IFFT\Big[X_o H_{mimic} H_r\Big] \\ &= IFFT\Big[X_o (H_o/H_r) H_r\Big] \\ &= IFFT\Big[X_o H_o\Big] \end{aligned} \tag{2.31}$$

where $IFFT$ represents the conversion from the frequency domain to the time domain using the Inverse Fast Fourier Transform. As can be seen by equation 2.31, the resulting output is a time domain signal that has the response of the original input signal and the original speaker. Meaning, the reference speaker now has the response of the original speaker playing the original sound signal. Physically, this process takes the difference between the response of the reference speaker and the original speaker to remove the response of the reference speaker.

The technique of using an inverse filter is not limited to mimicking speakers. Mäkivirta et. al [17] describe its use to equalize a room's modal response at low frequencies using a loudspeaker. Mouchtaris et. al [31] use this technique for removing the response of headphones used in immersive audio rendering.

2.4 Psychological Acoustics

Sound can very easily be measured in experiments to obtain different qualities and properties of that sound. These experiments are repeatable, yielding the same level of accuracy time and time again. The importance of these experiments is placed not only on the repeatability of the measurements, but also on the basis that the information obtained is quantifiable in scientific measurements, dimensions, and terms. When introducing humans as the "measurement" device, the results change from being scientifically accurate to subjective perceptions of the sound. A few of the many factors which influence a person's perception of the sound will be discussed later in this section.

Visual cues play a big role in the way sound is heard and interpreted and are generally considered to be the dominant stimuli. An example of the dominant effect visual cues have over auditory cues would be the television. The sound from a television comes from a loudspeaker that is generally below the screen. Even thought its placement is not in the screen, the sound is perceived as coming from the actors mouth. With this type of cue determination, the visual stimulus "captures" the location derived from audio cues and is called visual capture [22]. An effect similar to visual capture is called the ventriloquism effect and is where the apparent position of the virtual sound source is influenced when correlated visual objects are present [22, 32]. An example of this would be where sound is played through one speaker, but there are multiple other speakers nearby. The true source would be difficult to distinguish from the others.

Significant research has been conducted with auditory perception and the factors contributing to it. One of these factors is the direct-to-reverberant ratio [22, 33, 34]. It has been claimed that this ratio can be an absolute cue, meaning enough information can be gathered from it to determine absolute distances rather than relative distances [34]. It is argued that the listener contains, a *priori*, a knowledge of environmental acoustic properties, in order for the direct-to-reverberant ratio to provide an absolute distance cue. Without this knowledge, it may only be possible to determine relative distances using the direct-to-reverberant ratio.

2.4.1 Psychoacoustic Test Design

In designing a psychoacoustic test, factors, such as where to conduct the test, what stimulus to use, and whether to use headphones or speakers, among others, need to be considered. These factors are generally dictated by what the test is trying to determine, for example; determining the perception of reverberation in a room, determining source location, etc.

According to Kuttruff [24], there are two methods that can be employed for investigating the subjective effects of complex sound fields. One is to simulate sound fields in an anechoic chamber, while the other is to directly judge the perception of acoustic qualities of completed halls. With these two methods of investigating the subjective effects of complex sound fields, Kuttruff is referring to room acoustics, more specifically reverberation, and its subjective effects on sound quality, speech intelligibility, spaciousness, source location, and so on.

Selecting the appropriate method of presenting the sound source is also dependent on the type of test to be conducted. Loudspeakers are often used for testing the precedence effect [35] (The precedence effect is known as the ability to localize a sound source from the first sound that reaches a listener's ear.) and for assessing auditory distance perception [33]. Headphones are often used for virtual acoustic testing when real-time implementation is desired [14].

The stimulus, or sound source, used for the psychoacoustic experiment depends greatly on what the test is supposed to determine. Clifton et. al [36] conducted tests to determine what the precedence effect tells listeners about room acoustics. In this experiment the authors used white noise bursts for the sound source. Multiple bursts were separated by a short variable delay to create a train of bursts. Zahorik [33] used single noise bursts as well as a recorded speech syllable for assessing auditory distance perception. Niaounakis and Davies [37] used anechoically recorded music for their sound source in testing the perception of reverberation time in small listening rooms. Each of these three experiments used different sound sources, even though each test was affected by the inherent reverberation in the room the tests were conducted in.

The psychoacoustic tests detailed in Chapter 5 were designed to represent the implementation conditions of the virtual acoustic showroom system. In other words, the location of the test, the method of presenting the stimulus, and the stimulus given are all representative of the conditions a customer would have in the showroom.

Each psychoacoustic test was conducted in the prototype virtual showroom, which was constructed to be as anechoic as possible, as will be discussed in Section 3.1. With the showroom having an anechoic treatment, the acoustic properties of the showroom will not interfere with the auralization of different acoustic environments.

The stimulus was presented to the listener through multiple loudspeakers, making the properties of the showroom very important. Multiple loudspeakers were used so that the auralization of the virtual acoustic environments could be obtained without the use of headphones. The loudspeakers also allowed the majority of area in the showroom to take on the acoustic properties of the virtual acoustic environment. During the implementation of the showroom system, the customers will have the ability to walk around in the showroom, without being attached to a set of headphones. Also, the customers will be using the showroom system to listen to speakers, not headphones. Using headphones to simulate speakers could reduce the customers perception of the accuracy of the system.

Music was used for the stimulus in each psychoacoustic test because music will be used once the showroom system is implemented. White noise bursts or anechoically recorded music could have been used as the stimulus, however, neither of those would be used during the implementation of the system.

Setting the factors of the psychoacoustic tests to be representative of what a customer would be faced with the implementation of the showroom system, creates realistic test conditions. The realistic test conditions provide a bases for validating the perception of the virtual acoustic showroom system.

2.4.2 Statistical Analysis

As with any human subject testing conducted, a statistical analysis should be used to analyze the data. The type of statistical analysis will depend on the data the test is generating and what is to be proved by the test. Two statistical analysis methods used for the psychological tests in chapter 5 are the Binomial Distribution and the Paired t-Test. Both of these methods, described below, are according to Vining [38].

A binomial experiment can be described by the following five conditions:

1. The experiment consists of n identical attempts or "trials."
2. Each trial results in one of two outcomes: one outcome that is considered a "success" and the other considered a "failure."
3. The probability of a success on a single trial is p and remains the same from trial to trial. As a result, the probability of a failure is $q = 1 - p$.
4. The trials are independent.
5. The value of Y is of interest, which represents the total number of success among the n trials.

The probability function for a binomial random variable Y is

$$p(y) = \begin{cases} \binom{n}{y} p^y \cdot q^{n-y} = \frac{n!}{y!(n-y)!} p^y \cdot q^{n-y} & \text{for } y = 0, 1, 2, \ldots, n \\ 0 & \text{otherwise} \end{cases} \tag{2.32}$$

where $\binom{n}{y}$ is used to denote the number of ways to get y success from n total attempts. The normal approximation to the binomial provides a firm basis for estimating and testing the true proportion of "success" in a population, p. An estimate of p is the sample proportion

$$\hat{p} = \frac{Y}{n} \tag{2.33}$$

where, again, Y is the number of success and n is the sample size. Then, the standard normal random variable Z for the normal distribution is given as

$$Z = \frac{\hat{p} - p_0}{\sqrt{(p_0 \cdot q_0)/n}} \tag{2.34}$$

where p_0 is the nominal value of p and $q_0 = 1 - p_0$. The value of Z is the test statistic used in hypothesis testing, which will be described later.

The paired t-test is used to compare two samples that are not independent of each other. For this comparison, the ith difference between samples is first taken

$$d_i = y_{1,i} - y_{2,i} \tag{2.35}$$

where y_1 and y_2 refer to the data sample sets 1 and 2 respectively. Let δ be the true mean difference between the two data sets, then the estimator of δ is the sample mean difference, $\bar{d}$.

$$\bar{d} = \frac{\sum\limits_{i=1}^{n} d_i}{n} \tag{2.36}$$

Now the sample variance of the d_i's is given as

$$s_d^2 = \frac{n \sum\limits_{i=1}^{n} d_i^2 - \left(\sum\limits_{i=1}^{n} d_i \right)^2}{n \cdot (n-1)}. \tag{2.37}$$

With this the test statistic, t, is defined as

$$t = \frac{\bar{d} - \delta_0}{s_d / \sqrt{n}} \tag{2.38}$$

where δ_0 is the estimated difference between the two data sets for hypothesis testing.

Hypothesis testing a process in which a nominal claim, H_0, is made and an alternative claim, H_a, is suggested. The appropriate test statistic is then used to determine if the nominal claim (or null hypothesis) should be rejected based on a significance level of α. For example, if the hypothesis tests, for the paired t-test, are $H_0 : \delta = \delta_0$ and $H_a : \delta < \delta_0$ and it was found that $t < -t_{n-1,\alpha}$, where the value for $t_{n-1,\alpha}$ is found in a table. Then the null hypothesis would be rejected, which suggests that there is sufficient evidence to conclude $\delta < \delta_0$.

Instead of reporting values for the test statistics, t or Z as above, statistical software packages give p-values, which are defined as the probability of seeing the particular value of the test statistic if H_0 is true. When the p-value is less than α, H_0 is rejected. This allows different values of α to be selected by the user for determining the appropriate significance level to use for the test.

Both the binomial and the paired t-tests were used for analyzing the data obtained from the psychological tests conducted in chapter 5. They were appropriate for the testing based on the experiments design. When designing psychological tests, the statistical analysis method must be appropriate for that test.

2.5 Conclusion

Each section in this chapter has thoroughly discussed concepts and ideas that are imperative to creating a virtual acoustic showroom system. Speaker responses, including modeling sound radiation from a monopole and a baffled piston, and the distinction of near field and far field was introduced. A room's acoustic qualities, and the measures of such qualities, were defined as well as a detailed description of the image source method. The method of inverse filtering for its use in speaker mimicking was defined. Finally, different aspects of psychoacoustic testing were discussed, including test design and statistical analysis of the test data.

The next chapter will discuss the development and empirical verification of virtual acoustic environments using theory presented in Section 2.2 on room acoustics.

Chapter 3

Adding Response of Virtual Environment

One of the objectives for this research is to employ a system which would allow a user to select a listening environment that is comparable to the room for which they are purchasing speakers. This would give the user a more accurate representation of how a set of speakers sounds after they are installed in their home. For this system, a special showroom was constructed that has a minimal amount of environmental influence on the system due to the acoustic properties of the room, and is presented in Section 3.1. The simulated acoustical environments are calculated using the image source method and are used to pre-filter the signal going into the speakers in the showroom as described in Section 3.2. Using absorption coefficients for building materials typically found in home construction, different acoustical environments can be simulated, as shown in Section 3.3. Once the acoustical environment filters are created, they are verified prior to implementation into the system (Section 3.4) and again after implementation into the system (Section 3.5).

3.1 Virtual Showroom Prototype Design and Validation

The goal of the virtual showroom is to provide a customizable acoustical environment where users will enter and are able to select a listening environment. The acoustic properties of the virtual showroom are very important in the overall performance of the virtual acoustic system. In order to simulate acoustic environments in the showroom, the room must be as anechoic as possible. If the showroom is reverberant, less reverberant virtual acoustic environments may not be auralized, as will be discussed in Section 3.1.1. Reverberation in a room can not be removed, making it necessary to construct a prototype of the showroom and test its acoustic properties to determine if it is sufficiently anechoic. After verifying the acoustic properties of the virtual showroom, the implementation and validation of the simulated acoustic environments will be conducted within the showroom. These acoustic environment simulations will be for the bookshelf system, the DVD surround sound system, and the car simulation system.

For simulating the acoustic environments of cars, a prototype of the car simulator must also be built and its acoustic properties verified. The functionality of the car simulator is similar to the showroom, as it will be used in simulating acoustic environments. It is, however, a subset system to the showroom, where its key function is source location, which means the speakers are placed in appropriate locations for a wide range of vehicles. Since it is not a stand alone system, the car simulator must be located within the showroom where the surrounding acoustic properties are controlled. The design and verification for these two prototype systems are discussed in the following sections.

3.1.1 Room Design and Validation

There are two main concerns addressed in the design of the showroom: acoustic properties and appearance. First is the acoustic properties of the room. If the room is overly reverberant, it would be very difficult, if not impossible, to simulate other acoustic environments in it. This can be demonstrated by observing the T_{60} times for both the showroom and the simulated environment. The overall sound level in a room with a short reverberation time will decrease quickly. Having a long reverberation time will take longer for the sound level to decrease, as the definition of reverberation time indicates. When the two environments are overlapped by simulating one environment within the showroom, the longer reverberation time will dominate the overall sound level decay. If the reverberation time in the showroom were longer than that of the simulated environment, the reverberation time would be dominated by that of the showroom, and the simulated environment would not be audible. Therefore, the showroom needs to be as anechoic as possible. The second concern addresses the showroom's appearance. It will be used, in part, as a marketing tool and needs to be appealing and inviting to the users. Therefore, the showroom was designed to look like a typical living room or home theater. Figure 3.1 shows a picture of a small portion of the prototype room.

The room was built in $4ft \times 8ft$ modular sections and then assembled together. Each of the four walls were comprised of 3 sections, making the outside dimensions $12ft \times 12ft \times 8ft$. Each section had a $2in \times 6in$ frame, a $4ft \times 8ft$ sheet of $3/4in$ plywood on the backside, $4.5in$ fiberglass insulation in the frame, and was covered with an interweave fabric on the interior. Three $4ft \times 12ft$ sections were built to fit across the top of the room for a ceiling. One wall section had a door framed into it for access in and out of the room. The room was assembled on top of a thick carpet pad and carpet to aid in insulating outside noise and in achieving the desired anechoic treatment. An electrical outlet and lights were wired into two sections for use inside the room. Trim was used to cover up the seams between sections and give it a finished look.

Using the method of measuring T_{60} times detailed in Section 2.2.1, the performance

Figure 3.1: Virtual showroom prototype.

of the acoustic treatment of the showroom was evaluated. Two sets of data were taken using three randomly placed microphones for each set. White noise, filtered in octave bands from $250Hz$ to $8000Hz$, was used to collect eight T_{60} measurements for each microphone. An average T_{60} time was then taken between the three microphones and two data sets for each octave band. The results are displayed in Figure 3.2, along with two simulated rooms where the T_{60} of each was calculated using equation 2.12. Both simulated rooms are the same size as the showroom, but one has an anechoic (dead) acoustic treatment and the other has a medium acoustic treatment. The absorption coefficients for these two room are given in Table 3.2 on page 50.

As can be seen from the figure, the reverberation time for the showroom is ≤ 0.2 seconds over the frequency range of $500Hz$ to $8000Hz$. Therefore the acoustic treatment is creating an anechoic environment above $500Hz$. Below $500Hz$ the reverberation time

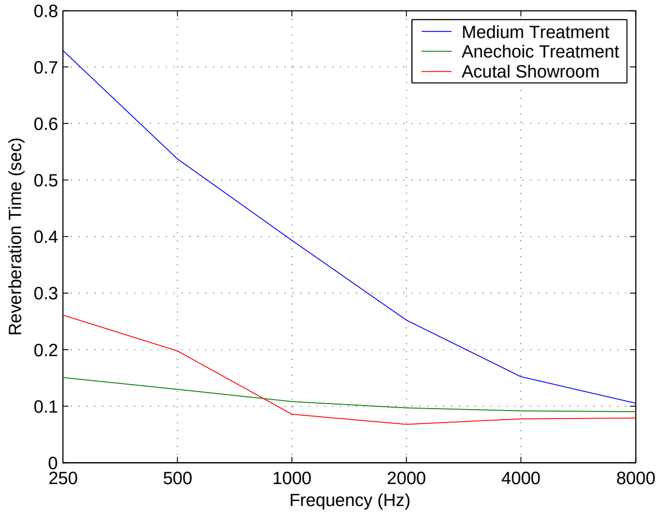

Figure 3.2: T_{60} measurement of virtual showroom compared with T_{60} calculations for two other rooms.

increases by nearly 0.1 seconds, which can be attributed to the thickness of the showroom walls. In an anechoic chamber the lowest frequency in which the chamber is considered to still provide an anechoic environment is where the thickness of the anechoic wedge is equal to 1/4 of a wavelength [23]. Since the walls of the showroom are $6in(0.1524m)$ thick, a cutoff frequency of approximately $563Hz$ is obtained. The acoustic treatment of the virtual showroom provides a good compromise between achieving a true anechoic environment and creating a realistic living environment. It is clear that an anechoic environment above $500Hz$ was achieved.

3.1.2 Car Design and Validation

The acoustical design considerations used in building the virtual showroom were also used when building the car simulator. The main consideration is making the simulator as anechoic as possible. Again, this is to avoid unnecessary interference caused by the acoustical properties of the car. The simulator is meant to represent a vehicle, and therefore, the properties of the simulator should be similar to those found in typical vehicles, for both source

Figure 3.3: Virtual car simulator.

positioning and psychological purposes.

The size of the simulator was set to be approximately equal to that of an SUV or a mini-van, allowing different speaker and seating arrangements to be tested. A total of nine speakers were placed in the simulator corresponding to one in each of four door locations, two in the front dash representing either dash or A-pilar speakers, two in the rear dash locations, and one sub woofer below the rear dash in a trunk or cargo area. Two bench seats from a van were mounted into the simulator for seating locations. These locations corresponded to front and rear seats in a vehicle. A $2in \times 4in$ frame construction was used and was covered by $1in$ thick melamine foam. To finish the simulator, and make it visually pleasing, a light gray interweave fabric was used to cover all exposed foam. The finished simulator with speaker boxes in appropriate locations is shown in Figure 3.3.

To validate the construction of the car simulator, two sets of measurements are required, allowing a comparison between the response of the speakers both with, and without, the car simulator to be made. First, the car simulator was set up in the anechoic chamber with all nine speakers in place (Figure 3.4). Four microphones, used to acquire the response of the car, were located in the approximate locations of four passengers' heads. One of the nine speakers was designated as the source speaker and was the only one that played white

Figure 3.4: Car simulator test setup in anechoic chamber.

noise. The source speaker was moved to each of the eight speaker locations, excluding the sub-woofer, and the response was measured at the four microphones. The same speaker was used throughout the test to avoid any possible speaker-to-speaker variation, which would degrade the results. When the source speaker was moved to a new location, the speaker currently occupying that location was moved to the source speaker's previous placement, and another measurement was taken. Once all eight measurements were complete, the car simulator was removed from the anechoic chamber, allowing the second test to be conducted.

This second test makes it possible to compare the response of the speakers with and without the car simulator. By removing the simulator from the chamber, but keeping the microphones and speakers in the same positions within the chamber, another set of data could be acquired and compared with the first test. If the car simulator's response is truly anechoic, then the two data sets would be the same, showing that the simulator does not interfere with the response of the speakers. Therefore, the second test was set up and conducted identically to the first test, but without the car simulator. Each of the microphones and speakers were placed in the chamber at the same X, Y, & Z coordinates, relative to the front left corner of the chamber, as in the previous test.

Once these eight measurements were acquired, the spatial average for both data sets were calculated and compared. The comparison between $50Hz$ and $2500Hz$ is given in Figure 3.5. It is expected to have some interaction due to the car, but should be kept to a minimum. The difference between the two responses, for the most part, is less than $3dB$ which is at the point of being just perceivable [20] and therefore acceptable.

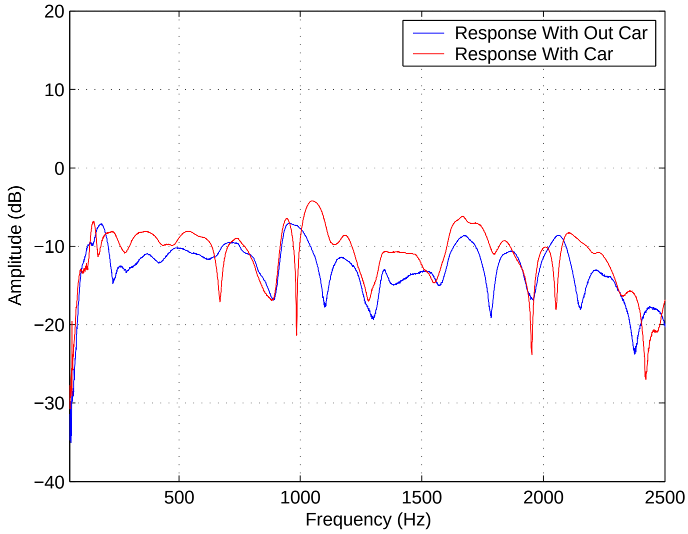

Figure 3.5: Verification of car simulator response.

3.2 Environment Filter Implementation

As discussed in Section 2.2.3 the image source method can be used in conjunction with absorption coefficients in order to simulate frequency dependent reverberation. If this reverberation is then saved as a finite impulse response (FIR) filter, it can be convolved with the signal going to the speakers, resulting in the auralization of a new acoustic environment. For detailed information on calculating the acoustical environment filters, see [39]. The calculation of each virtual acoustic environment filter begins with the calculation of each image source.

The calculation of the image sources, as discussed in Section 2.2.3, requires the dimension of the virtual room to be simulated, as well as both the source position and the listeners position. During the process of calculating the image sources, each source can be represented by a monopole that is a certain distance and relative angle to the listeners position. The length of reverberation contained within each filter is directly related to the number of sources used, and is set to be ≈ 0.2 seconds (8192 coefficients), based on computation limitations. To insure the length of reverberation between filters is the same, regardless of room size, the number of calculated sources is a function of the simulated room size. For a $35ft \times 35ft \times 35ft$ room, a total of 2197 sources are calculated. Since it is impractical to implement each monopole source with a speaker, a simplified setup using only four speakers was used. The four speaker setup was chosen over setups with more speakers, based on the minimalistic approach and that informal listening tests showed a realistic auralization within a substantial area of the showroom was achieved.

To divide the sources among the four speakers, the relative angle to the listener's position of each source is assessed and grouped with other sources in $90°$ increments, creating four quadrants. This grouping process preserves the directionality component of the sources. These four quadrants correspond to the four speakers positions shown in Figure 3.6. Each of the four groups of image sources are then treated as four individual FIR filters and are sent to the corresponding speakers.

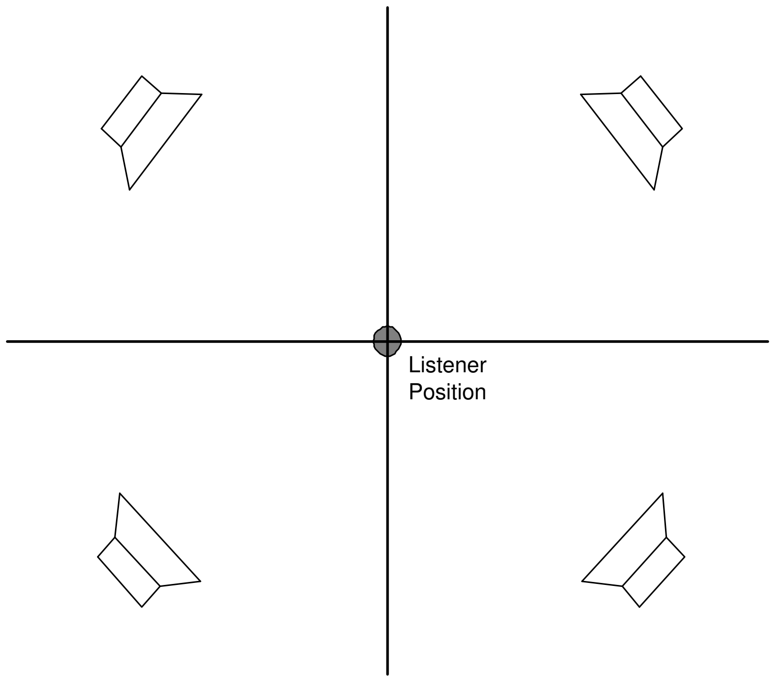

Figure 3.6: Four quadrants of the showroom for determining which speaker each image source is sent to.

Figure 3.7 shows the filter schematic for each of the four speakers in the virtual showroom. As seen in the figure, there is a filter for each of the speakers (H_1 - H_4) corresponding to each of the four quadrants of sources. The other two filters (H_5 and H_6) are crosstalk filters between the front two speakers. The filters for the front speakers (H_1 and H_2), contain both the direct sound and non-direct sound. The direct sound is simulated as originating from "Original Source 1" and "2." While the filters for the rear speakers (H_3 and H_4), only contain non-direct sound. This insures that the sound originates from the front of the showroom, and that precedence is given to the front speakers. This is very important since none of the speakers in the showroom are visible, and it would be confusing for the listener to hear sound radiating from the rear of the room prior to hearing it from the front of the room.

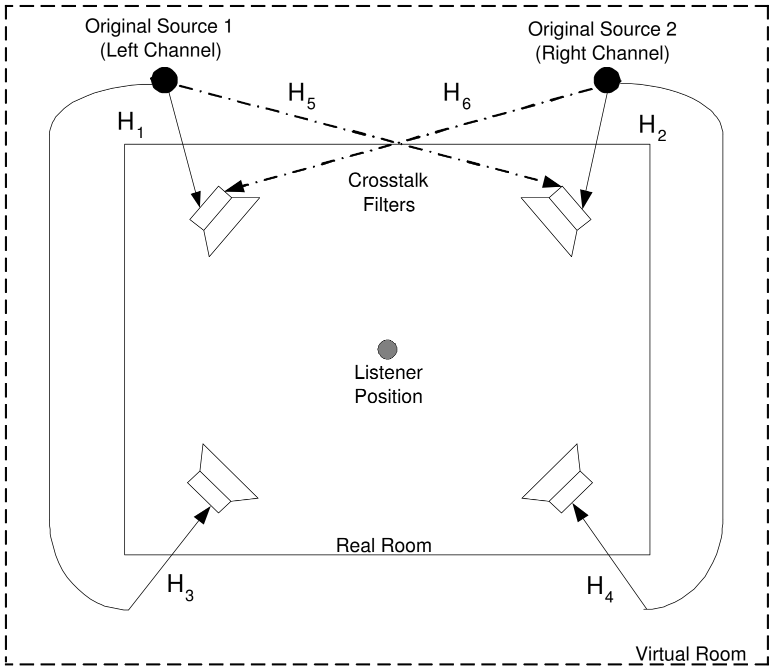

Figure 3.7: Filtering schematic for acoustic environments.

Using LabView, data is streamed from a CD and filtered by the six acoustic environment filters using fast convolution DSP techniques. Music CDs recorded in stereo have data represented in both left and right channels. To preserve the affect of stereo recordings, both the left and right channels of data are respectively filtered with the environment filters corresponding to the left and right speakers in the showroom. The user selects an audio track from the CD, and then chooses a virtual acoustic environment. Switching CD tracks or the virtual acoustic environments can be done at anytime during a simulation. There is a slight delay between the time a new selection is made and when it is being played for auralization. The delay corresponds to the size of the buffer used for reading the data off the CD and, at present, is < 1 second.

For the implementation of the DVD 5.1 and 6.1 surround systems as well as the car simulator, more than four speakers are required. The definitions of the surround sound systems, given in Section 2.1.1, denote the number of speakers needed in the virtual showroom as five or six depending on the system, plus one sub-woofer. It was determined empirically that using six speakers allows the same speaker setup to be used for either system, but with the 5.1 system, the rear center channel is not used. For the car simulator, it was determined empirically that a total of 8 speakers and one sub-woofer are needed. However, depending on the type of car and the placement of the speakers, only certain speakers will be used. This will allow multiple speaker configurations to be simulated with the same setup.

3.3 Room Types and Expected T_{60}

The acoustic functionality of a room depends heavily on its reverberation time. If the reverberation time is not appropriate for the activities conducted in the room, undesired results could occur. For example, if the reverberation time in a cathedral was 0.55 seconds at $500Hz$, it would be very difficult to hear a sermon without amplification. In having such a short reverberation time, the amplitude of the sound decreases so rapidly that intelligible speech may not be heard near the back of the cathedral. An optimal range for reverberation times in churches and cathedrals falls within $1.4-2.6$ seconds at $500Hz$, with cathedrals being on the upper end [40]. This long reverberation time is instrumental in carrying unamplified speech throughout the cathedral. The long T_{60} also gives an organ the "full" sound, which is more pleasurable than hearing an organ with very little reverberation.

An example of where a short reverberation time is desirable would be in a recording or broadcasting studio. Optimal T_{60} times for these environments are between 0.4 and 0.6 seconds at $500Hz$ [40]. This short reverberation time allows music and speech to be recorded, listened to, and broadcast in a more true fidelity context. Examples of reverberation times for concert halls, restaurants, recording studios, and more are given by Salter [25].

The goal of this virtual room simulation is to simulate typical environments where the speakers will reside. Knowing that most people purchasing home stereo equipment will install it in a typical living environment, it was necessary to simulate the acoustics for these typical living environments. To explore different typical living environments, a preliminary set of acoustic filters were created using the image source method (Section 2.2.3) and absorption coefficients for building materials found in most home construction (see [39] for a detailed description on calculating the filters). These filters consisted of all possible perturbations with the following criteria; three room sizes: Small, Medium, and Large; three room shapes: Square, Rectangular (portrait), and Rectangular (landscape); and three acoustic treatments: Dead, Medium, and Live. This resulted in 27 different acoustical environments to select from. The dimensions for the virtual room sizes are given in Table 3.1 and are typical for home-based listening environments. The types of environments are not strictly limited to these parameters, since using different absorption coefficients in conjunction with room dimensions could theoretically yield an endless number of possible environments.

Table 3.1: Dimensions for virtual rooms.

Room Size	Shape	Length m	Width m	Height m
Small	Square	4.57	4.57	2.74
	Rectangular (Portrait)	4.57	7.62	2.74
	Rectangular (Landscape)	7.62	4.57	2.74
Medium	Square	7.62	7.62	2.74
	Rectangular (Portrait)	7.62	10.67	2.74
	Rectangular (Landscape)	10.67	7.62	2.74
Large	Square	10.67	10.67	2.74
	Rectangular (Portrait)	10.67	13.71	2.74
	Rectangular (Landscape)	13.71	10.67	2.74

The acoustic treatments are defined, using absorption coefficients typical for home based listening environments, as Dead: thick pile carpet on sponge rubber underlay for the floor and expanded poly urethane foam on the walls and ceiling; Medium: heavy carpet on the floor, slightly vibrating walls, and a boarded ceiling; and Live: hardwood floor, hard walls, and a boarded ceiling. The absorption coefficients for these treatments are given in Table 3.2 [20]. T_{60} times for each of the 27 rooms are given in Table 3.3.

Table 3.2: Absorption coefficients (α) used for each virtual room.

Room Type	Surface	Frequency (Hz)							
		63	125	250	500	1000	2000	4000	8000
Dead	Floor	0.15	0.45	0.75	0.80	0.85	0.90	0.95	1.00
	Walls	0.30	0.40	0.50	0.60	0.75	0.85	0.90	0.90
	Ceiling	0.30	0.40	0.50	0.60	0.75	0.85	0.90	0.90
Medium	Floor	0.05	0.15	0.25	0.50	0.60	0.70	0.70	0.65
	Walls	0.15	0.01	0.07	0.05	0.04	0.10	0.40	0.80
	Ceiling	0.15	0.15	0.10	0.10	0.30	0.60	0.80	0.90
Live	Floor	0.10	0.15	0.12	0.10	0.07	0.06	0.07	0.07
	Walls	0.15	0.10	0.07	0.05	0.04	0.10	0.40	0.80
	Ceiling	0.15	0.15	0.10	0.10	0.30	0.60	0.80	0.90

Table 3.3: Reverberation time at $500Hz$ for simulated rooms. Verification of these reverberation times will be presented in Section 3.4.

Room Size	Shape	Treatment	T_{60} (seconds)
Small	Square	Dead	0.157
		Medium	0.609
		Live	1.360
	Rectangular (Portrait & Landscape)	Dead	0.199
		Medium	0.630
		Live	1.469
Medium	Square	Dead	0.194
		Medium	0.653
		Live	1.596
	Rectangular (Portrait & Landscape)	Dead	0.205
		Medium	0.662
		Live	1.658
Large	Square	Dead	0.218
		Medium	0.673
		Live	1.724
	Rectangular (Portrait & Landscape)	Dead	0.226
		Medium	0.679
		Live	1.764

3.4 T_{60} Verification Of Calculated Impulse Response

To ensure the accuracy of the acoustical environment filters, each filter can be analyzed in the time domain using the T_{60} technique (Section 2.2.1) prior to implementing them in the showroom. Since each of the filters represent the impulse response of a simulated environment, an analysis of the FIR filter should give the corresponding T_{60} of the room. If the reverberation time of a filter does not match that of the environment it is simulating, then the filter is not an accurate representation of that environment. Sabine's model was used since it is a standard measure for T_{60} times.

For this verification process, filters for a small square room with a medium acoustic treatment and a large square room with a live acoustic treatment were analyzed. To obtain the $60dB$ decay of the impulse response, the slope of approximately the first $20dB$ decay was selected, then the slope was extrapolated to the $60dB$ decay. Figure 3.8 shows the result for the small room with a medium treatment. From the figure, it is clear that the filter created to simulate the reverberation of the small room with a medium acoustic treatment closely matches the reverberation predicted by Sabine's equation (equation 2.12) for the same room.

The reverberation simulated by the filter for the large live room was not as close to the Sabine predictive model as the small room. Figure 3.9 shows that the reverberation time for the filter is approximately 0.6 seconds lower than that of Sabine's model at $500Hz$. It may be possible to attribute this difference to the method used for selecting the slopes for interpolation.

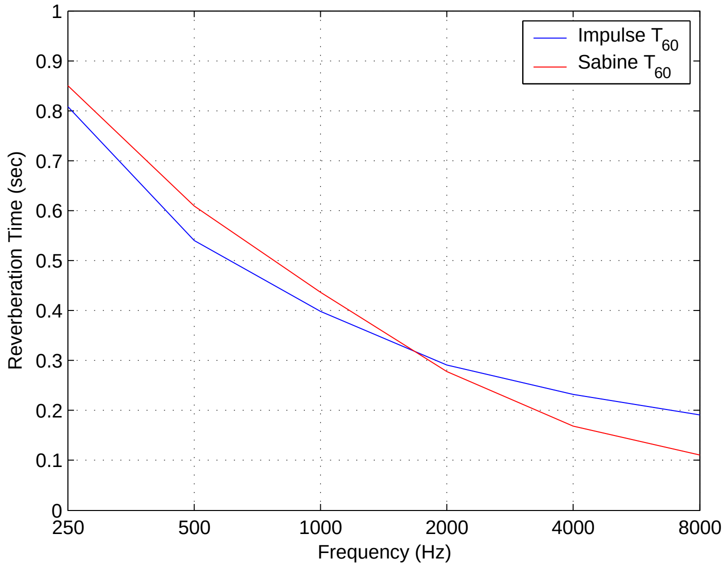

Figure 3.8: Comparison of calculated filter T_{60} and the calculated Sabine T_{60} time for the small medium room.

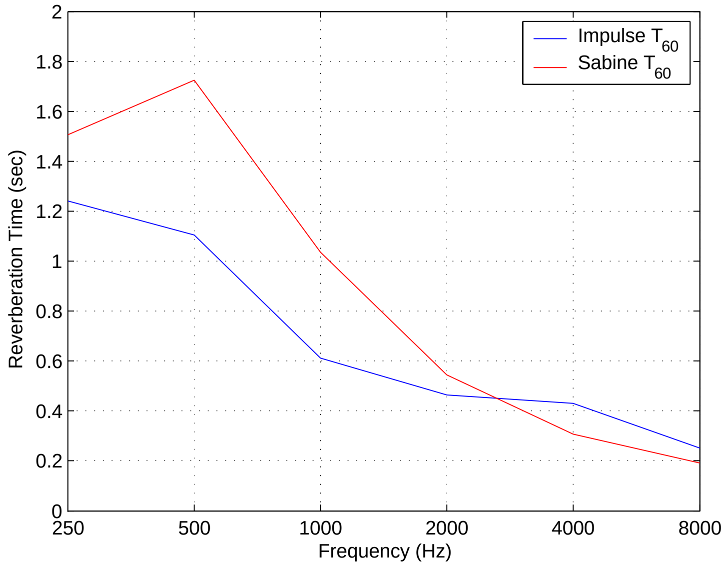

Figure 3.9: Comparison of calculated filter T_{60} and the calculated Sabine T_{60} time for the large live room.

The T_{60} time can change by a significant amount depending on which two points are chosen for interpolating the $60dB$ decrease. It is possible that choosing different points for this measurement would have yielded a time closer to the calculated time. However, the method of picking points was the same for each measurement so as to remain as consistent as possible. Comparing the slope of Sabine's model to the slope of a measured T_{60} at a single frequency shows this issue of sensitivity (Figure 3.10 at $1000Hz$). It can be visualized in Figure 3.10 that the selection of different points could yield the same slope as that of Sabine's. Also, from the figure, the two slopes appear to be very close to the same. However, the difference between them equates to 0.279 seconds for the $60dB$ decay. Therefore, a very small difference in slope leads to a large difference in reverberation time.

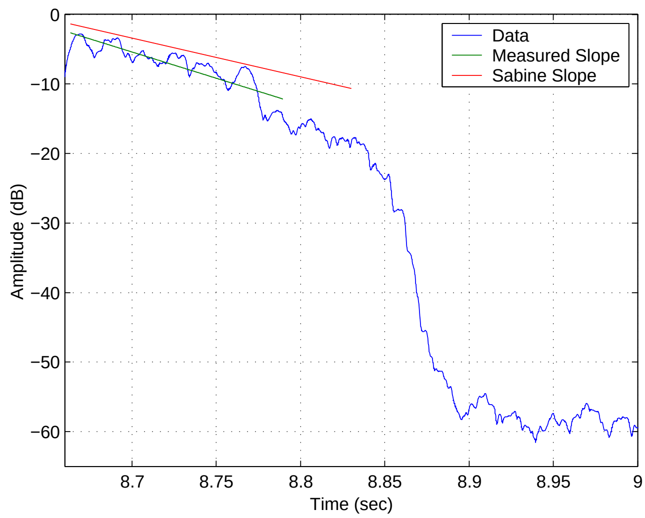

Figure 3.10: Comparison of Sabine T_{60} slope with a measured T_{60} slope at $1000Hz$.

The reverberation simulated by the filters above, for the most part, generated accurate representations of their Sabine predicted model. It is obvious that the small room filter matches the predicted model better than the large room filter, but that doesn't discount the validity of the large room filter. Selecting different points for estimating the reverberation time, as was previously mentioned, could lessen the differences between the filter and the predicted model for the large room.

3.5 T_{60} Verification Of Virtual Environment Implementation

Once the acoustical environment filters are implemented into the showroom system, it is again necessary to check the accuracy of the filters. This time, the method of validating the filters requires measuring the T_{60} time with microphones inside the showroom. These measurements can then be compared with the Sabine model and the T_{60} results of the filter (pre-implementation) obtained in the previous section. T_{60} measurements are typically given from the average times of several microphone placements within an enclosure. This helps ensure that the acoustical properties of the entire room is being taken into account. If one microphone were used at a single location, its placement in the room could dramatically affect the measured reverberation time, depending on whether it were close to an absorbant surface or a reflective surface. That being said, the filters were calculated for one point in the center of the room, the listener's position. The results presented here will use the T_{60} measurements at this location. The T_{60} was also measured at other locations and produced very similar results.

The sound pressure level in an anechoic chamber will decay at a fairly consistent rate once the source is shut off. However, when simulating a different environment with reverberation, the response of the showroom will dominate after the filters are complete. The filters used in the simulation are 8192 coefficients long, which corresponds to approximately 0.2 seconds at a sampling rate of $44.1kHz$. Having such a short filter length makes it difficult to select only the section of data that is resulting from the filter when the measured data is analyzed. To ensure that the data being analyzed is strictly from the filters, an unfiltered reference signal was used to determine the time when the source was turned off. The first $10dB$ fall of the decay process was used to determine the overall decay of $60dB$. This early portion of the decay process is known as the 'early decay time' (EDT) [24]. Using the EDT would also help to ensure that what was being analyzed was from the response of the filters, and not from the showroom.

3.5.1 Anechoic Chamber Verification

To minimize the amount of room interaction, a simulation of a few virtual acoustic environments was performed in an anechoic chamber. A set of filters were created to account for the dimensions of the chamber, as well as the positions of the speakers (sources) and the microphone (receiver). T_{60} measurements were then taken to verify the accuracy of the simulation. The environment filters used for this verification process were selected to be the small square room with medium acoustic treatment and large square room with live acoustic treatment. These were selected so that a comparison could easily be made between the results obtained from this test and the results of Section 3.4.

Figure 3.11 is a comparison of the small-square-medium room T_{60} measurements, the Sabine predicted T_{60} time, and the filter T_{60} time. The difference between the measured time and the predicted time for this room was 0.125 seconds and less, showing a very good match between the virtual and physical (theoretically predicted) environments. It is also apparent that the measured T_{60} time matches very well with that of the filter (pre-implementation). The two curves are almost identical except for frequencies in the $4kHz$ to $8kHz$ range, where they differ by less than 0.075 seconds.

The measurement of the large-square-live room (Figure 3.12) shows similar results but, has a maximum difference of 0.554 seconds at $500Hz$. In addition to the reasons mentioned in Section 3.4, the discrepancies between the measured and Sabine predicted T_{60} times can be attributed to assumptions and limitations of Sabine's equation for reverberation times. One such limitation is the placement of the source. These tests were conducted in an anechoic chamber that has an internal dimension of approximately $11.50ft \times 8.50ft$, in which the sources were placed in the corners of the chamber. With the absorbent material of the anechoic chamber being located so close to the source, the standard Sabine's model can no longer be used because the reverberation time is conditioned by the first reflections on the walls [41].

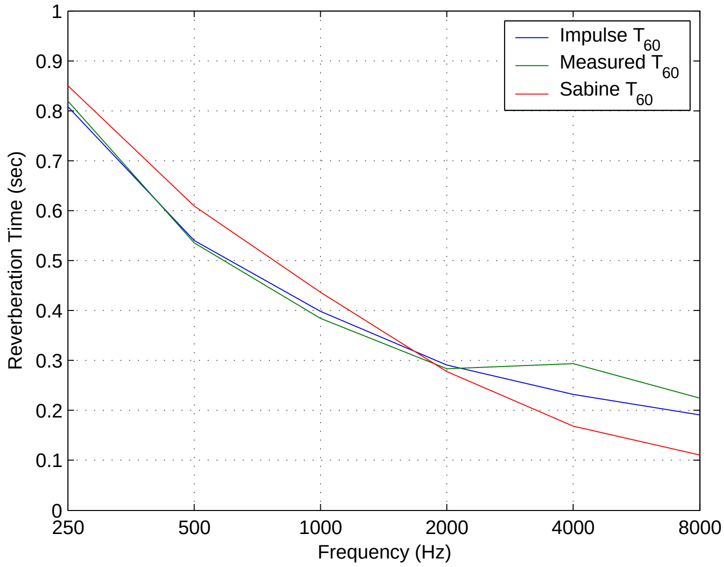

Figure 3.11: Comparison of anechoic measurements and the calculated Sabine T_{60} time for the small medium room.

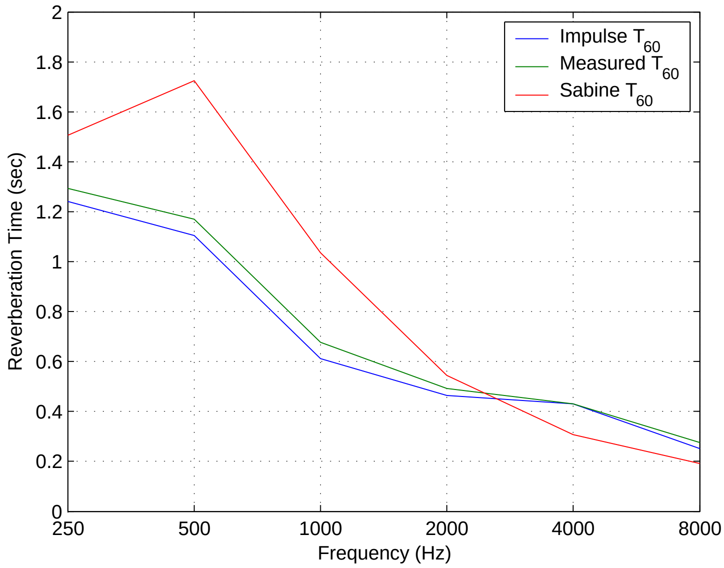

Figure 3.12: Comparison of anechoic measurements and the calculated Sabine T_{60} time for the large live room.

Another assumption made by Sabine's methodology is that the reverberation process is stochastic and ergodic. Being stochastic, all the reflections and room modes are treated as having statistical regularity, allowing them to be described in terms of their probabilities. In a rectangular shaped box, the reverberation is not complex or random enough to describe them in terms of probabilities because of the regularity of the room shape [42]. For ergodic spaces, the statistics for all points in the space are the same as for a single point over time. Given a space with a low T_{60}, there is insufficient time for mixing. (Mixing is the time it takes when there is statistically equal energy in all regions of the space.), and the space is therefore not ergodic [43].

It is very clear that the calculated acoustic filters can be implemented in an anechoic chamber and accurately simulate the desired acoustic environment. It is also apparent that using only four speakers for the auralization process is sufficient enough to produce very accurate results. With the accuracy of these measurements, it could be concluded that the environments created with virtual room acoustic simulation are representative of their physical counterparts. It can also be stated that the environments are being accurately simulated within the showroom according to the design of the filters. In other words, the reverberation contained in the filter is being accurately incorporated into the showroom.

3.5.2 Virtual Showroom Verification

As was shown in the previous section, this simulation of virtual acoustic environments can be implemented in an anechoic chamber. However, this system will be implemented in the virtual showroom and not an anechoic chamber. Since the virtual showroom is not totally anechoic, inverse filters can be applied to compensate for the effects of the room. The inverse filters were generated by measuring the transfer functions between the four speaker positions and the center of the showroom.

The use of the inverse filters made it possible to compare the T_{60} measurements in the showroom and those in the anechoic chamber, with minimal room interaction. Figures 3.13 and 3.14 show these comparisons. The method of testing used was the same method used in the anechoic chamber, and used the same virtual acoustic environment filters. Additional microphones were placed randomly in the showroom to aid in determining the robustness of this system.

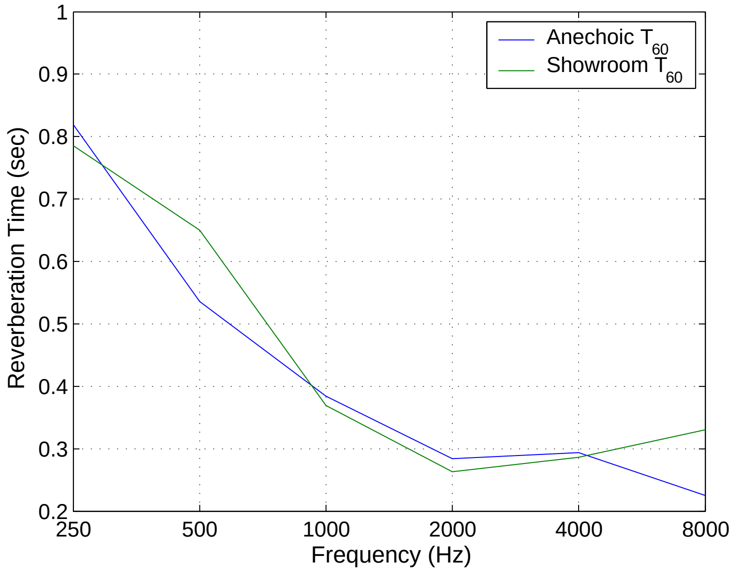

Figure 3.13: Comparison of anechoic and showroom T_{60} time for inverse filter verification of the Small room with medium acoustic treatment.

Figure 3.13 illustrates that the measurements in the showroom are within 0.11 seconds of those taken in the anechoic chamber. For Figure 3.14 the results are close, between $250Hz$ and $2kHz$. The major differences between them is limited to $4kHz$ and $8kHz$ bands and are within 0.24 and 0.32 seconds respectively. These discrepancies can, again, be attributed to the sensitivity issues stated in Section 3.5.1. It can be concluded that the inverse filter sufficiently minimizes the virtual showroom interaction, allowing the simulation to be conducted in the showroom rather than an anechoic chamber.

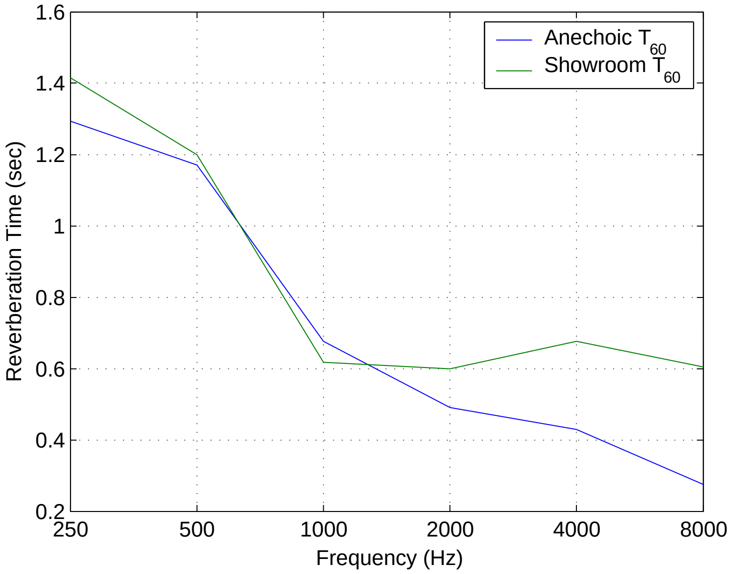

Figure 3.14: Comparison of anechoic and showroom T_{60} time for inverse filter verification of the Large room with live acoustic treatment.

3.6 Conclusion

This chapter focused on simulating acoustic environments for the virtual acoustic showroom. The showroom prototype had a verified anechoic acoustic treatment above $500Hz$, making it suitable for simulating acoustic environments. Results for the small room showed nearly the same T_{60} curve between the filter and the anechoic implementation. The T_{60} curve between the anechoic and showroom implementation also showed a good match, but with some divergence in the showroom nearing the $8kHz$ frequency range. For the large room, again, the T_{60} curve between the filter and the anechoic implementation is nearly the same. The T_{60} curve between the anechoic and showroom implementations diverge from each other in the higher frequency range ($4kHz - 8kHz$). Because both the small and large rooms show this divergence at higher frequencies, it may indicate that the reverberation is being over-estimated at the higher frequencies. This may require that some modifications be made to the way the filters are calculated. However, with the results presented, it is clear that the environment filters that are calculated using the image source method and absorption coefficients can be implemented into the virtual acoustic showroom to accurately simulate a range of acoustic environments for users to select from.

Chapter 4

Speaker Compensation

One of the major goals of this research is to accurately make one speaker sound like another speaker. To accomplish this, a method of filtering called inverse filtering was employed to pre-filter the signal sent to the speaker. For accurate mimic filtering, the frequency response function for both the reference speaker and the original speaker has to be measured (Section 4.1). Inverse filters were created and implemented for speaker mimicking as well as for minimizing the acoustic interaction of the virtual showroom with the speaker responses (Section 4.2).

4.1 Speaker Response Testing

During the testing phase of this project, multiple speakers were tested from different price ranges and size categories. In order to appropriately filter a speaker, it is imperative to know the frequency response function of that speaker. The frequency response function (FRF) is a measure of a speaker's magnitude over a range of frequencies. To obtain this response, white noise (random noise with all frequency content having the same amplitude) is sent to the speaker and the speaker output (radiated acoustic field) is measured with a microphone. The result can then be analyzed and filtered. For this measurement, each speaker was tested with white noise having a frequency content between $20Hz$ and $20kHz$ and a sampling frequency of $44.1kHz$ and 65536 data points per sample. The frequency content of the white noise was limited to the maximum audible range of humans [20]. Each speaker was analyzed within the same range, but did not extend below $50Hz$ due to both the lack of output of most speakers at low frequencies, and the inability to accurately measure the speaker's dynamics at those frequencies.

In addition to filtering, the response of each speaker was needed so that a reference set of speakers could be selected and used to mimic all of the other speakers. This selection was based on the set of speakers which had the most uniform amplitude response over the range of frequencies in interest. A uniform response is considered to be one where the variation in amplitude over the entire frequency range is $< 5 - 10dB$. If the response of a speaker was not uniform, it would be difficult for that speaker to mimic another speaker with a more uniform response. For example, if the reference speaker had a zero at $2kHz$ (a drop of $20dB$ or more in the transfer function over a narrow band of frequencies), it would not be possible to create a response at that frequency and mimic a speaker that did not have a zero there. Trying to create a response at a zero would require extra compensation, and in the process may overdrive the speaker at other frequencies, causing damage, non linearities, and other problems.

4.1.1 Speaker Testing in Anechoic Chamber

As mentioned in Section 2.2, measuring a speaker in an anechoic chamber produces a response that lacks almost all environmental context. Therefore, these measurements are considered to be the true response of the speaker down to the lower limit of the chamber. The wedges in the Durham anechoic chamber facility are $0.762m$ in depth. Using the quarter wavelength designation for the lower limit would make the chamber theoretically capable of absorbing sound waves down to $112Hz$.

The first set of tests taken in the anechoic chamber had three microphones placed in an arc with one being on axis, normal to the speaker front, and the others at $22.5°$ and $45°$ off axis. Assuming the directivity of a speaker to be symmetric about its centerline allowed for testing only one side off axis (This assumption was later verified using measurements in the virtual showroom.). The radius of the microphone arc was $1.83m$, with the center being located at the speaker cone to accommodate a listener who would sit $1.83m$ away from the speaker in the virtual showroom. The planned internal dimensions of the virtual showroom was set at $3.66m$, with a designated listener position in the center of the room. The speakers were placed on a suspended platform so the speaker could be raised or lowered, placing the center of the cone at the same height, relative to the floor and microphone for each speaker. Using these measurements, the response for each speaker was compared and a reference set was selected. The response for one of the chosen reference speakers is shown in Figure 4.1.

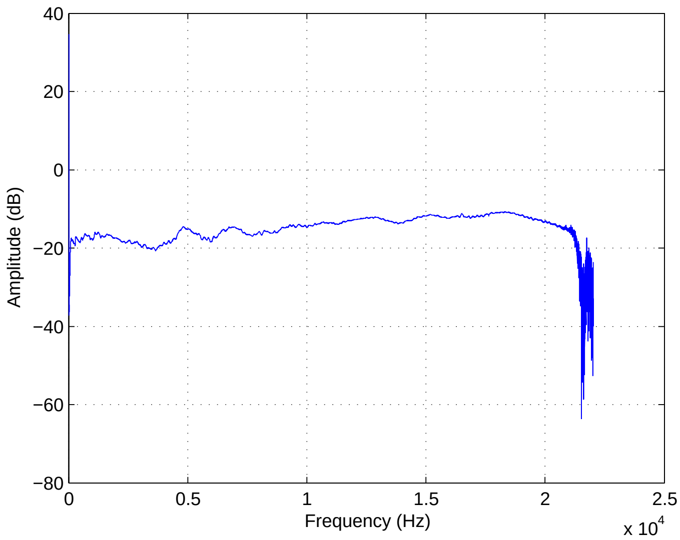

Figure 4.1: Frequency response of reference speaker.

As seen in Figure 4.1, above $20kHz$, the response quickly drops off and is comprised of noise. This is simply noise created from calculating the frequency response function and is not considered in the mimicking process. Since this measurement was taken in the anechoic chamber, low frequencies drop off and are much lower in magnitude than higher frequencies. This is shown in Figure 4.2, with a magnified view of the reference frequency response function, where below $100Hz$, the response begins to drop. It is clear from these two figures that the response of this speaker is relatively uniform and does not contain any zeros, therefore meeting the selection criteria for a reference speaker.

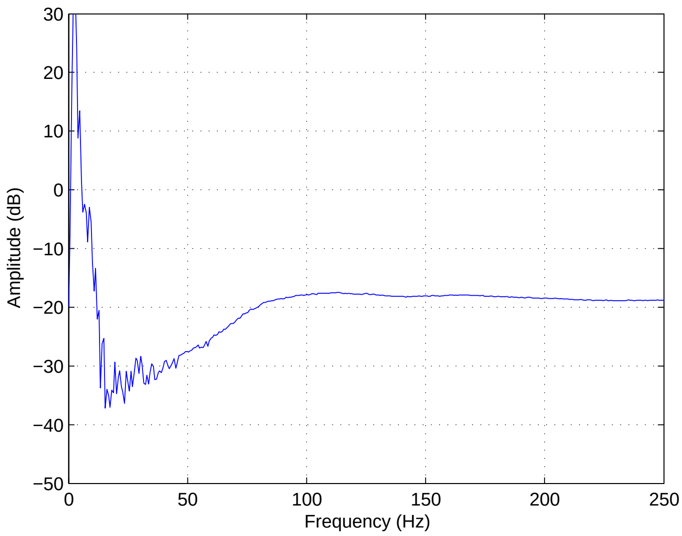

Figure 4.2: Low end frequency response of reference speaker.

It was later necessary to take a second set of measurements in the anechoic chamber to improve the mimicking process (Section 4.1.2). For this measurement a single microphone was placed on axis 0.61m from the cone of the speaker. The microphone was also raised or lowered to be at the same height as the center of the mid-range cone.

The two tests described above were used for all speaker types mentioned in Section 2.1.1, except the car speakers. To test the car speakers it was necessary to mount them in a baffle. The baffle was used for two main reasons: first, it allows the true response of the speaker to be measured without the effect of a backing volume due to an enclosure; and second, it prevents low frequency cancellation due to sound radiating around the back of the speaker and pushing on the cone. To measure the car speakers, each speaker was placed in the center of the baffle and a single microphone was placed on axis at a distance of 0.61m from the cone. After taking these measurements and selecting a reference set of speakers for both the home stereo portion and the car portion of this project, it was necessary to measure the speakers in the virtual showroom.

4.1.2 Speaker Testing in Virtual Showroom

The speaker testing conducted in the virtual showroom was for the purpose of determining a suitable method of measuring speaker responses to be used in creating accurate mimic filters. Several measurement techniques were investigated including: 1 - a single measurement location in the center of the room, 2 - a sphere of measurement locations (in the center of the room) and multiple speakers, and 3 - a single measurement location near the speaker.

The measurements obtained using technique 3 provided the best results for the purpose of creating mimic filters, and was subsequently chosen as the technique to use. Since the microphone for this technique was placed $0.61m$ from the speaker, it was not located in a position corresponding to the anti-node for the first axial mode (1,0,0) at $49Hz$, as was technique 1. This greatly reduced room mode interaction with the response of the speakers. This method was also very simplistic, using only one microphone and one speaker. The computational time it took to analyze the data and create a mimic filter was far less than what would have been needed for technique 2.

While investigating the potential of technique 2, the number of speakers needed to accurately mimic one speaker at each of the measurement locations was in question. If only a few speakers would be required, then the technique may be a viable option. To make this determination, singular value decomposition (SVD) was used with data obtained from the technique 2 experimental setup. Singular value decomposition takes one matrix and separates it into three matrices

$$[H] = [U][\Sigma][V]^T \tag{4.1}$$

where H is an $m \times n$ matrix, in this case, of transfer functions for each of the m speakers to each of the n measurement locations, U is an $m \times m$ unitary matrix of left singular vectors, Σ is an $m \times n$ diagonal matrix of singular values, and V is an $n \times n$ unitary matrix of right singular vectors [44]. After applying the singular value decomposition to H, the singular values are plotted as a function of frequency. When plotting singular values, the more important a singular value is, the closer to $0dB$ it will be. As the amplitude of a singular value decreases, so does its importance. For this analysis a $-20dB$ cutoff was used to distinguish important versus unimportant singular values. In other words, when a singular value drops $20dB$ below the first important value, it is no longer considered to be important.

The experimental setup for this technique consisted of an array of 5 speakers and 5 microphones. The speakers were arranged in a row where the center speaker was in the same location as the speaker to be mimicked. The microphones were placed in an arc with a radius of $0.5m$ and were evenly spaced from $0°$ – normal to the axis, $\pm 36°$, and $\pm 72°$ as shown in Figure 4.3. The axis was located in the center of the room at the listener's location

Figure 4.3: Spherical test setup with 5 microphones.

and the center line ($0°$), was normal to the speaker cone of the middle speaker. The height of this array was set so that the center line was at the same height as the center of the speaker cone. Rotating the array around the vertical axis in eight incremental steps of $45°$ produced a sphere of 40 points centered around a seated listener's head.

Each speaker was played individually at each of the spherical locations. The speaker in the center was measured first (Figure 4.4) and then replaced by a speaker that was to be used with the other four in the mimicking process. In creating the sphere of points, the microphones near the center line covered more area than those close to the vertical axis, making it necessary to weigh their contribution according to how much area they covered.

Figure 4.4: Minimization of points with multiple speakers and the spherical test rig.

Applying the SVD method, the matrix H was a 5×40 matrix of transfer functions from each of the 5 speakers to each of the 40 measurement locations, and was decomposed into its singular values. Figure 4.5 is a plot of the 5 singular values as a function of frequency. It is clear from this figure that up to about $500Hz$, only the first three singular values are important, using the $20dB$ cutoff. If the frequency requirement for the mimic was limited to $500Hz$ and lower, it may have been a viable option to use multiple speakers. However,

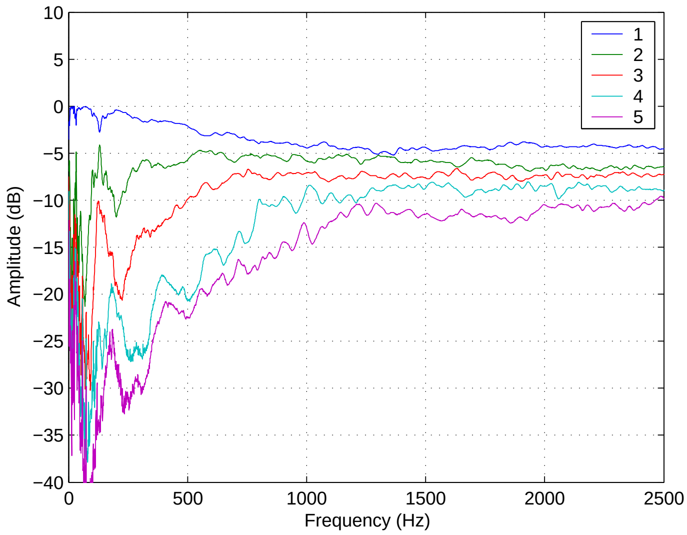

Figure 4.5: Singular values as a function of frequency.

from about $1000Hz$ up to the frequency limit of $20kHz$, all of the singular values are near the same value and are subsequently considered important. This suggests that there are many more degrees of freedom that are not being captured by using only five speakers. In order to recreate all possible sound-pressure combinations of the 40 microphone positions, a substantial number of additional speakers would be required. For this reason, simpler measuring techniques were evaluated.

As with technique 1, technique 3 used a single microphone placed on axis with the speaker cone. The difference between these two techniques was their distance from the cone. As mentioned earlier, the measurement location for technique 1 was in the center of the room where the listener would be located, while the measurement location for technique 3 was $0.61m$ from the cone. The single microphone measurement provided for a good mimic at the measurement location. However, by moving the measurement location very close to the speaker, the area affected by the mimic was increased by a significant amount since the microphone is in the near field, and everything that is radiated, within certain limits, beyond the measurement location is mimicked. This larger area now encompasses the listener's position, where as before in technique 1, the listener's position was on the very edge, or at the beginning, of the affected mimic area. This affected area was within about $\pm 30°$ of the axis normal to the speaker cone. Not only does the mimic area cover a larger portion of the room, but it accomplishes this using only one speaker, thereby greatly reducing potential processing requirements. Because of the results obtained from this measurement location, all subsequent measurements were conducted using technique 3, and having a position of $0.61m$ from the speaker, normal to the cone. Details on the mimicking process is discussed in Section 4.2.

4.2 Speaker Mimicking and Filtering

The method of mimicking one speaker with another speaker requires the use of the frequency response functions of both the reference and original speakers. As stated earlier, placing the microphone in the center of the room resulted in a large interaction of the room's acoustic properties, causing very poor results for the mimics. The method of measuring multiple points for multiple speakers would have required more speakers than was acceptable, especially since a goal of the project was to use a minimal number of sources with a minimal amount of computation time. Placing a microphone on axis in the near field produced excellent mimic results while using only one source and was chosen as the measuring technique for the mimicking process.

The measurements used for preliminary testing and for determining the mimic filters were taken in the virtual showroom, although it will be shown that anechoic chamber measurements can, and should, be used for creating the mimic filters. Using the virtual showroom measurements would mean that every speaker to be mimicked would have to be measured in every virtual showroom built. Each time a new speaker is added to the list, a measurement of it would first need to be taken in the virtual showroom. Having multiple showrooms would mean multiple measurements, resulting in an inefficient method of cataloging speaker responses for filters. Anechoic measurements would allow each speaker's response to be cataloged and distributed to any showroom desired.

4.2.1 Speaker Inverse Filtering

The method of inverse filtering (Section 2.3) is used in creating the mimic filters as outlined by this section. By taking the inverse transfer function of the original speaker response, and sending it to that speaker, the original response is theoretically canceled out. If the response of another speaker is included with the inverse of the original response, the resulting response is of the second speaker. In other words, by "removing" the response of the reference speaker, the response of a secondary speaker can be added in. This is the premise for creating the filters used to mimic speakers. An analogy of this process could be given using three sine waves, two of which are 180° out of phase with each other and the third is arbitrary. The net result of combining all three signals would simply be the third arbitrary sine wave, since the two signals that are 180° out of phase with each other would completely cancel each other out.

It can be shown by using equations 2.29 and 2.31 that the reference speaker B, with response H_B can be given a signal $H_{mimic} = H_A/H_B$, resulting in a response of H_A, where H_A is the frequency response of the original speaker. This process will be given the notation $B \rightarrow A$, where speaker B is filtered such that its response is equivalent to that of speaker A.

To create the filter H_{mimic}, frequency response functions of both the reference speaker and the original speaker are needed. This is where the measurements of Section 4.1 are used. Both measurements are analyzed in MATLAB in the frequency and time domains. In the frequency domain, the original speaker response is divided by the reference speaker response, giving the mimic response. A complex digital filter frequency response is then calculated and multiplied to the mimic response to remove computational noise from the mimic filter. The length of this filter is the length of the original data measured, and in this case it is 32768 coefficients long. Although accurate, its length would require a sizable computational time when convolving it with a signal being sent to the speaker. Therefore, it is desirable to reduce the filter length and shorten the computational time.

In the first attempts to reduce the number of coefficients, the resulting filter did not produce an accurate mimic. This was mainly because the number of coefficients needed to represent the response below $60Hz$ is very high and the data below $60Hz$ is not accurate. By estimating the roll off below $60Hz$, the number of coefficients needed is greatly reduced. It was found that reducing the number of coefficients in the filter from 32768 points down to 8192 points still produced an accurate inverse filter. This reduction also reduced computation requirements and lag time in the system. Figure 4.6 shows the roll-off of an inverse filter and the curve fit used to make the filter more efficient in its use of coefficients.

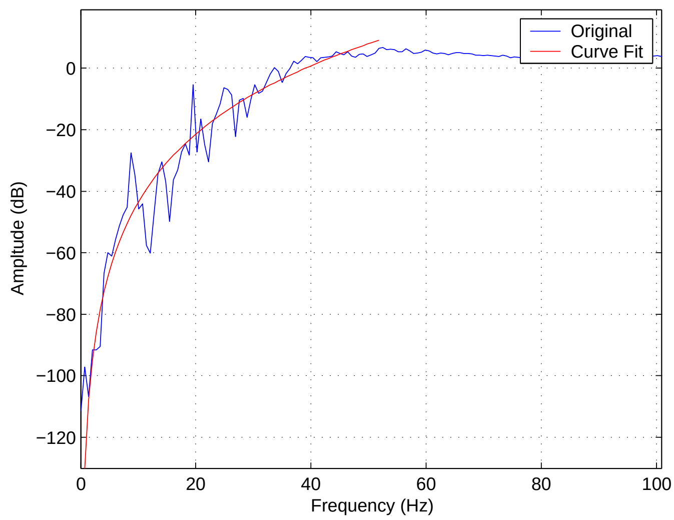

Figure 4.6: Curve fit of inverse filter roll off for reducing the number of filter coefficients.

After the number of coefficients are reduced, the inverse filter is converted into the time domain and saved as a finite impulse response filter. Using LabView, data is read from a CD and is convolved with the finite impulse response filter, as discussed in Section 3.2. This signal is then sent to the reference speaker.

For verification purposes, the transfer function $H_{B \to A}$ was measured and compared with H_A. Figures 4.7 and 4.8 show this comparison, where the original speaker, H_A, is the blue line, $H_{B \to A}$ is the green line, and the difference between those two measurements is the red line. The reference speaker used for these measurements is the same one shown in

Figure 4.1.

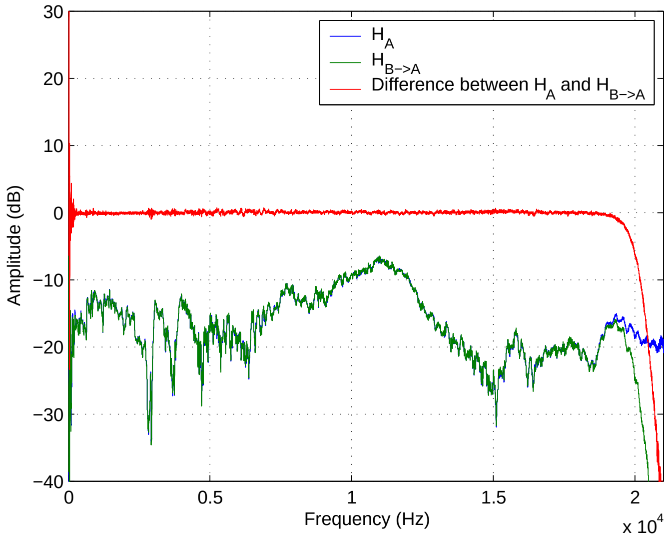

Figure 4.7: Verification of a speaker mimic filter created using virtual showroom measurements. H_A is the original speaker and $H_{B\to A}$ is the reference speaker mimicking the original speaker.

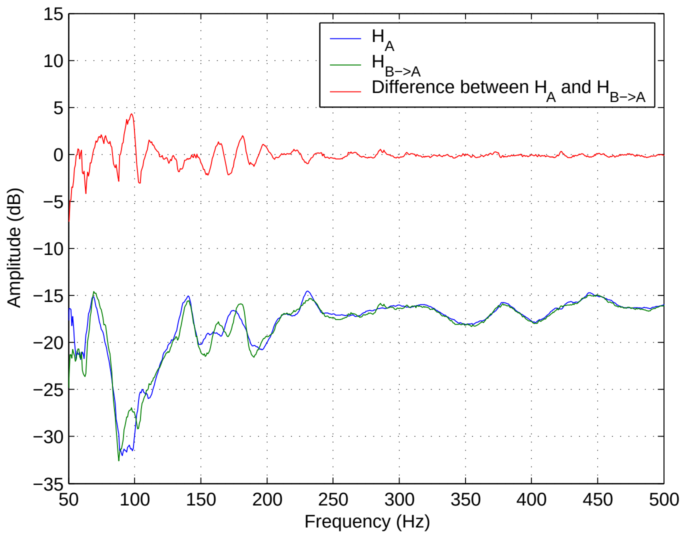

Figure 4.8: Verification for the low end of the speaker mimic filter created using virtual showroom measurements.

For frequency range of $50Hz$ to $20kHz$, the largest difference between $H_{B\rightarrow A}$ and H_A is shown to be less than $5dB$ at $98Hz$, and at frequencies above $250Hz$ the difference is limited to approximately $0.5dB$. According to Beranek [45] the minimum perceptible change in sound pressure level that a person is able to detect is about $1dB$ for a pure tone between $50Hz$ and $10kHz$, if the tone level is greater than $50dB$ above the threshold for that tone. If the level is less than $40dB$, then changes of about $3dB$ are necessary in order to be audible. Given that for broad band noise (such as white noise), the necessary change in sound pressure level is greater than for pure tones, the differences shown in the figures should not be audible. Therefore, it is evident that the mimicking process using the reduced number of coefficients works with a high degree of accuracy.

4.2.2 Spatial Position of Speakers

As previously mentioned, using anechoic measurements is an important part of making it feasible for multiple virtual showrooms to exist. Therefore, a set of mimic filters were created using anechoic measurements in order to determine if they will produce accurate results when implemented in the virtual showroom. It would be reasonable to assume that if a speaker accurately mimics another speaker in an anechoic chamber, then the same filter could be used in the virtual showroom and maintain a high degree of accuracy. In verifying the accuracy of the filters in the virtual showroom, two fundamental issues were discovered, causing an increase in amplitude below $250Hz$ when anechoic filters were used. First, the spatial position of speakers within an enclosure greatly effects the mimic response obtained using anechoic filters. And second, the physical placement of the speakers in the virtual showroom is critical during the verification process.

To investigate the increase in amplitude for the low frequencies, a floor standing speaker and a bookshelf speaker were used as the reference for an anechoic mimic filter. The sizes of speaker drivers in each of the enclosures is as follows: the floor standing speaker had a $3.81cm$ tweeter, $8.89cm$ mid-range driver, $15.24cm$ mid-bass driver, and a $30.48cm$ sub-woofer; the bookshelf speaker had a $3.81cm$ tweeter and a $15.24cm$ mid-bass driver; and the original speaker had a 2.54cm tweeter and a 13.34cm mid-bass driver. The mimic filters were created using anechoic measurements for both the reference speakers and the original speaker. To ensure the filters were properly calculated, $H_{B \to A}$ was measured in the anechoic chamber for both reference speakers and then compared to H_A, which was also measured in the anechoic chamber. After confirming that both filters accurately mimicked the original speaker in the anechoic chamber, the reference speakers were individually placed in the virtual showroom. During the verification process in the virtual showroom, the original speaker was placed on top of the reference speaker and the microphone was appropriately raised or lowered to be at the height of the center of the mid bass driver for each speaker.

A second experimental setup in the virtual showroom was used to gather more data and insight into the amplitude increase over the low frequency range. This second setup was similar to that of the one just discussed, except for the placement of the original speaker. For this test, the original speaker was placed in the same location as the reference speaker during the verification process. In other words, H_A was measured in the virtual showroom, the original speaker was removed, the reference speaker was positioned in its location, and then $H_{B \to A}$ was measured. In doing this, the microphone position was held stationary and the height of the mid bass driver for each speaker was set to the height of the microphone.

The results from the first experimental set up are displayed in Figure 4.9. It is clear from the figure that the mimic using the floor standing speaker is significantly greater in amplitude than the original speaker, prompting the experiments discussed in this section. The response of the bookshelf mimic closely matches the response of the original speaker, suggesting that the sub-woofer in the floor standing speaker should not be used for mimicking bookshelf speakers.

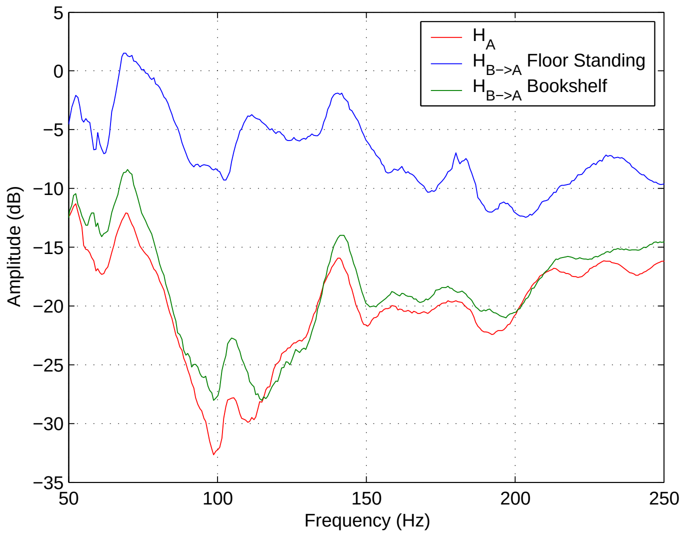

Figure 4.9: Comparison between floor standing and bookshelf speakers for use as the reference speaker using anechoic mimic filters. $H_{B\to A}$ Floor Standing, $H_{B\to A}$ Bookshelf and H_A were measured in the virtual showroom.

Even though using the bookshelf speaker as the reference clearly gave much more accurate results over the floor standing speaker, the results were still not as accurate as those presented in Figure 4.8. The fact that the virtual showroom mimic filters were still more accurate than the anechoic mimic filters prompted the second experimental setup. Figure 4.10 shows the results from this test. The differences between $H_{B\to A}$ and H_A were further reduced from the previous experiment by placing the two speakers in the same position in the virtual showroom for verification. This helps to confirm that speaker driver placement in an enclosure and placement relative to room location does change the accuracy of anechoic mimic filters.

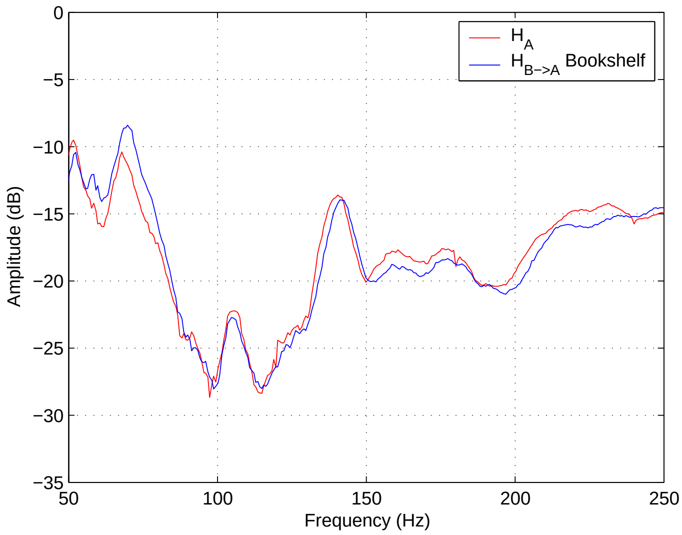

Figure 4.10: Comparison of an anechoic mimic filter ($B \rightarrow A$) and original speaker placed in the same location in the virtual showroom. $H_{B \rightarrow A}$ Bookshelf and H_A were measured in the virtual showroom for verification.

It can be gathered from these two experiments that the virtual showroom mimic filters take the interaction of the room, speaker placement, and speaker size into consideration, while anechoic mimic filters cannot predict that interaction. For example, it is clear that the floor standing speaker accurately mimics the bookshelf speaker using a virtual showroom mimic filter (Figure 4.8), while the anechoic mimic filter increases the bass response by over $15dB$ at $150Hz$ (Figure 4.9). The explanation for this can be found by observing the sub-woofer location of the floor standing speaker. In the virtual showroom this sub-woofer was located in a lower corner of the room. This placement would automatically increase its radiated sound power level by $3dB$ over the rest of the speakers in the enclosure [21]. It is important to note that the other speakers in that enclosure were located in the middle of the $z - axis$ of the room, and would interact with some of the acoustic modes of the showroom, while the original speaker was located on top of the reference speaker, not corresponding to room mode excitation positions.

The implications of this are quite important. First, transfer functions of speakers taken in an anechoic chamber can be used in creating accurate mimic filters if; the original and reference speakers are close to the same size, and the two speakers are placed in the same location, in the showroom, for the verification process. Secondly, a floor standing speaker produces an inaccurate increase in low frequency response. Therefore, a bookshelf speaker with a separate sub-woofer would be the best reference speaker setup. And third, the rooms dynamics play an important role in the mimicking process and should be addressed (see Section 4.2.3).

4.2.3 Removal of Room Response

In room acoustics, the response of a room can be very influential on the quality of sound in that room. Since the virtual showroom was constructed to be anechoic, which was shown to be true for frequencies $\geq 500Hz$ (see Section 3.1.1), the biggest contributor to the response of the virtual showroom are the acoustic modes. By removing the response of the showroom that is dictated by its acoustic modes, the effects that those modes have on anechoic mimic filters, as discussed in Section 4.2.2, could theoretically be eliminated. The concept of removing a speaker's response by using an inverse filter can be expanded to include removing the response of the virtual showroom.

To determine if the modes of the showroom modify the response of the speakers in the room, several bookshelf speakers were placed in the same position and their frequency response functions were measured. Using several different speakers will help ensure that the portions of the response functions that are identified as corresponding to the room's acoustic modes are truly from the showroom and not part of the speaker's response. Figure 4.11 shows two different speakers measured in the same location with no filtering, where each of the peaks labeled correspond to the interaction of a room mode with the speaker response. It can be seen that the modes of the room modify the response of the speakers at these lower frequencies.

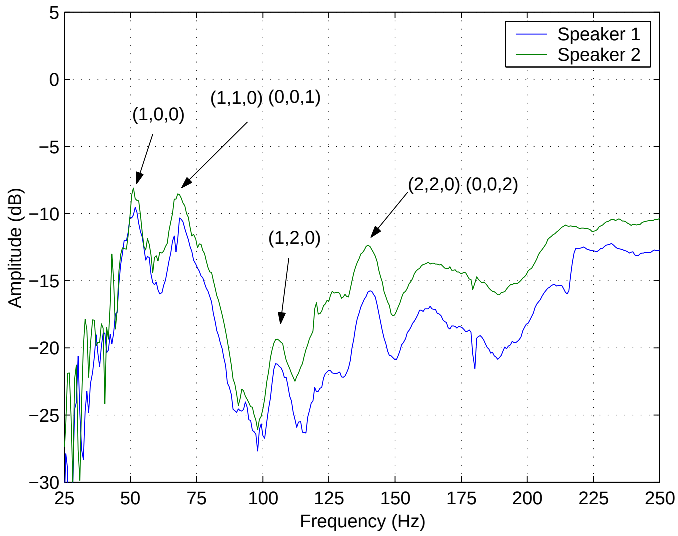

Figure 4.11: Virtual showroom acoustic modal interaction.

If the acoustic modes of the virtual showroom did not modify the response of the speakers, then their frequency response functions would not be similar. Also, the modes that are excited by the location of the speakers were calculated using equation 2.18 and found to match the frequencies corresponding to the peak amplitude response of both speakers. The calculated modal frequencies and measured modal frequencies are given in Table 4.1. The measured modal frequencies were within 5% of the calculated modal frequencies. The small differences between the calculated and measured modal frequencies was anticipated because of the construction of the virtual showroom walls. The equation used to calculate these frequencies depends on the distance between the walls of the room. The absolute distance between the walls is hard to precisely define acoustically, due to the absorption of the insulation.

Table 4.1: Comparison of calculated modes to measured results.Measured frequencies are within 5% of the calculated frequencies.

n_x	n_y	n_z	Calculated Frequency (Hz)	Measured Frequency (Hz)
0	1	0	48.9	51
1	0	0	48.9	51
1	1	0	69.2	68.5
0	0	1	70.3	68.5
1	2	0	109.4	105
2	1	0	109.4	105
2	2	0	138.4	140
0	0	2	140.67	140

To remove the room interaction, the concept of the inverse filter was applied to the virtual showroom's acoustic dynamics [46]. Creating the room inverse filter is similar to removing the response of the reference speaker for mimicking. The reference speaker's frequency response function was measured in the anechoic chamber, and then again in the showroom. Care was taken to ensure the location of the microphone was the same for both measurements, relative to the speaker. Knowing that the true response of the speaker was measured in the anechoic chamber, deviations from that response, when the response is measured in the showroom, will show the interaction of the virtual showroom's acoustic properties. Therefore, dividing the response measured in the chamber by the response in the showroom will create an inverse filter of the virtual showroom. The filter is then combined with the signal going to the reference speaker and measured for verification.

Figure 4.12 presents a comparison between response measurements for verifying the virtual showroom inverse filter. The comparisons use the frequency response function of the reference speaker, where H_A is the response in the anechoic chamber, H_B is the response in the virtual showroom, and $H_{B \to A}$ is the response in the showroom with the inverse filter. As can be seen in the figure, the amplitude of $H_{B \to A}$ was reduced down to the approximate amplitude of H_A. That is, between $50Hz$ and $250Hz$, the amplitude was reduced by $10 - 20dB$. At the frequencies listed in Table 4.1, $H_{B \to A}$ has corresponding peaks in amplitude response which suggest there is still modal content present. However, the amplitude of those peaks are $< 3 - 5dB$ of the anechoic response, above $75Hz$, and are not audible. It can be concluded that the response of the virtual showroom was effectively removed.

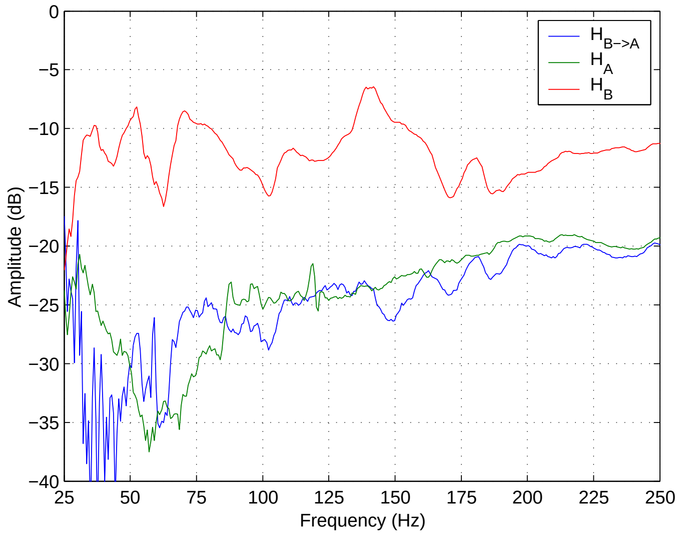

Figure 4.12: Virtual showroom response removal using an inverse filter. H_A is the anechoic response, H_B is the virtual showroom response, and $H_{B\rightarrow A}$ is the virtual showroom response using the inverse filter.

Figure 4.13 is a schematic depicting how the room inverse filter is implemented. The room inverse filter and the mimic filter are multiplied together in the frequency domain and are then transformed into the time domain. Once in the time domain, the combined filter signal is convolved with music from a CD. The new signal, consisting of both filters and the music, is then sent to the reference speaker, resulting in the reference speaker mimicking the original, removing the response of the virtual showroom, and the music from the CD.

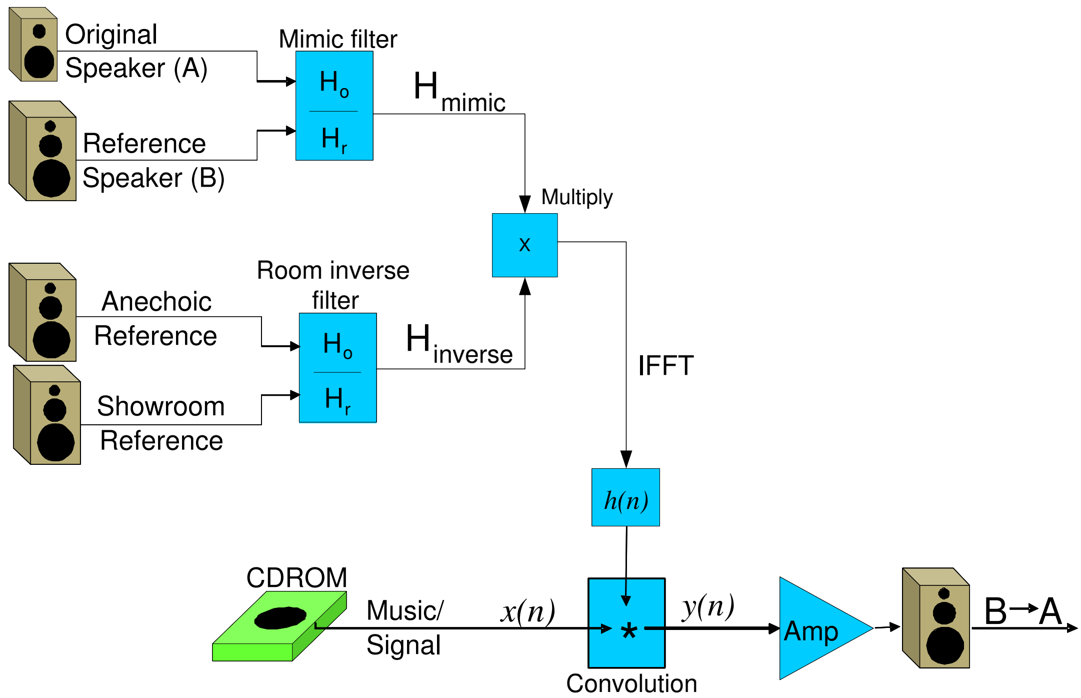

Figure 4.13: Mimic filter and room inverse filter schematic.

In summary, measuring the reference speaker in both an anechoic chamber and the showroom would allow for the room response to be removed before applying mimic filters. The use of the inverse filter will increase the accuracy of anechoic mimic filters used in the virtual showroom.

4.3 Conclusion

The focus of this chapter was on the speaker compensations used in creating the virtual acoustic showroom system. The chapter detailed the testing methods for measuring the frequency response function of a speaker, how to use that response in creating an inverse filter for the mimicking process, and for removing the response of the virtual showroom. It was shown that a speaker could be accurately mimicked by another speaker using anechoic mimic filters under the following conditions: The measurement location must be close enough to the speaker (near field) so that the area affected by the mimic encompasses the listener's position and is not at an anti-node of the virtual showroom. The number of coefficients in the inverse filter can be reduced by estimating the roll-off of the low end of the inverse filter's response by a curve fit. The reference and original speakers must have similar spatial position of speaker drivers within the enclosure, and be co-located in the showroom for verification. Finally, the interaction of the showroom must be reduced by using an inverse filter of the room. The accuracy of the mimicking process was determined experimentally, but still in question is whether a listener will agree with these results or not. This will be investigated and discussed in chapter 5.

Chapter 5

Psychological Testing

Chapters 3 and 4 showed empirically the ability to accurately create different virtual acoustic environments and mimic one speaker with another. Those measurements however, do not indicate whether or not the user will perceive them as they are intended. This chapter deals with the psychological human interaction aspect. Section 5.1 discusses the testing methods related to changing the acoustic environments where the results from the test are shown to be less conclusive for some environments than for other environments. Section 5.2 discusses the testing methods related to speaker mimicking and it will be shown that the mimicking process is perceived as being accurate.

5.1 Room Response Verification

For clarity, the following conventions will be used throughout the rest of this chapter:

- Virtual acoustic showroom system - Refers to the entire virtual acoustic system, which includes the virtual showroom, the computer system for mimicking speakers, the computer system for simulating acoustic environments, and the speakers. It is a reference to the complete "package" that would be installed in a store.

- Virtual showroom - Refers to the prototype of the showroom that was built for testing according to Section 3.1. The virtual showroom is the room that would be used in the virtual acoustic showroom system.

- Physical room or environment - Refers to a real/physical room.
- Virtual environment - Refers to a simulation of a physical environment. The virtual environment has two components: a virtual visual cue and a virtual acoustic cue.
- Virtual visual cue or environment - Refers to the visual portion of the virtual environment that is simulated using a 3D modeling program.
- Virtual acoustic cue or environment - Refers to the acoustic portion of the virtual environment that is simulated using the image source method of Section 2.2.3 [39].

Testing the different virtual acoustic environments that were simulated in Section 3.3, in general, was much more difficult than for speaker mimicking. One key factor complicating the tests is the lack of a direct comparison. It would be pointless to simulate an acoustic environment equivalent to the physical environment it is played in, as no useful information could be gathered. It could be possible to simulate an acoustic environment in a listening room and then go into the physical room and make a comparison. However, that would require the listener to very precisely remember the virtual acoustic environment while relocating into the physical environment, where such memory is very limited [24].

Another complicating factor is visual cues. When the listener is in the virtual showroom, the room's size and acoustic treatment are quite apparent. In the virtual showroom it is possible to listen to a virtual acoustic environment and say it sounds "different," or maybe "more lively or reverberant," but to say it sounds "like a room of X dimensions and Y acoustic treatments" would be very difficult. Having visual cues that conflict with the virtual acoustic cues could make it difficult for the listeners to validate the accuracy of the virtual acoustic cues without prior knowledge of the simulated environment. To address the issue of conflicting cues, a head mounted display was employed.

A head mounted display (Figure 5.1) is a 3D virtual environment visual display. Attached to the top of the display is a head tracking unit with which the movements of the user are tracked. Using this technology would allow the listener to be placed in a virtual visual environment that could be paired with its corresponding virtual acoustic environment. With this, testing of multiple environments by a single listener would be possible.

Figure 5.1: Head mounted display used for psychological testing of virtual acoustic environments.

5.1.1 Main Objective

As with the speaker mimicking process, the method used for creating virtual acoustic environments has been shown, empirically, to produce accurate results. It is desired that the same results be demonstrated through psychological testing. The objective of this test is to prove that the virtual acoustic environments are perceived by the listeners as accurately representing the physical room they are modeled after. If a statistically significant number of people who are tested in this system correctly identify the virtual acoustic environments, it could be concluded that, subjectively, this system is accurately replicating the physical acoustic environments.

5.1.2 Test Design

The premise for this psychological test is that several virtual environments will be presented to the listener. Some of the virtual acoustic environments will match the virtual visual environments, and some will not. The listener will be asked to identify which virtual acoustic environments and virtual visual environments are matching pairs. There are six virtual environments presented during the test that have matching virtual acoustic and virtual visual environments. These virtual environments correspond to a small and a large room in size with dead, medium, and live acoustic treatments according to the definitions given in Section 3.3. The test instructions and questionnaire for this test can be found in Appendix A.1.

When the listener is presented with a virtual environment (consisting of both a virtual visual environment and a virtual acoustic environment), the virtual acoustic environment filter is convolved with an audio clip before being sent to the speakers (see Section 3.2). Once the audio clip has been convolved with the virtual acoustic environment filter, it has the acoustical properties as if it were being presented in the physical room the virtual acoustic environment is simulating. Music was chosen as the audio clip over, say white noise bursts (see Section 2.4.1), because music will be used for the virtual acoustic showroom system when the system is implemented. Therefore, by using music, this psychological test is testing real implementation scenarios and not idealized laboratory scenarios.

At the start of the test, the listener is presented with a virtual environment for reference and familiarization, which is not one of the six mentioned above. This reference virtual environment, according to the definitions in Section 3.3, is medium in size and has a medium acoustic treatment. Figure 5.2 shows the virtual visual environment that is associated with the reference virtual environment. The music clip that is used with the reference virtual environment is of a jazz quartet and is played for approximately a minute and a half. This length of time is given to the listener to help them familiarize themselves with the head mounted display, and correlate the virtual visual environment with the virtual acoustic environment. After presenting this virtual environment, the rest of the test is explained to the listener.

Figure 5.2: 3D Model of reference room for head mounted display.

The listener was instructed to take note of the materials on the floor, walls, and ceiling in order to understand the acoustical properties of the virtual visual environment. As can be seen in Figure 5.2, there is heavy carpet on the floor, plaster on the walls, and a boarded ceiling. The speakers, television, plants, and couch were included in the virtual visual environments to aid in the listeners perception of the room size. Another consideration into the appearance of the virtual visual environments was the placement of the virtual speakers relative to the couch. For each virtual visual environment the virtual speakers were positioned such that the sound emitting from the actual speakers in the virtual showroom would appear to be coming from the virtual speakers. Figure 5.3 shows a diagram of the virtual showroom's physical setup with the virtual visual environments overlaid on it. As seen in the figure, all of the virtual speakers are in line with each other according to the listener's position in the center of the couch. This was done to reduce confusion between the virtual visual environments and the virtual acoustic environments. For example, if the speakers in the virtual showroom were placed to the sides of the listener, but the virtual speakers were placed in front of the listener, then the listener would have a difficult time believing that the

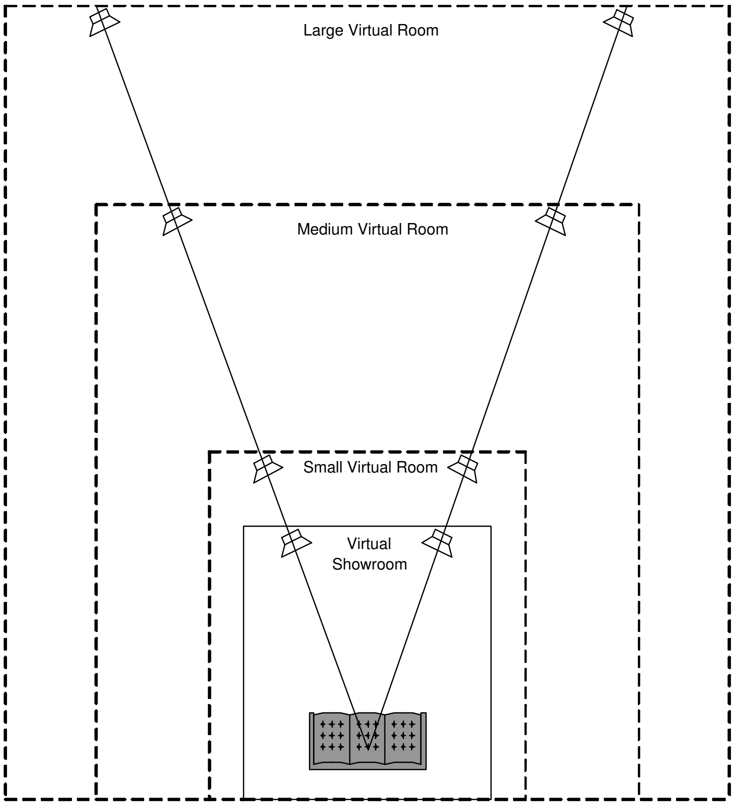

Figure 5.3: Speaker placements in virtual visual environments.

virtual visual environment was accurate for the virtual acoustic environment.

As mentioned before, this psychological test used six virtual environments. However, a total of 18 different cases were given to the listeners. With each virtual visual environment presented to the listener, three virtual acoustic environments were also presented. Table 5.1 shows the test matrix used for selecting the virtual acoustic environments to be used in conjunction with the virtual visual environments. It is clear from the table that the size of each virtual environment was held constant between the two virtual cues. For example, virtual environments 1, 2, and 3 all have the same dimensions and have acoustic treatments A, B, and C respectively. Therefore when presenting virtual visual environment 3, the virtual acoustic environments presented would be A, B, and C, where C would be the correct match. This was done so that the size of the virtual visual environment would not be an issue and focus could be given to the surface treatments. The large virtual visual environments presented to each listener are given in Figure 5.4. It is clear from the figure that each of the acoustic treatments are distinguishably different.

Table 5.1: Test matrix of virtual visual and virtual acoustic cues.

		Virtual Acoustic Cue					
		Small Dead	Small Medium	Small Live	Large Dead	Large Medium	Large Live
Virtual Visual Cue	Small Dead	X	X	X			
	Small Medium	X	X	X			
	Small Live	X	X	X			
	Large Dead				X	X	X
	Large Medium				X	X	X
	Large Live				X	X	X

(a) Large Dead Room (b) Large Medium Room (c) Large Live Room

Figure 5.4: Large virtual visual environments.

The order in which each of the virtual acoustic environments were selected for a given virtual visual environment was randomized. Also randomized was the order in which the virtual visual environments were presented. This randomization was used to reduce the effect of the learning curve of a listener. It is highly likely that after the listener has been presented with the first few virtual visual environments, they have a better understanding of what to look and listen for during the remainder of the test. This type of an effect could skew the results if the order were the same between all of the listeners.

The music used for this portion of the test was from a pop band. A clip of music approximately 20 seconds in length was used for each test case. The length was set to be long enough to help the listeners gain confidence in their judgment. Because the filtering method used reads music data from a CD and filters it in real time, a CD was created with three instances of the same music clip. For each instance of music, the appropriate virtual acoustic environment was convolved with the music and sent to the speakers. Between each music clip an announcement was given to the listener stating "Audio clip A", "B", or "C" so the listener would not loose track of which one was which. At the end of audio C the listener was asked to report which clip acoustically matched that of the virtual visual environment visible in the head mounted display. After the listener made their selection, the next virtual visual environment was presented. A few second delay between the start of the virtual visual environment and the first virtual acoustic environment was used to enable the listener to briefly think about what the new virtual visual environment would sound like.

5.1.3 Data and Statistics

The data obtained from this test was analyzed as four different tests. Each test was to determine if a significant relation to factors within the test existed. These tests were:

1. Test of the probability of getting a correct answer for each virtual environment.
2. Test of the acoustic treatment effects for dead versus medium.
3. Test of the acoustic treatment effects for medium versus live.
4. Test of the effect of room size for small versus large.

For test 1, a χ^2-Square distribution was used, while the remaining tests were analyzed as Paired t-Tests. All of the tests had the same null hypothesis, $H_o : p = 0.33$, and alternative hypothesis, $H_a : p > 0.33$. A significance level of $\alpha = 0.05$ for the first test was used. But because the three T-tests are treated as individual tests but were taken from the same data set, a conservative statistical approach called the Bonferroni correction was used. The Bonferroni correction takes m individual tests and designates that a significance level of α/m should be used for each test in order to achieve an overall significance level of α [47]. In this case an overall level of $\alpha = 0.05$ was desired, therefore for each test, a significance level of $\alpha \approx 0.0167$ was used. A total of thirty volunteers participated in this test and were primarily engineering undergraduates.

From test 1, a statistical significance was found to exist for the small dead virtual environment and the large dead virtual environment. Their P-values were found to be 0.0017 and < 0.0001 respectively. Table 5.2 gives the statistical data obtained from this test for all six virtual environments. Given that the P-values for both of the dead virtual environments (small and large) were smaller than the significance level of $\alpha \approx 0.05$, it can be concluded that the listeners confirm the accuracy of those virtual acoustic environments given their corresponding virtual visual environment. The other virtual acoustic environments did not show any significance and therefore the accuracy of those virtual acoustic environments could not be verified with this test.

Table 5.2: Results for the probability of getting the correct answer for each virtual acoustic environment (Test 1).

Room Size	Acoustic Treatment	Total Correct	χ^2-value	P-value
Small	Dead	18	9.8915	0.0017
	Medium	12	0.6649	0.4149
	Live	6	2.2931	0.1300
Large	Dead	23	25.8722	< 0.0001
	Medium	14	2.5343	0.1114
	Live	12	0.6649	0.4149

Observing the results from the T-test analyses also showed some statistical significance. The statistical data for these 2, 3, and 4 are presented in Table 5.3.

Table 5.3: Statistical results for tests 2 through 6.

Test	N	Mean	Standard Deviation	Standard error	t-value	P-value
2	30	0.25	0.5374	0.0981	2.55	0.0164
3	30	0.133	0.5403	0.0986	1.35	0.1870
4	30	-0.144	0.258	0.0471	-3.07	0.0046

Test 2 was a test of the acoustic treatment effects for dead versus medium. This test looked at the percentage of correctly chosen dead virtual environments and medium virtual environments, regardless of size (small or large), to determine which virtual acoustic environment (dead or medium) was more likely to be selected. The P-value for this test was 0.0164 which is less than $\alpha = 0.0167$ (using the Bonferroni correction) and is therefore statistically significant. Presenting the dead virtual environment yields the correct response from a listener more frequently than the medium virtual environment.

Test 3 is similar to test 2, except the comparison is between the medium virtual environment and the live virtual environment. The results from this test did not show any significance between the two virtual environments as to which would be correctly chosen more frequently.

Tests 2 and 3 were concerned with the acoustic treatments of the rooms, regardless of their size. Test 4 addresses the size of the rooms, regardless of their treatment by comparing the small virtual environments versus the large virtual environments. The P-value of this test was 0.0046 which is less than $\alpha = 0.0167$ (again, using the Bonferroni correction) and is therefore also statistically significant. This means that the large virtual environments will have correct answers more frequently than the small virtual environments.

Aside from the statistical analysis, it is worth noting which virtual acoustic environments were chosen for each virtual visual environment by the listeners. Observing this data and identifying the number of times the correct virtual acoustic environment was selected versus the number of times the other two were selected may give some insight into the system as perceived by the listeners. Figures 5.5 and 5.7 are diagrams showing the percent that each virtual acoustic environment was selected given each virtual visual environment for the large and small rooms respectively.

In Figure 5.5 it is evident that there are patterns in the responses given by the listeners. First is the extreme cases of the dead virtual visual with the live virtual acoustic and visa versa were selected very few times. Secondly, the majority of responses given followed a trend of being correct with a slight pull towards having the virtual acoustic environments less reverberant. This trend can be further visualized in Figure 5.6.

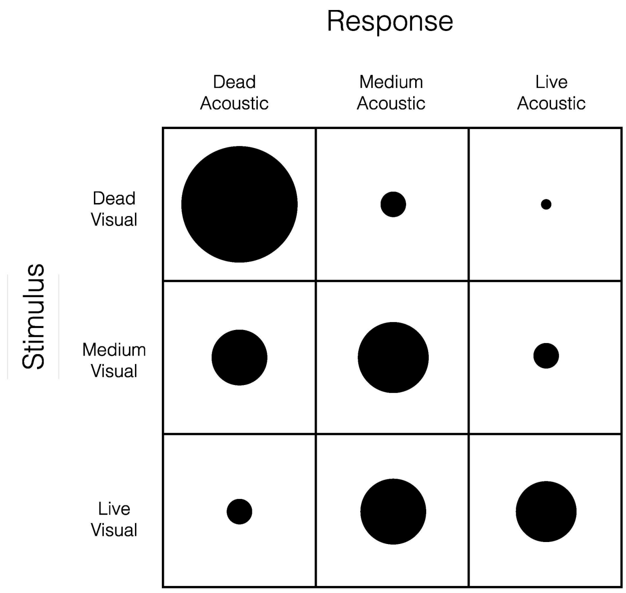

Figure 5.5: Stimulus versus response comparison of large virtual environments. The size of each circle represents the percent each response was selected for a given stimulus.

In the figure, the average T_{60} times for the three virtual visual environments are compared with the calculated T_{60} times for their corresponding physical room. The average times were gathered from taking the T_{60} time associated with the virtual acoustic environments for a given virtual visual environment, multiplying each T_{60} by the number of instances it was selected by the listeners, and then take the average of those numbers. For example, given the dead virtual visual environment, the dead virtual acoustic T_{60} is 0.2217 seconds and was selected 23 times. The medium virtual acoustic T_{60} is 0.6864 seconds and was se-

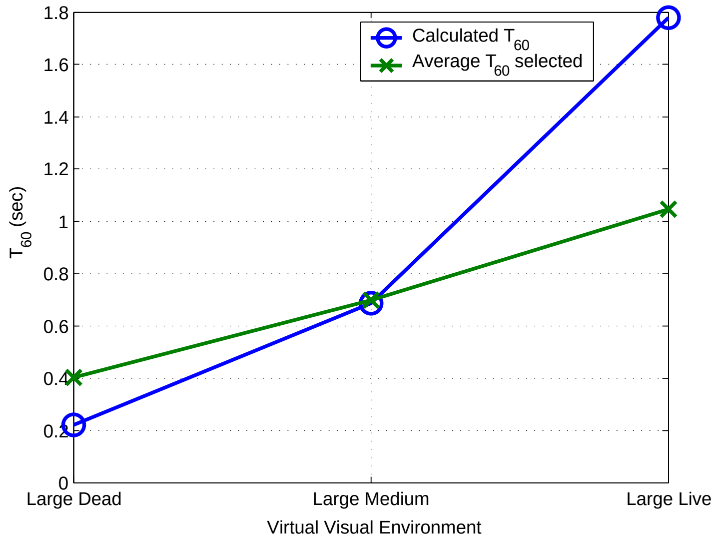

Figure 5.6: Comparison of calculated T_{60} and average selected T_{60} for large virtual environments.

lected 5 times. Finally, the live virtual acoustic T_{60} is 1.7784 seconds and was selected twice. Therefore: $1/30\big[(0.2217 \times 23) + (0.6865 \times 5) + (1.7784 \times 2)\big] = 0.403$ seconds.

Figure 5.6 shows that the slope of the average T_{60} times selected is smaller than the slope of the calculated T_{60}. This indicates that the live virtual acoustic environment, overall, may have been perceived as being too reverberant. The dead virtual acoustic environment average T_{60} is slightly more than the calculated T_{60}, but by looking at the data presented in Figure 5.5, it is evident that the majority of the listeners selected the correct virtual acoustic environment. The few who selected either of the other two virtual acoustic environments raised the average by a small amount. The results from the large virtual environments give a good indication that the virtual acoustic environments are being perceived, for the most part, as accurate by the listeners. The results from the small virtual environments are not as conclusive.

It is clear from Figure 5.7 that the responses given tend to be focused on the dead and medium virtual acoustic environments, regardless of the stimulus. This suggests that the live virtual acoustic environment was perceived as being too reverberant. The average selected T_{60} time for this case, given in Figure 5.8, confirms this.

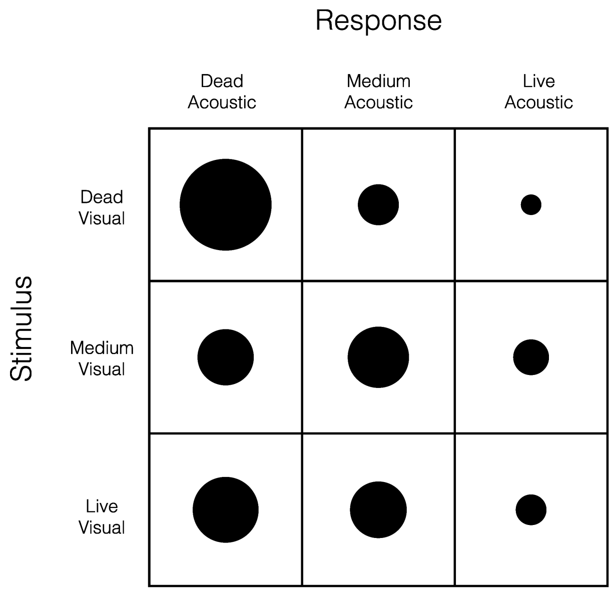

Figure 5.7: Stimulus versus response comparison of small virtual environments. The size of each circle represents the percent each response was selected for a given stimulus.

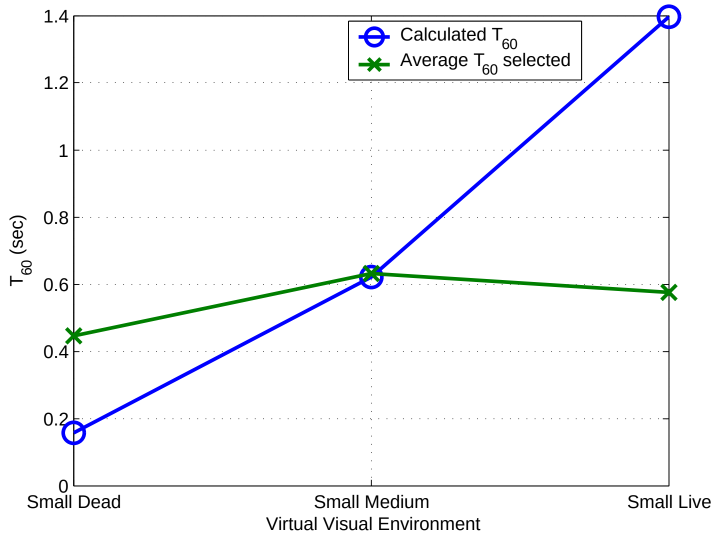

Figure 5.8: Comparison of calculated T_{60} and average selected T_{60} for small virtual environments.

The inconsistency of the average T_{60} slope between the virtual acoustic environments suggests there also may have been some confusion with the virtual visual environments that were being displayed.

The fact that the reverberation of the live virtual acoustic environments for both cases (large and small) were perceived as being too high, may be explained by the way the filters were calculated. The acoustic filters take into consideration the properties of each room, such as the walls, floor, and ceiling. They do not, however, consider any furniture in the room, and by adding furniture into a room, the amount of reverberation would be reduced. This relates to the notion that most of the participants are probably not used to rooms without any furniture, which could have led to the selection of the less reverberant acoustic cues. Figure 5.9 shows a comparison between the average selected T_{60} and the Sabine predicted time both with and without the absorption of two couches. As shown in the figure, the inclusion of the couches shifts the reverberation time to nearly the values selected by the listeners. This suggests that the listeners selected virtual acoustic environments that were less reverberant due to furniture in the virtual visual environments.

The lack of couches, or furniture in general, in the virtual acoustic environments may explain why the less reverberant virtual acoustic environments were selected more frequently. This does not explain why there is such a difference between the two cases. Comparing the results between both cases clearly shows that some interaction is taking place in the small virtual environments that is not happening in the large virtual environments. Otherwise, the results should be more consistent. The statistical analysis of this psychological test showed that some interactions were statistically significant where as others were not. Looking at the data, though, provided insight into trends and patterns that will help draw conclusions from this test, as well as recommendations for future tests.

5.1.4 Interpretation of Results

The results from this psychological test were not as conclusive in verifying the accuracy of the virtual acoustic environments as originally anticipated. However, some very useful in-

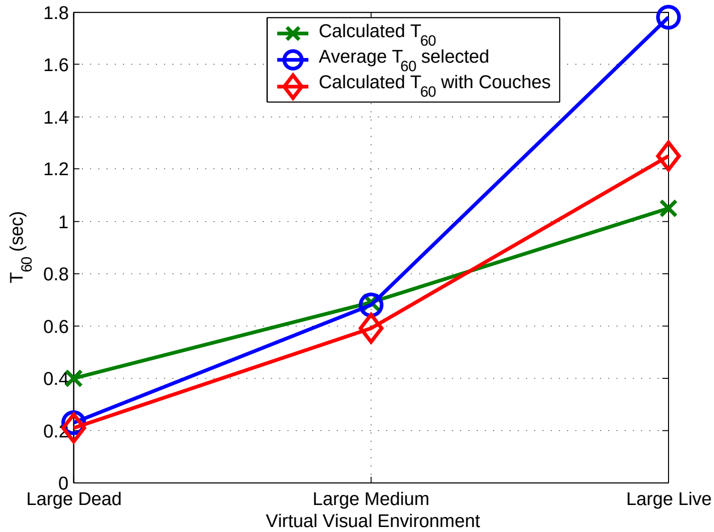

Figure 5.9: Comparison of average selected T_{60} and the calculated Sabine T_{60} for the large environments with and without couches. The T_{60} shown by the red line includes the absorption of two couches within the room.

formation about the system was obtained. The data provided statistical significance for the accurate perception of the dead acoustic environments. The other acoustic environments appear to be more reverberant than what the participants perceived as being correct, although this was not statistically substantiated. Also statistically significant was the observation that the large virtual environments were more likely than the small virtual environments to obtain correct answers from the listeners.

The data presented for the large virtual environments (figures 5.5 and 5.6) show that the right trends and patterns exist to verify the accuracy of the virtual acoustic filters, but both figures suggest that the reverberation time for the medium and live virtual acoustic environments is perceived to be too high. Lowering the amount of reverberation time should bring the two T_{60} slopes closer together and increase the number of correct responses for those two environments.

The data presented for the small virtual environments also suggests that the reverberation time for both the medium and live virtual environments is too high. However, there is another interaction occurring within the small virtual environments that is contributing to the patterns seen in figures 5.7 and 5.8. This interaction could simply be the size of the virtual environment in combination with the high reverberation time. When viewing the small virtual environments, the size the room appears to be is very small, and the listeners may not have expected a room of that size to have as much reverberation as it did.

The average selected T_{60} time for the large virtual environments was shown to be less than the predicted T_{60} of those environments. However, the average selected T_{60} is close to the predicted T_{60} of those environments with the inclusion of two couches, Figure **??**. Mentioned earlier, the virtual visual environments included a couch for the listener to sit on, virtually. But in the large virtual visual environments, a second couch was visible to aid in the listener's perception of size. This supports the notion that the reverberation in the environments was too high. It also suggests that the listeners selected the virtual acoustic environments that were more appropriate for the virtual visual environments in that the visual environments had furniture that would have reduced the reverberation in that environment.

If trained listeners had been used for this test, they may have provided the affirmation in the accuracy of the virtual acoustic environment. It was reported by Neher et. al [48] that listeners can be trained for evaluating spatial sound reproduction. By training the listeners, they have an improved ability, compared to their previously untrained ability, to correctly identify source distance using spatial sound reproduction. It has also been reported by Bech [49] that one trained listener could be used instead of 7 untrained, and the same level of statistical confidence could be achieved. A few of the listeners used in this psychological test were graduate students doing work in the Vibration and Acoustics Laboratory (VAL) at Virginia Tech. During the test, each of them selected the correct virtual acoustic environments with a higher accuracy rate over the other listeners. This may be attributed to their knowledge of the acoustic consequences of changing materials in a room. Therefore if the few graduates from VAL could be considered to be trained listeners, their data may better support the objective of this test.

Overall, there are a number of factors that could have affected the accuracy of this test. One major factor could have been the participants prior experiences in rooms similar to those given in the test. According to Naef at. al [50], "From daily experience we know that the response of a small bathroom sounds significantly different than that of a large church. Even though we may not be consciously aware of the acoustic properties, we immediately notice when something is wrong." If the listener has never been in a room of the size presented, without furniture, but with music playing, they may not have understood the environment. This may have caused the listener to think there was a problem with the virtual environment. This does not mean they understand what is causing the perceived problem, or how to identify what it is they perceive as being wrong.

Another factor could have been the virtual visual environments used with the head mounted display. A few participants had some confusion about the materials and textures that were on the floor and ceiling. If the participant did not realize what the material was, it is highly unlikely that they understood the acoustic consequences of that material being in the room. A factor also dealing with the head mounted display was the tracking unit that was used. The head tracker had a tendency to drift during the test, causing the user to have to physically look to their left, or tilt their head, in order to see forward in the virtual display. The participant was informed about that possibility prior to the test, but it would have been difficult for them to concentrate on the acoustic cue when the virtual room reference has been shifted.

The choice of music used for the virtual acoustic environments could have been another factor. A clip of pop music was used for this test, but there could have been significance in using classical music, or even white noise bursts instead. When testing a listener's perception of reverberation, often the stimulus given is anechoically recorded music [37, 48]. Anechoically rerecorded music was not used for this test because when the virtual acoustic showroom system is implemented, the users will use any music CD of their choice, none of which will be anechoically recorded. This, and the factors mentioned above, may have contributed to the lack of statistically significant data. Tests conducted in the future could be designed to reduce the possibility of these factors and achieve better results.

To prove that the virtual acoustic environments are perceived as being correct, additional tests with improved designs would be required. Changing the virtual acoustic filters to include furniture, particularly a couch or two, may improve the verification results. Having each participant select a room from a list of rooms that most closely matched their room, and test them only with that, could decrease the complexity of the test. If the only acoustic environment given to each participant was one that represented their room, there would be no need for the head mounted display and the problems with the head tracking unit would be eliminated. There have not been significant number of studies in virtual acoustics using virtual visual displays, more of which would be very useful in the development of virtual reality applications [22]. An important change in the test would be the number of participants. This test only used thirty participants, which was the bare minimum for the statistical analysis. Increasing the number of participants would greatly increase the power of the statistical analysis, giving more insight into the system. On the other hand, using trained listeners would not require more participants to achieve a higher statistical confidence.

5.2 Speaker Mimicking Verification

In trying to aurally verify the speaker mimicking process a side-by-side comparison was used. This early testing showed that there was a perceivable difference between the reference speaker mimicked as the original ($B \rightarrow A$) and the original speaker (A) – even when the difference between the measurements indicated there should not be any audible difference. Since both speakers were visible, the visual cues created perceived differences between the reference speaker and the original speaker. A black cloth was placed in front of the speakers to eliminate the visual cues, so the visual cues would not override the acoustic cues. The cloth used was chosen so as to not interfere with the response of the speakers. Subsequent tests suggested that the elimination of the visual cues greatly reduced the perceived difference for the comparison. This finding could be related and contributed to either visual capture or the ventriloquist effect described in Section 2.4.

5.2.1 Main Objective

The objective of this test is to verify the mimicking process psychologically. It is desirable to prove that the perceived variability between $B \rightarrow A$ and A is within the speaker variability of a selection of A speakers. By proving this, it could be concluded that not only do measurements of the frequency response function of $B \rightarrow A$ match that of A, but also, users will not be able to distinguish between $B \rightarrow A$ and A.

5.2.2 Test Design

First it was important to establish that the difference between A and B was significant enough to be distinguished by the listener as being separate speakers. If a listener could not tell that they were different speakers, then they would most likely be unable to distinguish any differences between $B \rightarrow A$ and A. This would cause a major problem, however, if most listeners were unable to distinguish A from B, because there would be little to no validation of the mimicking process. Therefore, a control test was set up for that purpose.

The control test consisted of $1 - A$ speaker and $3 - B$ speakers placed in the front left corner of the virtual acoustic showroom. They were arranged in a 2×2 matrix, where their placement was randomized and changed for each listener. A clip of music, approximately 20 seconds long was played through each speaker in an order that was also randomized between listeners. However, for each listener all four speakers were played through twice in the same order. This would allow the listener to make a better comparison between the 1^{st} and 4^{th} speakers. After the order was played through the second time, the listener was asked to record which speaker $(1 - 4)$ sounded the most different and why. The results from this control test are given in Section 5.2.3.

The experiment to test the speaker to speaker variability was set up in the front right corner of the virtual acoustic showroom and was almost identical to that of the control test. For this test there were $3 - A$ speakers and $1 - B \rightarrow A$ speaker, shown in Figure 5.10. All of the testing procedures were the same as for the control test. And again the listener was asked to record which speaker sounded the most different and why. The results are in Section 5.2.3. The test instructions and questionnaire for this test can be found in Appendix A.2.

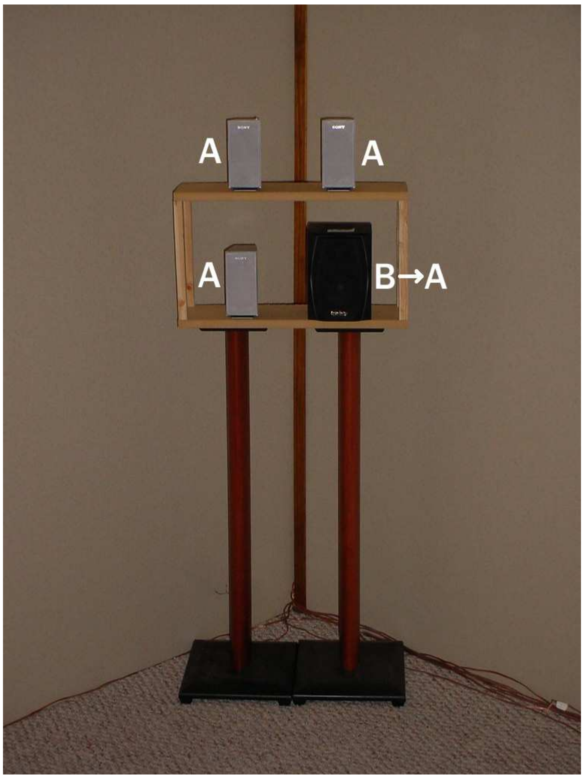

Figure 5.10: Psychological test setup for testing speaker to speaker variability

5.2.3 Data and Statistics

A group of twenty-four volunteers were given both the control and the variability test. These tests were designed and analyzed as binomial tests with a null hypothesis of $H_o: p = 0.25$, an alternative hypothesis of $H_a: p > 0.25$, and a significance level of $\alpha = 0.05$. The statistical analysis provided P-values and confidence intervals for interpreting the test data. P-values give the probability of seeing the particular value of the test statistic ($p = 0.25$) if H_o is true [38]. If the P-value is less than the significance level, then the null hypothesis will be rejected at the $(1 - \alpha) \times 100\%$ confidence level. The confidence intervals are given at 95%

($\alpha = 0.05$) and show that the values of p supported by the data fall between the upper and lower confidence level.

For the control test 21 out of the 24 listeners chose speaker A as being the most different. While the other three each chose a separate speaker B as being the most different, so that speaker B_1, B_2, and B_3 were each chosen once by one of the three listeners. Figure 5.11 shows the percent that each speaker was selected. A P-value of < 0.0001 was found, thereby rejecting the null hypothesis that speaker A would be selected 25% of the time. With a 95% confidence interval the statistical analysis put the true percentage of people that would choose speaker A as being between 67.63% and 97.34%. This confirms that there is an audible difference between these two speakers and most people can make that distinction. Table 5.4 shows the statistical data obtained.

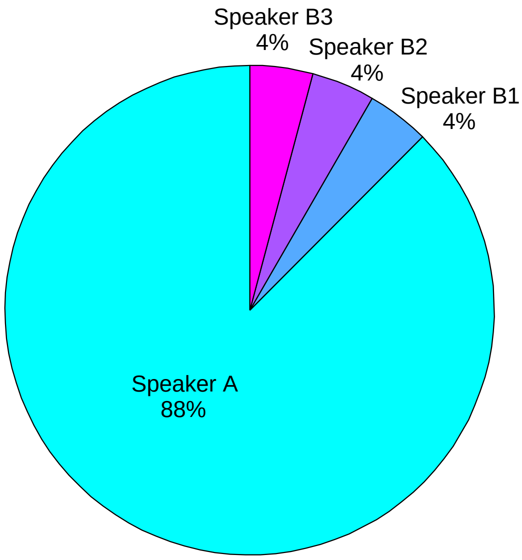

Figure 5.11: Control test pie chart.

In the variability test the objective was to have minimal differences between all four speakers so that there is an equal probability of selecting $B \to A$ or any of the three A speakers. With this, it would be expected to have each speaker chosen approximately the same number of times if the mimicking process was successful. As it turned out, one third of the listeners (8 of the 24) chose $B \to A$ as being the most different. Out of the other 16 listeners, 6 chose A_1, 7 chose A_2, and 3 chose A_3. Figure 5.12 shows the percent that each speaker was selected. The P-value for this test was found to be 0.467, which fails to reject the null hypothesis. This, however, does not discount the possibility of reaching the 25% selection value with a larger sample size, especially since this P-value is substantially larger than α. As the P-value increases towards 1, the sample p becomes the same value as the test statistic, therefore confirming the null hypothesis. The confidence interval given by the analysis put the true percentage of people that would choose speaker $B \to A$ as being between 15.63% and 55.32%. This is a large range in which the true value could be in, but by increasing the sample size, the range around the true value would be reduced. The results are given in Table 5.4.

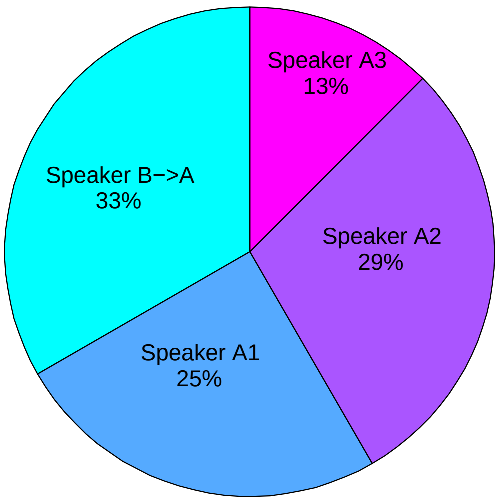

Figure 5.12: Variability test pie chart.

Table 5.4: Results for speaker mimicking psychological test.

Test	X	N	Sample p	95% Confidence Interval	Exact P-Value
Control Test	21	24	0.87500	0.676389, 0.973441	< 0.0001
Variability Test	8	24	0.33333	0.156302, 0.553219	0.467592

5.2.4 Interpretation of Results

The objective of this test was to prove that the variability between $B \rightarrow A$ and A is perceived as being within the variability between a selection of A speakers. The statistical results gathered from the experiments are sufficient enough to conclude that the variation between speaker $B \rightarrow A$ and A is perceived as being within the variability between A speakers, and therefore verifying the speaker mimicking process. Meaning, that by using the inverse filtering method of Section 2.3, one speaker can be used to mimic another and be verified both empirically and psychologically.

5.3 Conclusion

The tests detailed in this chapter were focused on the human perception of the mimicking process and simulating the virtual acoustic environments. It was shown that the mimicking process is accurate enough to have the perceivable variation between $B \rightarrow A$ and A within the variation of a set of A speakers. The main psychological effect for that test was to remove the conflicting visual cues. It was also shown that the dead virtual acoustic treatments are perceived as being accurate, even though the test results did not place any statistical significance on the accuracy of the other virtual acoustic environments. It was also shown that there are obvious trends in the data that suggest the reverberation time for the medium and live virtual acoustic environments is too high. The lack of furniture being taken into consideration in the virtual acoustic environments was shown to be a very likely cause for this trend. Modifications to the testing procedure may yield the desired results, but more investigations would have to be made before it can be concluded that listeners perceive the virtual acoustic environments as being accurate.

Chapter 6

Conclusion

This thesis has discussed the development of a virtual acoustic showroom that simulates both audio loudspeakers and listening environments. The two main components of the system, the speaker mimicking and the simulation of virtual acoustic environments, were both detailed in their development and implementation into the complete virtual acoustic showroom system. Both components were proven to produce accurate and desired results empirically. They were both then tested for human perception using psychoacoustic tests. The mimicking process again yielded desired results proving that this process is perceived by listeners to be accurate. The psychoacoustic testing of the simulation of virtual acoustic environments did not give results that proved, definitively, the accuracy of this process. It did, however, show that the perception of the medium and live acoustic environments was too reverberant and provided insight into how to improve the system for future implementations. Each chapter is briefly summarized below. Also, improvements for future work will be discussed.

6.1 Chapter Summaries

This thesis began with a brief introduction into virtual acoustics and listed a few methods that are used for the auralization process. The motivation for the work was to benefit speaker retail stores by substantially reducing display inventory by using a new-technology based marketing tool. The current research was reviewed and it was found that very little validation of virtual acoustic systems have been analyzed. Finally the scope outlined the steps to be taken in this research.

After having given a background on previous research into virtual acoustic simulations, chapter 2 presented a detailed description on the theory behind simulating audio loudspeakers and acoustic environments. Sound radiation from a monopole and a baffled piston were described in regards to a speaker's response as well as the distinction between near field and far field radiation. A room's acoustical qualities and the measures of such qualities were defined. The image source method was discussed in detail giving the background on calculating virtual acoustic environments. The method of inverse filtering for its use in speaker mimicking was defined. Also, different aspects of psychoacoustic testing were discussed, including test design and statistical analysis of the test data.

Chapter 3 focused on simulating acoustic environments for the virtual acoustic showroom. The showroom prototype was verified to have an anechoic acoustic treatment above $500Hz$, making it suitable for simulating acoustic environments. Results for the small room with medium acoustic treatment and the large room with live acoustic treatments showed nearly the same T_{60} curve between the filter and the anechoic implementation. The T_{60} curve between the anechoic and showroom implementation also showed a good match, but with some divergence in the showroom nearing the $4kHz - 8kHz$ frequency range. There is a crossover point at about $2kHz$ for the T_{60} curves (filter, anechoic, and showroom) for both of these rooms; below that point they underestimate the reverberation time and above that point they over estimate the reverberation time, when compared with the Sabine predicted reverberation time. This may require some modifications to be made to the way the filters are calculated. However, with the results presented, it is clear that the environment filters that are calculated using the image source method and absorption coefficients can be implemented into the virtual acoustic showroom to accurately simulate a range of acoustic environments for users to select from.

After having proven that the virtual acoustic environment filters can be accurately implemented in the showroom, the process of mimicking speakers was investigated. Chapter 4 detailed the testing methods for measuring the frequency response function of a speaker and how to use that response in creating an inverse filter for the mimicking process. It was also shown that the inverse filtering method could be used to reduce the modal response of the virtual showroom. It was proven that a speaker could be accurately mimicked by another speaker using anechoic mimic filters under the following conditions: The measurement location must be close enough to the speaker so that the area effected by the mimic encompasses the listener's position and is not at an anti-node of the virtual showroom. The number of coefficients in the inverse filter can be reduced by estimating the roll-off of the low end of the inverse filter's response by a curve fit. The reference and original speakers must have close to the same spatial displacement of speaker drivers within the enclosure, as well as being co-located in the showroom for verification. Finally the interaction of the showroom must reduced by using an inverse filter of the room.

The accuracy of the mimicking process, as well as the simulation of the virtual acoustic environments, was determined experimentally, but the accuracy needed to be validated using psychoacoustic tests. Chapter 5 detailed psychoacoustic tests focusing on the human perception of both these items. For the mimicking process, it was shown that the perceivable variation between $B \rightarrow A$ and A is within the variation of a set of A speakers. Therefore, validating the accuracy of the mimicking process. For the simulation of the virtual acoustic environments, it was shown that there are obvious trends in the data that suggest the reverberation time for the medium and live virtual acoustic environments are perceived as being too high, where the dead virtual acoustic treatments are perceived as being accurate.

This system as it stands, the mimicking of speakers and the simulation of acoustic environments, provides a new and unique way for stereo retail stores for marketing speakers.

6.2 Future Work

While the virtual acoustic showroom system accurately mimics speakers and simulates listening environments, improvements can be made to the system. The areas of improvements include the calculation of the virtual acoustic filters, the psychoacoustic validation of the system, and the delay while switching between environments.

The image source method is used to calculate the early and late (diffuse) reflections of the virtual acoustic environments. If, instead, a hybrid method were used, where the late reflections are simulated with recursive filters, the time required to calculate each virtual acoustic environment filter would be reduced. The reduction in calculation time comes directly from the fact that only the first several reflections would have to be calculated individually. Currently each reflection, early and late, are calculated individually. Other hybrid methods may also be used allow the affects such as diffusion and diffraction to be taken into account. These affects may improve the perception of the virtual acoustic environments, but at a cost of increasing computation time. Another improvement for the virtual acoustic environment calculations would be to include furniture. The inclusion of furniture will reduce the reverberation times of the virtual acoustic environments, making them more representative of actual living environments.

Improved psychoacoustic tests could be used to further validate the perception of the virtual acoustic environments detailed in this thesis. First, fewer virtual environments would be presented to each listener. Perhaps only a single virtual environment, representing their own living environment. Doing this, some virtual environments would be presented to many listeners, while others would only be presented to one or two listeners. However, the information gathered from this test design would be whether the listeners thought the presented virtual acoustic environment sounded like their own living environment. Second, a better head tracking unit would be used with the head mounted display. The drifting that occurred during the tests, was distracting to the listeners, and may have contributed to the selection of the wrong virtual acoustic environments. Finally, the music selection would consist of anechoically recorded music. The music used was produced with reverberation added into it. This additional reverberation may have compounded the issues of having too much reverberation in the virtual acoustic environment filters.

Other enhancements to the virtual acoustic showroom could be made, increasing its appeal to customers. Using a flat panel display in the showroom, that displays a picture of the environment corresponding to the virtual acoustic environment the customer is listening to, This will be beneficial because the head mounted display will not be used in the showroom. Along the lines of the flat panel display would be projectors that present the image of the speaker the customer is listening to. Projecting the image onto the surface that is covering the reference speakers, will aid in the perception that those are the speakers being played.

Before this virtual acoustic showroom system is installed in a retail store, it would be beneficial to conduct more studies in it. Mainly to test its performance as a showroom. This type of study could be conducted by opening a demo store, where a large number of consumers would be readily available, in which a consumer study could be performed. The system could then be modified as needed to enhance its effectiveness as a showroom.

Appendix A

Psychoacoustic Test Forms

A.1 Room Response Psychoacoustic Forms

A.1.1 Informed Consent Form

Informed Consent for Participant of Investigative Project Virginia Polytechnic Institute and State University

Title of Project: Acoustical Environment Cue Matching with Three-dimensional Visual Displays
Principal Investigator: Dr. Marty Johnson
Co-Investigator: Christopher Collins

1. THE PURPOSE OF THIS RESEARCH/PROJECT

   You are invited to participate in a study of interaction in virtual environments. This research studies the ability for people to match audio and visual cues in a three-dimensional virtual world. This study involves experimentation for the purpose of evaluating and improving the process of creating virtual acoustic environments.

2. PROCEDURES

   You will be asked to perform a set of tasks using a virtual environment system. These tasks consist of comparing acoustic and visual cues in a 3D environment. You will wear a head-mounted display (HMD). Your role in these tests is that of evaluator of the virtual acoustic cues. We are not evaluating you or your performance in any way; you are helping us to evaluate our system. All information that you help us attain will remain anonymous. You may be asked questions during and after the evaluation, in order to clarify our understanding of your evaluation.

   You may also be asked to fill out a questionnaire relating to your background with such systems, and to take a short test of hearing ability.

The session will last about 30 minutes. The tasks are not very tiring, but you are welcome to take rest breaks as needed. You may also terminate your participation at any time, for any reason.

You will be given full instructions before beginning. It is important that you understand the instructions before beginning. If anything is unclear, be sure to ask us questions.

3. RISKS

   The proposed experiments are straightforward tests of performance using standard virtual environments displays, trackers, and input devices. Participation involves sitting in a virtual acoustic facility (while wearing the head-mounted display). All light and sound intensities are well within normal ranges. The only foreseeable physical risks are slight eye strain, dizziness, or mild nausea. There are no known mental risks.

   If you experience any eye strain, dizziness, or nausea during a session, then between tasks please remove the HMD and take a rest break. The experimenter will explain when you can take such rest breaks. If you are having trouble with any task, please tell us. If dizziness or nausea becomes uncomfortable, you will be allowed to leave with no penalty.

4. BENEFITS OF THIS PROJECT

   Your participation in this project will provide information that may be used to improve the design of virtual environments hardware and/or software. No guarantee of benefits has been made to encourage you to participate. You may receive a synopsis summarizing this research when completed. Please leave a self-addressed envelope with the experimenter and a copy of the results will be sent to you.

   You are requested to refrain from discussing the evaluation with other people who might be in the candidate pool from which other participants might be drawn.

5. EXTENT OF ANONYMITY AND CONFIDENTIALITY

   The results of this study will be kept strictly confidential. Your written consent is required for the researchers to release any data identified with you as an individual to anyone other than personnel working on the project. The information you provide will have your name removed and only a subject number will identify you during analyses and any written reports of the research.

6. COMPENSATION

   Your participation is voluntary and unpaid.

7. FREEDOM TO WITHDRAW

   You are free to withdraw from this study at any time for any reason.

8. APPROVAL OF RESEARCH

   This research has been approved, as required, by the Institutional Review Board for projects involving human subjects at Virginia Polytechnic Institute and State University, and by the Department of Computer Science.

9. SUBJECT'S RESPONSIBILITIES AND PERMISSION

   I voluntarily agree to participate in this study, and I know of no reason I cannot participate. I have read and understand the informed consent and conditions of this project. I have had all my questions answered. I hereby acknowledge the above and give my voluntary consent for participation in this project. If I participate, I may withdraw at any time without penalty. I agree to abide by the rules of this project

Signature:
Date:
Name (please print):
Contact Phone:
Email address (OPTIONAL):

Should I have any questions about this research or its conduct, I may contact:

Investigator: Dr. Marty Johnson
Phone (540) 231-4797
Professor, Mechanical Engineering Department VAL (231-4162)
email: martyj@vt.edu

Co-Investigator: Christopher Collins
Phone (540)231-6790
Email: chcollin@vt.edu

Review Board: David M. Moore,
Office of Research Compliance, CVM Phase II (0442) 231-4991

cc: the participant, Dr. Johnson and Christopher Collins

A.1.2 Use Questionnaire

Please help us to categorize our user population by completing the following items.

Gender (circle one): Male Female

Age:

Do you wear glasses or contact lenses (circle one)?
No Glasses Contact Lenses

Do you have any hearing problems (circle one)? If yes please explain. Yes No

Occupation (if student, indicate graduate or undergraduate):

Major / Area of specialization (if student):

How frequently do you listen to music (circle the best answer)?

1. a. not at all
2. b. once a month
3. c. once a week
4. d. several times a week
5. e. daily

Do you consider yourself an audiophile (circle one)?
Yes No Don't know

Rate your familiarity with computers: (circle one)

1. not at all familiar
2. not very familiar
3. somewhat familiar
4. fairly familiar
5. very familiar

How often do you use computers...
...for work? (circle the best answer)

1. not at all
2. once a month
3. once a week
4. several times a week
5. daily

...for fun? (circle the best answer)

1. not at all
2. once a month
3. once a week
4. several times a week
5. daily

Have you ever used a virtual reality (VR) system? If so, please describe it (what type of display was used, what kind of application (e.g. game, architectural walk-through) was running, how did you interact with the system, etc.).

A.1.3 Test Instructions

The listeners were told they would be presented with one virtual visual room. While viewing that room, they would be presented with three virtual acoustic environments. One of the three acoustic environments would match the virtual visual room visible in the head mounted display. They were told to imagine that they were physically sitting in that visual room and think about what it would sound like. They were then told to take note of the materials on the floor and the ceiling and think about what affect those materials have on the acoustic properties of the visual room. After the third virtual acoustic environment, they were told to identify which one was the match. They were then told that a total of six virtual visual environments would be presented, each of which would have three virtual acoustic environments.

The listeners were then warned about the tendency for the head mounted display to drift. But, all the information they needed about the virtual visual room could be gained by looking forward. And by only looking forward, the tendency for the head mounted display to drift would be greatly reduced.

A.2 Speaker Mimicking Psychoacoustic Forms

A.2.1 User Questionnaire

Test 1

A	B
C	D

Which speaker sounds the most different?

Why?

Test 2

A	B
C	D

Which speaker sounds the most different?

Why?

A.2.2 Test Instructions

For both the control and variability test:

The listeners were told that there are four speakers behind the black cloth in a 2×2 matrix. Each of the speakers will be played through in a random order twice. Both times, the order would be the same, to allow a better comparison between the first and the last speakers. They were asked to report which speaker sounded the most different from the other three, and why. They were told that once they selected the speaker that was the most different, the test would begin, with another set of speakers in the other front corner of the showroom.

Bibliography

[1] Mendel Kleiner, Bengt-Inge Dalenbäck, and Peter Svensson. Auralization – an overview. *Journal of the Audio Engineering Society*, 41 (11) 861–875, 1993.

[2] Sean E. Olive, Peter L. Schuck, Sharon L. Sally, and Marc E. Bonneville. The effects of loudspeaker placement on listener preference ratings. *Journal of the Audio Engineering Society*, 43 (9) 651–669, 1994.

[3] Damian T. Murphy, Chris J. C. Newton, and David M. Howard. Digital waveguide mesh modeling of room acoustics: Surround-sound, boundaries and plugin implementation. In *Proceedings of the COST G-6 Conference on Digital Audio Effects (DAFX-01)*, 2001.

[4] Diemer de Vries and Marinus M. Boone. Wave field synthesis and analysis using array technology. In *Proceedings of the 1999 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, 1999.

[5] Thomas Funkhouser, Ingrid Carlbom, Gary Elko, Gopal Pingali, Mohan Sondhi, and Jim West. A beam tracing approach to acoustic modeling for interactive virtual environments. *Computer Graphics*, 32(Annual Conference Series):21–32, 1998.

[6] Marinus van der Voorden, Lau Nijs, Gerrit Vermier, and Gert Jansens. The applicability of ray-tracing based simulation tools to predict sound pressure levels and reverberation times in 'coupled spaces'. In *Proceedings of the 1999 International Building Performance Simulation Association (ISBPA) Conference on Building Simulation*, 1999. www.ibpsa.org/proceedings/bs99/papers/D-17.pdf.

[7] Jont B. Allen and David A. Berkley. Image method for efficiently simulating small-room acoustics. *Journal of the Acoustical Society of America*, 65 (4) 943–950, 1979.

[8] Jeffrey Borish. Extension of the image model to arbitrary polyhedra. *Journal of the Acoustical Society of America*, 75(6):1827–1836, 1984.

[9] Peter David Yardley. Image source method for simulating wave motion in rectangular enclosures. Master's thesis, Department of Computer Science, The University of Sheffield, UK, 2001.

[10] Jyri Huopaniemi, Lauri Savioja, and Matti Karjalainen. Modeling of reflections and air absorption in acoustical spaces – a digital filter design approach. In *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA'97)*, 1997.

[11] Lauri Savioja, Jyri Huopaniemi, Tommi Huotilainen, and Tapio Takala. Creating interactive virtual acoustic environments. *Journal of the Audio Engineering Society*, 47 (9) 675–705, 1999.

[12] William Grant Gardner. The virtual acoustic room. Master's thesis, Massachusetts Institute of Technology, Cambridge, Massachusetts, 1992.

[13] Jens Holger Rindel and Claus Lynge Christensen. Room acoustic simulation and auralization – how close can we get to the real room? In *WESPAC 8, The Eighth Western Pacific Acoustic Conference, Melbourne*, 2003.

[14] Lauri Savioja, Jyri Huopaniemi, Tommi Huotilainen, and Tapio Takala. Real-time virtual audio reality. In *Proceedings International Computer Music Conference*, pages 107 – 110, August 1996.

[15] Ville Pulkki and Taqpio Lokki. Creating auditory displays with multiple loudspeakers using vbap: A case study with diva project. In *Proceeding from the Fifth International Conference on Auditory Display*, 1998.

[16] Bill Gardener. A realtime multichannel room simulator. *Journal of the Acoustical Society of America*, 92(A)(2395), 1992.

[17] Aki Mäkivirta, Poju Antsalo, Matti Karjalainen, and Vesa Välimäki. Modal equalization of loudspeaker-room responses at low frequencies. *Journal of the Audio Engineering Society*, 51 (5) 324–343, 2003.

[18] Vance Dickason. *The Loud Speaker Design Cookbook*. Audio Amateur Press, Peterborough, New Hampshire, fifth edition, 1997.

[19] Murray P. Rosenthal. *How To Select And Use Loudspeakers And Enclosures*. Hayden Book Company, Inc., 1987.

[20] David A. Bies and Colin H. Hansen. *Engineering Noise Control Theory and Practice*. Spon Press, second edition, 2002.

[21] Lawrence E. Kinsler, Austin R. Frey, Alan B. Coppens, and James V. Sanders. *Fundamentals of Acoustics*. John Wiley & Sons, Inc., fourth edition, 2000.

[22] Durand R. Begault. *3-D Sound For Virtual Reality And Multimedia*. Boston: Academic Press, 1994.

[23] Dr. techn. Per V. Brüel. Technical review 96–02 – anechoic chambers. http://www.bruel-ac.com/tr/tr9602/tr9602.html.

[24] Heinrich Kuttruff. *Room Acoustics*. Applied Science Publishers LTD, London, second edition, 1979.

[25] Charles M. Salter Associates Inc. *Acoustics*. William Stout Publishers, San Francisco, 1998.

[26] R. O. Neubauer. Estimation of reverberation time in rectangular rooms with non uniformly distributed absorption using a modified fitzroy equation. *Building Acoustics*, 8 (2) 115–137, 2001.

[27] Peter David Yardley. Image source method for simulating wave motion in rectangular enclosures. Master's thesis, The Department of Computer Science, University of Sheffield, United Kingdom, 2001.

[28] Fokke van Saane. Altiverb impulse responses. http://www.xs4all.nl/ fokkie/IR.htm.

[29] Patrick Gaydecki. A versatile real-time deconvolution dsp system implemented using a time domain inverse filter. *Measurement Science And Technology*, 12 82–87, 2001.

[30] Eric W. Weisstein. Fast fourier transform. From MathWorld–A Wolfram Web Resource. http://mathworld.wolfram.com/FastFourierTransform.html.

[31] Athanasios Mouchtaris, Panagiotis Reveliotis, and Chris Kyriakakis. Inverse filter design for immersive audio rendering over loudspeakers. *IEEE Transactions on Multimedia*, 2 (2) 77–87, 2000.

[32] Pavel Zahorik. Auditory display of sound source distance. In *Proceedings of The 2002 International Conference on Auditory Display*, 2002.

[33] Pavel Zahorik. Assessing auditory distance perception using virtual acoustics. *Journal of the Acoustical Society of America*, 111 (4) 1832–1846, 2002.

[34] D. H. Mershon and E. King. Intensity and reverberation as factors in the auditory perception of egocentric distance. *Perception and Psychophysics*, 18 409–415, 1975.

[35] Ruth Y. Litovsky, H. Steven Colburn, William A. Yost, and Sandra J. Guzman. The precedence effect. *Journal of the Acoustical Society of America*, 106 (4) 1633–1654, 1999.

[36] Rachel K. Clifton, Richard L. Freyman, and Jennifer Meo. What the precedence effect tells us about room acoustics. *Perception & Psychophysics*, 64 (2) 180–188, 2002.

[37] T. I. Niaounakis and W. J. Davies. Perception of reverberation time in small listening rooms. *Journal of the Audio Engineering Society*, 50 (5) 343–350, 2002.

[38] G. Geoffrey Vining. *Statistical Methods for Engineers.* Duxbury Press, Pacific Grove, California, 1998.

[39] Jan Troge, Christopher Collins, Jamie Carneal, and Marty Johnson. Development of a real-time virtual room acoustic simulation. Technical report, Vibration and Acoustic Laboratory, Virginia Polytechnic Institute and State University, 2003.

[40] M. David Egan. *Architectural Acoustics.* McGraw-Hill, Inc., 1988.

[41] J. Duncourneau and V. Planeau. The average absorption coefficient for enclosed spaces with non uniformly distributed absorption. *Applied Acoustics*, 64 (2003) 845–862, 2003.

[42] Barry Blesser. An interdisciplinary synthesis of reverberation viewpoints. *Journal of the Audio Engineering Society*, 49(10):867–903, October 2001.

[43] W. B. Joyce. Sabine's reverberation time and ergodic auditoriums. *Journal of the Acoustical Society of America*, 58(3):643–655, 1975.

[44] David Lay. *Linear algebra and its applications.* Addison-Wesley, Reading, Massachusetts, second edition, 1998.

[45] Leo L. Beranek. *Acoustics.* American Institute of Physics, Inc., 1993.

[46] Louis D. Fielder. Analysis of traditional and reverberation-reducing methods of room equalization. *Journal of the Audio Engineering Society*, 51(1/2), 2003.

[47] Eric W. Weisstein. Bonferroni correction. From MathWorld–A Wolfram Web Resource. http://mathworld.wolfram.com/BonferroniCorrection.html.

[48] Tobias Neher, Francis Rumsey, and Tim Brookes. Training of listeners for the evaluation of spatial sound reproduction. In *AES 112th Convention of the Audio Engineering Society, Journal of the Audio Engineering Society*, 2002. preprint 5584.

[49] Søren Bech. Selection and training of subjects for listening tests on sound-reproducing equipment. *Journal of the Audio Engineering Society*, 40 (7/8) 590–610, 1992.

[50] Martin Naef, Oliver Staadt, and Markus Gross. Spatialized audio rendering for immersive virtual environments. In *Proceedings of the ACM symposium on Virtual reality software and technology*, pages 65–72. ACM Press, 2002.

Vita

Christopher M. Collins was born in Eugene, OR on June 27^{th}, 1977 to Debra and Michael Collins. He graduated with a Bachelor's of Science degree in Industrial and Manufacturing Engineering from Oregon State University in March, 2002. In August, 2002 he began to pursue his Master's of Science degree in Mechanical Engineering with a concentration in acoustics at Virginia Polytechnic Institute and State University.