UNIVERSITY OF MIAMI

SPATIALLY RELOCATED FREQUENCIES AND THEIR EFFECT ON THE LOCALIZATION OF A STEREO IMAGE

By

Robert G. Hartman

A Research Project

Submitted to the Faculty
of the University of Miami
in partial fulfillment of the requirements for
the degree of Master of Science

This research was supported by Delphi Automotive Systems

Coral Gables, Florida

May 2003

UNIVERSITY OF MIAMI

A research project submitted in partial fulfillment of
the requirements for the degree of
Master of Science

SPATIALLY RELOCATED FREQUENCIES AND THEIR EFFECT
ON THE LOCALIZATION OF A STEREO IMAGE

Robert G. Hartman

This research was supported by Delphi Automotive Systems

Approved:

_________________	_________________
Ken Pohlmann Professor of Music Engineering	Dr. Edward Asmus Associate Dean of Graduate Studies
_________________	_________________
Colby Leider Assistant Professor of Music Engineering	Dr. Mohamed Abdel-Mottaleb Professor of Electrical Engineering

HARTMAN, ROBERT (M.S., Music Engineering Technology)

Spatially Relocated Frequencies and Their Effect on the Localization of a Stereo Image (May 2003)

Abstract of a Master's Research Project at the University of Miami.

Research project supervised by Professor Ken Pohlmann.
No. of pages in text: 125

The ability of humans to detect the location of a sound is generally referred to as localization. Sound waves interact with the head, body, and pinnae creating temporal and spectral differences between the left and right ear canal signals. The brain uses these differences to interpret a probable number of sound events and their respective locations. There are three major cues: interaural time differences (arrival, phase, envelope), interaural level differences, and the monaural pinnae influences. The physical presence of these cues depends mostly on the spectral content of the sound and its spatial origin relative to the listener. Perceptually, the localization cues exhibit a relative dominance that varies significantly with frequency. This research explores the relative importance of low and high frequency localization cues during free field listening. More specifically, it compares the horizontal shift of a stereo image caused by spatially relocating low versus high frequency bands of the audible spectrum. It is shown that contrary to the popular belief that low frequencies are "hard to localize," the horizontal position of a stereo image is most significantly affected by moving low-to-mid frequencies as opposed to high frequencies. This can most likely be attributed to the overall perceptual dominance of low frequency interaural phase differences.

DEDICATION

This work is dedicated to all of those around me who unselfishly gave their support and understanding while I completed this thesis. First, I send all my love and appreciation to my immediate family who have been my biggest fans while also challenging me to accept nothing less than my dreams. To my parents, whose hard-working attitude and modest lifestyle have taught me to work for what I want and be generous with what I have. To my little sister, who has encouraged my successes and maintained a loving relationship despite our differences. To Amy Beth, thank you for all of your sacrifices, and for being pleasant in a difficult situation - this would have been a lot tougher without your support. Finally, to my grandparents and extended family - thank you for staying involved in my life and always being there to tell me how proud you were. Your love and kindness has shaped the person that I am today, and given me the confidence and determination to complete this project.

ACKNOWLEDGMENT

I would like to especially thank my advisor and mentor, Ken Pohlmann, who has generously given of his precious time and provided me with several great opportunities during my time at UM; I hope we will again work together on future projects. To Kevin Heber of Delphi Automotive Systems, who took the initiative to get involved with my research and provided useful equipment and financial contributions. Thank you also to my thesis committee, for their time and cooperation with this work. Of course this research would not have been possible without all of the student volunteers who generously gave their time with minimal begging on my part. Also to the faculty and staff of UM including Joe Abbati, Luis Ruiz, and Paul Griffith who were flexible enough with equipment, facilities and schedules to allow me to complete the needed listening tests.

Table of Contents

List of Figures

List of Tables

Chapter 1: Introduction

Over the last century, auditory scientists have made great progress in understanding the complexities of human hearing. However, as with any sensory behavior, there will always be uncertainty between what is perceived and what can be physical measured. Physically, the analysis of human hearing can be reduced to studying the signals entering the left and right ear canals. Yet perceptually, these two complex signals contain many encoded messages that the brain deciphers into audible qualities such as loudness, pitch, timber and spatial origin.

This ability for complex processing is especially applicable during localization - determining the spatial origin of a sound event. It has been shown through experiments that humans do not have an absolute sense of a sound's location. Because of this, scientists explicitly differentiate between the "sound event," where the sound physically originates, and the "auditory event," where the sound is perceived to originate.

For a single sound source, the auditory and sound event often occur in close proximity. However for multiple sound sources, the overall perceived event depends on many factors. For instance, if each sound event is unique in location, pitch and timbre - they are typically perceived as independent sounds with their own spatial origins. Yet, if the sounds are similar enough, they might be integrated into one perceived auditory event with one collective spatial location. In this multi-source situation, the overriding perception is based on the agreement of the localization cues and the correlation of the sound sources.

The research for this thesis has focused on studying the localization of a stereo image when certain frequencies of the signal have been spatially relocated in a multi-source sound system. This is a fairly common occurrence in most consumer electronic systems. Often the signal is split into several frequency bands, each sent to a transducer in a different spatial position. Consider some "three-way" loudspeaker systems shown in Figure 1, which shows the spatial relocation of low (L), middle (M), and high (H) frequencies for a stereo, automotive, and home theater surround audio system.

Note that in the home stereo system of Figure 1a, the left loudspeaker enclosure has three transducers which share the same horizontal position but have different vertical positions. The low frequencies come from the vertically lowest portion of the enclosure whereas the tweeter is significantly above it at the top. An automotive audio system (Figure 1b) represents an even more severe condition, having different vertical and horizontal origins for all three frequency bands. Moreover, in surround sound home theater systems (Figure 1c), the lowest frequencies are often separated to a powered subwoofer in an altogether different horizontal and vertical position. This is deemed acceptable because low frequencies are said to be "hard to localize." This popular, yet somewhat oversimplified observation will be discussed in more detail later.

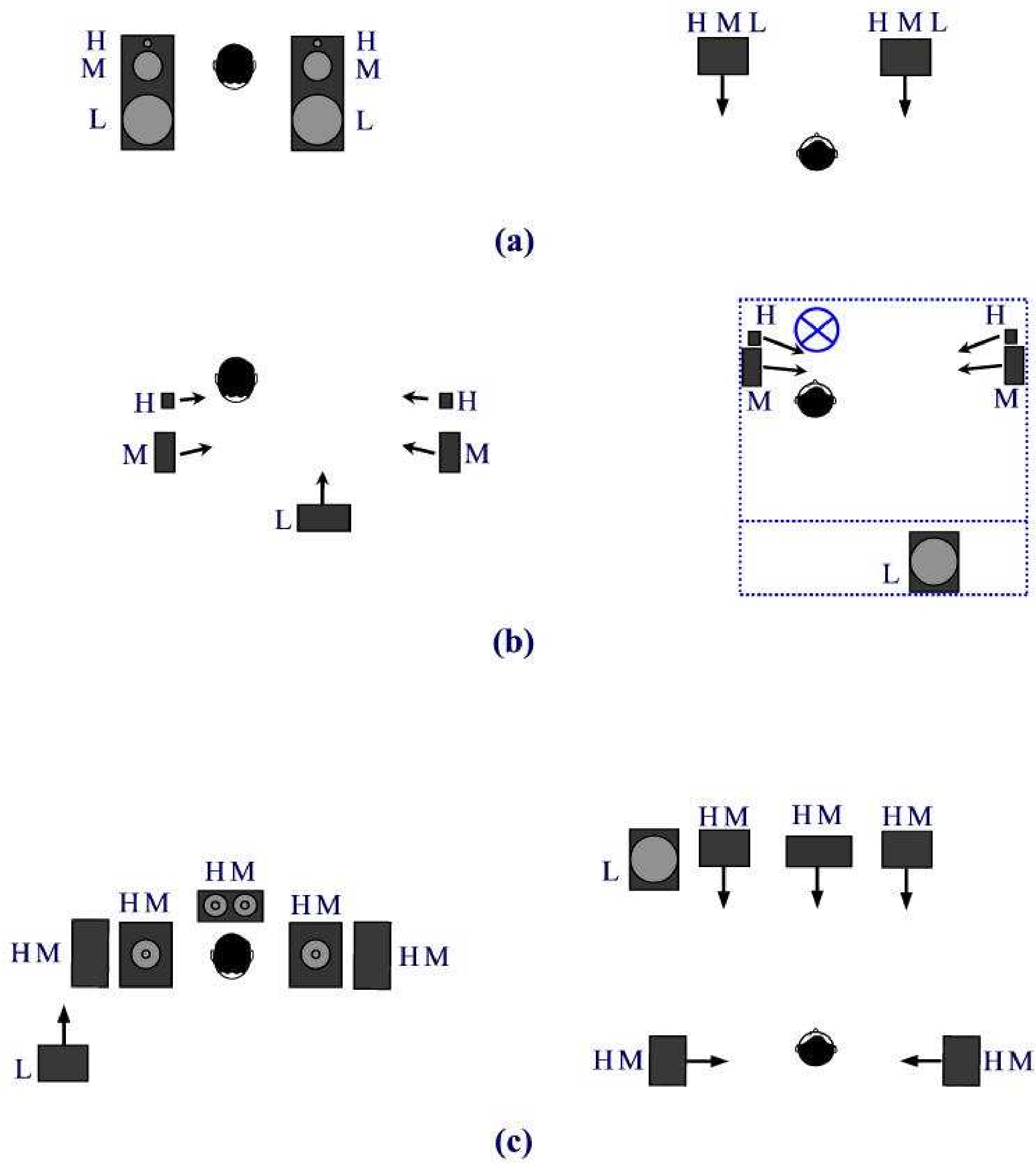


Figure 1: Views of a listener from the Rear (left) and Top (right) and the spatial relocation of high (H), mid (M), and low (L) frequencies. Shown are typical (a) Stereo, (b) Automotive, and (c) Home Theater Surround loudspeaker systems

As mentioned, the focus of this research was to study the perceived location of a stereo image when portions of the signal are moved to a different spatial location. This idea originated in the suggested reading of a paper on Digital Theater System's (DTS) surround sound encoding called "Coherent Acoustics." In the paper, Smyth (1999) specifically mentions, "experimental evidence suggests that it is difficult to localize mid-to-high frequency signals above about 2.5 kHz, and therefore any stereo imagery is largely dependent on the accurate reproduction of only the low-frequency components of the audio signal" (p. 18). This obviously seems to contradict popular opinion, and warranted a preliminary investigation into the claims Smyth made.

Early investigation into this topic included well-known texts such as Blauert (1999), Yost (1987) and Begault (1994), as well as some previous UM graduate research by West (1998) and Ballman (1990). There are also many good summary articles, such as those from Hartmann (1999) and Kendall (1995). These and other sources, suggest that Smyth's (1999) hypothesis of a sound's spatial location being dominated by low over high frequencies localization cues is well established among auditory scientists.

To further study this idea, it was investigated whether high frequencies could be limited to a mono-tweeter, the same way that low frequencies are sent to a mono-subwoofer. A typical stereo loudspeaker system was set up, adding a separate tweeter directly in front of the listener (see Figure 2). The stereo signal was processed so that all frequencies above 10 kHz were sent to the central tweeter, while frequencies below 10 kHz were played from their respective left/right speakers. Using a wide range of music, it became clear that while the spectral balance was kept mostly intact, the image localization of high-pitched instruments was sometimes different.

Figure 2: Preliminary experiments tested "mono-ized" high frequencies

In fact, the image's sound stage position seemed to depend on the spectral energy distribution of its frequencies relative to the 10 kHz crossover. Essentially, the more high frequency energy the instrument had, the farther towards the tweeter (center sound stage) it was pulled. For example, a hard-left-panned cello was correctly localized at the left because of dominant lower frequency energy, while a left-panned cymbal crash was now heard somewhere between left and center.

This result mandated a new direction for the thesis. One option would have been to investigate how noticeable this type of image shift actually was. After all, would most people even realize the cymbal image had shifted towards the center? While this line of questioning might make interesting marketing data (and was actually performed for a small portion of this thesis), it would be difficult to scientifically explain the results because of so many potential variables.

Instead, this thesis compared the relative impact that spatially relocating low versus high frequencies have on the perceived horizontal location of a stereo image. The experimental setup would be a less severe version of that shown in Figure 2, bringing a full range center speaker much closer to one of the stereo pair. Also, rather than moving only high frequencies, the objective would be to move various portions of the audible spectrum to this offset speaker to see which had the greatest effect on the perceived location. The details of the experiment will be covered later in this paper.

To present this research, the topic of localization is first introduced. This includes a discussion of the various localization cues and the results of historical experiments measuring human accuracy, precision, and sensitivity to those cues. Yet for practical purposes, it is more important to consider the physical nature of the cues and their relative perceptual significance. Also, the topic of auditory scene analysis is introduced as it relates to this research. Scene analysis shows how conflicting spatial and spectral cues, as well as the acoustic space, might impact the "integration" (fusion) and localization of sound events. Next, the specific experimental setup and methodologies used in this research are reviewed. This is followed by the results, which are presented and analyzed in various forms. Finally, conclusions are made and future areas of research involving this topic are suggested.

Chapter 2: Localization

One can rarely read a publication on the topic of localization without seeing some reference to the early twentieth century work of Lord Rayleigh (1907) and his duplex theory of sound. Rayleigh felt that humans rely on two types of cues for localization: interaural time differences (ITD) and interaural level differences (ILD). However, his theory did not allow scientists to understand how localization occurred when ITD and ILD were zero or equivalent, which occurs in several situations. Thus, a complement to Rayleigh's explanation is that monaural cues from the pinnae must provide additional help with localization when the duplex theory does not apply.

Since Rayleigh's time, many experiments have been performed to further our understanding of this topic. The experimental setup typically falls into one of two broad categories, free field (localization) or with headphones (lateralization). While free-field testing is obviously more natural, headphone testing tends to be more popular because it allows the isolation of each localization cue and also removes any effects from the room. Yet headphone testing also has its drawbacks, such as the creation of internalized auditory images (i.e. headphone images are perceived inside the head).

For either type of testing, it is important to introduce some basic terms used to describe the position of the sound and auditory events (see Blauert, 1999). This position is typically described by a horizontal angle (azimuth, φ) and vertical angle (elevation, δ) from a point directly in front of the listener. Note from Figure 3 that azimuth and elevation both start at zero directly in front, and increase in a counterclockwise fashion.

The starting point is the intersection of two imaginary planes, the horizontal and median planes. The horizontal plane is the extension of the interaural axis, containing the center of the ear canal entrance and the lower portion of the eye sockets. It essentially splits the head into an upper and lower portion. The median plan bisects the head into left and right portions, whereas the frontal plane creates the front and rear halves. All three planes are perpendicular to one another and intersect at the center of a symmetrical head (see Figure 3).

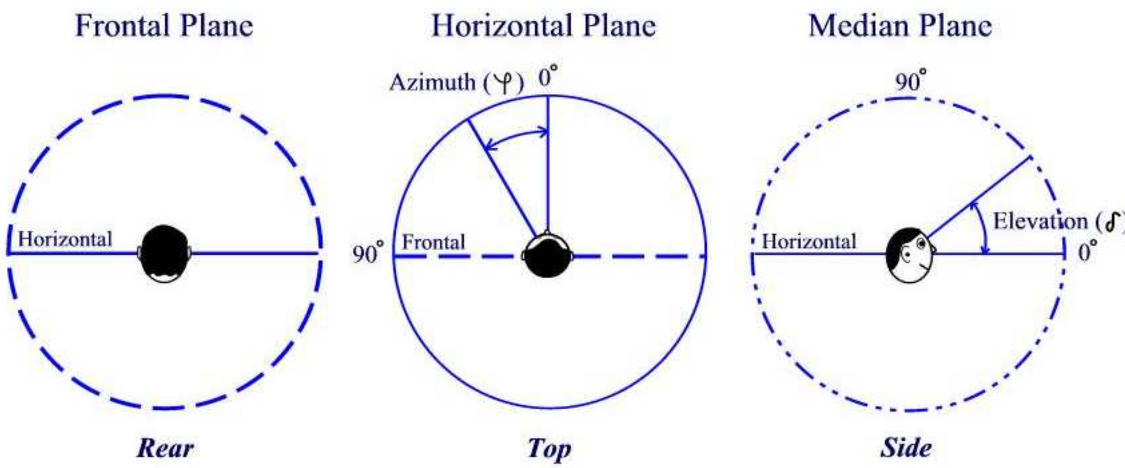


Figure 3: Views of the auditory planes, azimuth angle and elevation angle

Interaural Difference Cues

As was mentioned, two major localization cues involve the interaural level and time differences between the left and right ear canal signals. Yet, the term "interaural time difference" is somewhat nebulous because it can represent arrival, phase, and envelope temporal differences. Therefore the terms will be defined as below:

- **Interaural Arrival Time Difference (IATD):** The difference in arrival time between left and right ear signals. This is due to the constant speed of sound with varied path length differences (see Figure 4). The sound typically reaches one ear and then must additionally go around the head to the opposing ear.

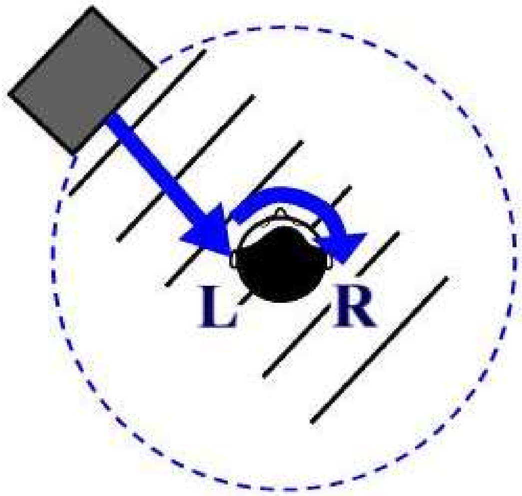


Figure 4: ITDs are caused primarily by path length differences

- **Interaural Phase Difference (IPD):** The difference in phase between left and right ear signals caused by different arrival times. For a periodic sound ($T = \frac{1}{f}$), IPD can have two different physical values: either IATD or T - IATD.

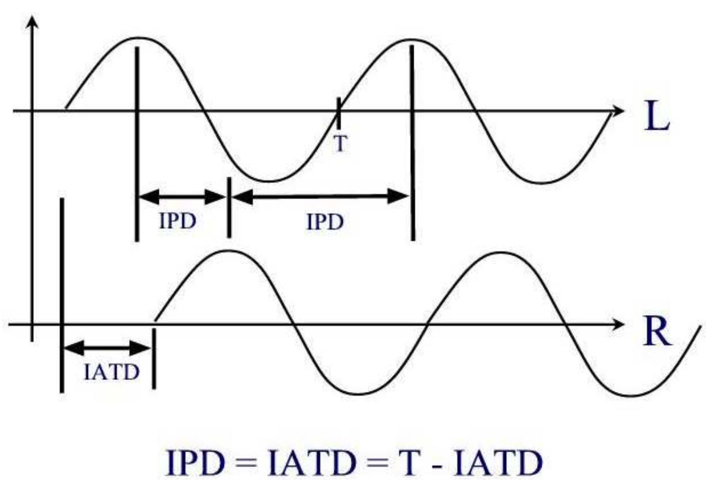
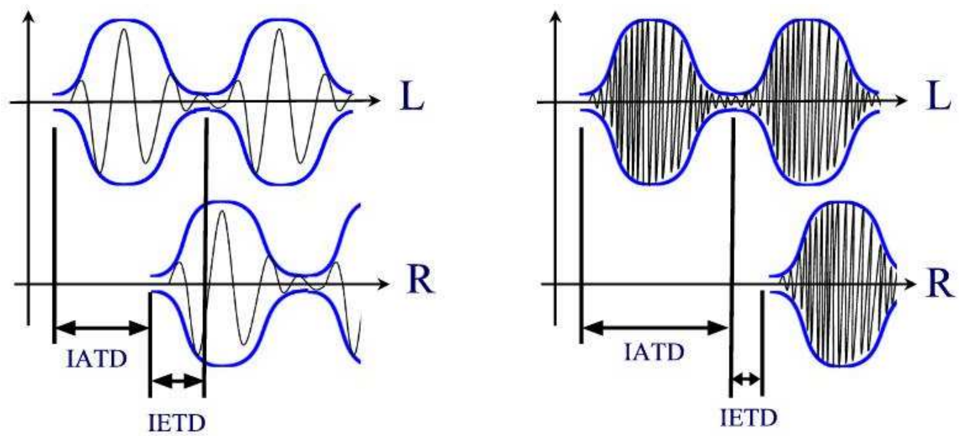


Figure 5: Interaural Phase Difference (IPD) has two physical values due to periodicity

- **Interaural Envelope Time Difference (IETD):** The temporal difference of the modulation pattern between the two ears. IETD is independent of the carrier frequency. Similar to IPD, it also exhibits two physical values (see Figure 6).



IETD = IATD = T - IATD

Figure 6: Interaural Envelope Time Difference (IETD) also has two physical values due to periodicity

- **Interaural Time Difference (ITD):** A generic term used to describe any of the above time differences. Typically refers to the one that dominates the signal frequencies under discussion. According to Blauert (1999), continuous sounds under 1.6 kHz would be dominated by IPD, while IETD has a definite influence above 1.6 kHz. Although IATD directly affects IPD, its only direct influence is to impulsive sounds.

With regards to interaural level differences (ILD), those frequencies with long wavelengths as compared to the 17.5 cm diameter head are relatively undisturbed. As the frequency of the sound increases (decreasing wavelength), it will begin to either reflect off or refract around the head (see Figure 7). ILD is additionally dependent on source position. This is because of the asymmetrical characteristics of the head and body, and also the properties of acoustical waves and the barriers they encounter.

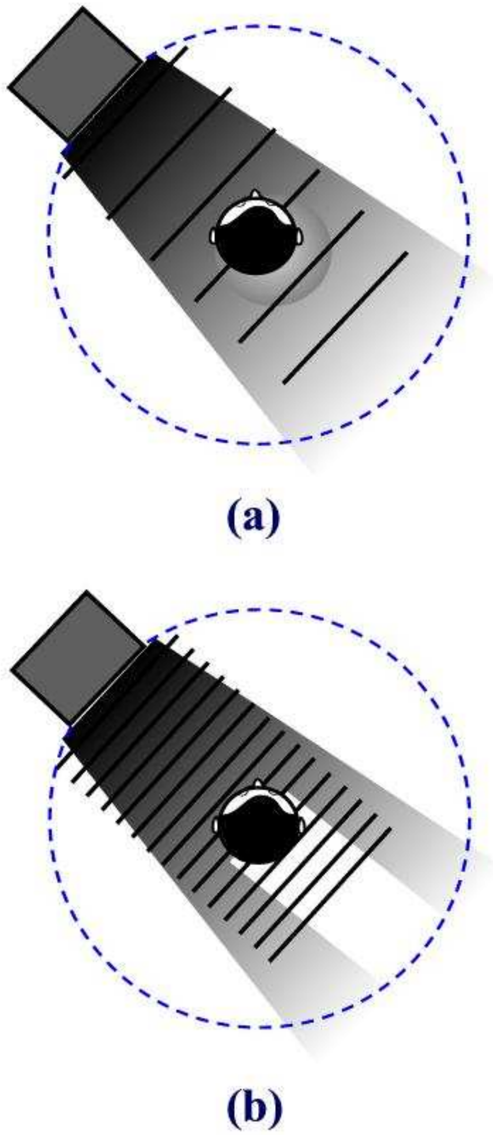


Figure 7: Interaural Level Differences caused by reflection of sound off head is (a) minimal for low frequencies (b) significant for high frequencies

Monaural Cues

Having discussed the interaural cues, it is also important to realize the role played by monaural cues. These cues are important because for every sound source position, there is a unique group of points that shares the same path to the ears (Durrant & Lovrinic, 1984). These points are more commonly know as the "cone of confusion," and are represented by a hyperbola in the horizontal plane and a cone in three-dimensional space (see Figure 8). Two sound sources located on this cone would provide identical interaural cues; thus the monaural cues allow listeners to differentiate between them.

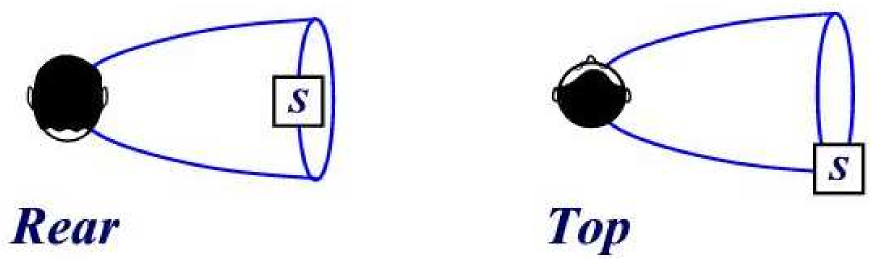


Figure 8: The cone of confusion is a set of points which provides identical interaural cues

Scientists have known for some time that it is possible to localize sounds with only one ear. Angell and Wite (1901) compared the localization abilities of a normal binaural hearing individual to one who was entirely deaf in one ear. They found that the monaural individual's localization ability on the side of the non-deaf ear was "not greatly inferior" (p. 236) to the normal hearing individual. However, hearing on the side of the deaf ear was "extremely uncertain" (p. 243).

For both subjects, front/back confusions occurred often and in general, complex sounds (whistles and bells) were more accurately localized than pure tones (tuning forks). Their final conclusion stated that in monaural hearing, the external ear was responsible for contributing "qualitative peculiarities," (Angell & Wite, 1901, p. 246) to the sound, which allowed proper localization to occur.

A more detailed analysis of the origin of monaural cues was not published until Batteau (1967) and Blauert (1969). Blauert described the operation of the pinna as a directionally dependent filter. He stated that it enhanced or reduced various spectral portions of the input signal depending on the angle of vertical and horizontal incidence. Blauert's experiments suggested that these spectral influences dominated localization in fixed-head experiments; and that the actual location of the sound source had little to do with its perceived location.

For instance, because sounds originating from overhead exhibit a peak in the 7 kHz range, a sound that is played in the horizontal plane with an artificial peak at 7 kHz is perceived to originate from overhead. Blauert defined several sections of the auditory frequency range that behave this way. He called them "preference bands," and showed that the relative intensity of these bands is what dictates fixed-head localization.

Batteau (1967) also discussed the influences of the pinna, but in terms of time-based reflections. He showed that sound will reflect off the individual folds and cavities of the pinna, causing replication of the original signal with very small time delays. Batteau measured an almost linear relationship between azimuth and monaural pinna delay ranging from 10 μsec (on axis with the ears) to 90 μsec (directly in front) (p. 163). He also showed that changing the elevation of the sound source influenced the amount and concentration of pinna delay.

Wright, Hebrank, and Wilson (1974) reinforced the plausibility of Batteau's theory by showing that humans are sensitive to time delays as short as 20 μsec (p. 960).

However, Middlebrooks (1997) points out that time delays essentially cause spectral amplitude modifications due to phase interactions of the original and delayed signals. Thus, researchers since Batteau's time have focused on "spectral modifications, rather than on time delays per se" (Middlebrooks, p. 78).

Localization Blur

Our ability to detect changes in a sound source's position is experimentally measured as the minimum audible angle (MAA), also called "localization blur." There are various methods of experimentation, but in essence, the localization cues are varied from a fixed point and the MAA is calculated to be the minimal amount of change that a statistically significant number of listeners can detect. The MAA can be measured for both horizontal and vertical directions.

Blauert (1999) has summarized much of the localization blur experiments, including influential work from Stevens and Newman (1936) and Mills (1958). From this summary, Blauert suggests that our most acute sense of localization is directly in front (0° azimuth). In that position he states "the absolute lower limit for the localization blur is, as shown, about 1°" (p. 38). Schmidt, Vangemert, De Vries, and Duyff (1953) also state that changes in azimuth for pure tones close to the median plane were "considerably less than one degree" (p. 16).

Precision in locating a source is affected by its spatial location and frequency content. With regards to source location, MAA in the horizontal direction (azimuth) is generally considered to be more accurate than that of the vertical plane (elevation)

(Strybel and Fujimoto, 2000). On the horizontal plane, MAA is smallest directly in front of the listener where it intersects the median plane. As the source moves around the head, MAA slowly increases to a maximum on axis with the ears, and then decreases again as the sound continues towards the rear of the listener. In the vertical plane, MAA is again most accurate directly in front near the horizontal plane. It similarly increases to its maximum directly above the listener's head before decreasing to its secondary minimum directly behind the listener.

MAA also varies with the signal's spectral content. Testing, such as Mills (1958) has shown that for pure tones, the middle frequency range generally has a larger MAA than either low or high frequencies. In addition, narrowband signals and sinusoids are intrinsically more difficult to localize than wideband signals because of the limited number of localization cues the brain has to consider.

This discussion above describes the general nature of localization blur, but in reality it is more complex. To get an idea for this complexity, consider Figure 9 from Blauert (1999). This is the result of test subjects aligning the azimuth of a sinusoidal (solid) and octave-band (dotted) sound source to that of three wideband sources fixed at azimuths of 0º and ± 40º from midline.

Notice how the perceived azimuth varies with frequency and also between the two narrow band test signals. While the results seem to change somewhat unpredictably with frequency, the 0º incident typically has smaller variations than either the 40º or 320º (i.e. -40º) positions. Also, the results fall into a finite area around the wideband source, which is an indication of localization blur. The white noise should be considered the absolute position of the source, whereas the difference in azimuth for the sinusoid/octave band represents the localization blur. For example, a 5kHz sinusoid (solid) at ~44º azimuth is perceived to share the same location as the white noise source at 40º. This suggests that the MAA for a 5 kHz sinusoid is approximately 4º. In comparison, a 5kHz octave band (dotted) seems to have a MAA around 14º (located at ~26º azimuth) when compared to the same white noise source.

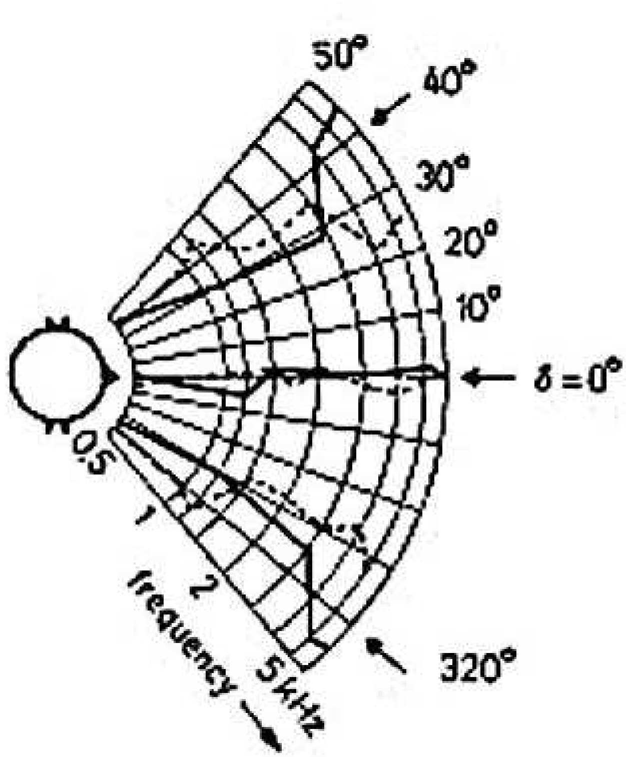


Figure 9: Horizontal plane localization of sinusoidal (solid) and narrow band noise (dotted) as compared to a reference sound of wide-band noise at 0, 40, and 320 degree azimuth locations. Shown versus frequencies to 5 kHz. Reprinted from Blauert (1999) with permission from the MIT press.

Chapter 3: Localization Cue Salience

It is also important to consider the relative hierarchy that localization cues have on the resulting auditory event. First, the acoustical nature of sound waves dictates that the physical level of a localization cue will vary with spatial position and frequency. Perceptually, research has shown that localization cues vary in importance relative to one other and with frequency. Understanding both the physical and perceptual significance of the localization cues should help suggest which frequencies could be spatially relocated with a minimal impact on the position of the resulting auditory image.

Physical Aspects

To understand the relationship between free field listening conditions and the physical localization cues they create, first consider the interaction of two different frequency tones with a listener (Figure 10). The sound pressure wave travels towards the subject and interacts with the body and head as an acoustic barrier. Depending on the frequency content and angle of incidence, portions of the wave may be reflected off the listener or may refract around them with minimal interaction.

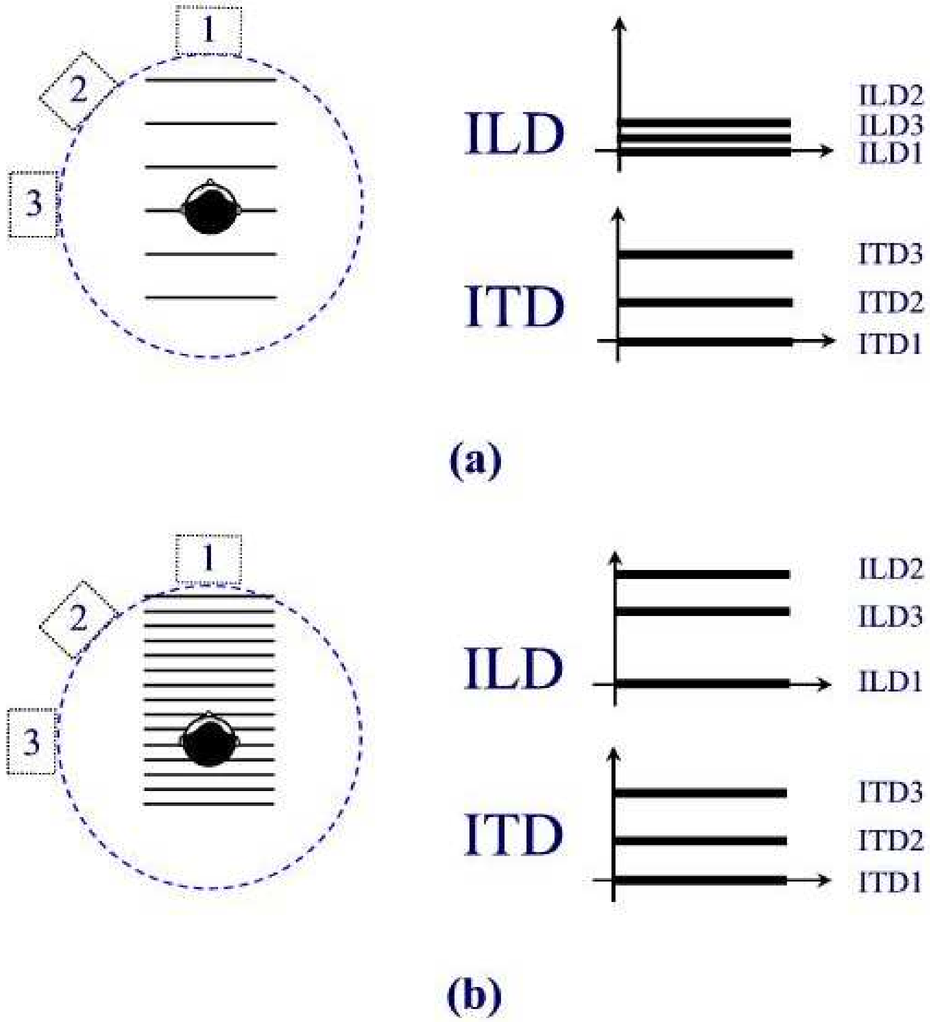


Figure 10: ITD and ILD variations with speaker position (azimuth)

As is shown in Figure 10, a sound source located on the median plane (position 1) has no ILD or ITD for either frequency case. This is because the sound source has a similar path to both ears. Yet, when the source is moved counter-clockwise on the horizontal plane, the two interaural cues have varied changes.

In particular, note that the ITD gradually increases because of the increased path length difference, reaching its maximum at position 3. ITDs are independent of acoustic effects, which is why they are the same in both frequency cases. However, because ILDs are caused by the complex interaction of sound with the listener, position 2 will generally exhibit a greater level difference than position 3. In fact position 3 is unique, because in some conditions the diffraction of sound can cause the opposing ear to be louder than the incident ear (Blauert, 1999, p. 71). Also notice that the ILD is considerably less for low frequencies (Figure 10a) than for high frequencies (Figure 10b). This is because at low frequencies, sound will refract around the head whereas high frequencies will be reflected, creating an acoustic shadow on the opposing ear.

A similar analysis of individual ITD cues shows how each varies with frequency and spatial position (see Figure 11). Notice that IATD is dependent only on spatial position and is independent of frequency changes. On the other hand, despite the same spatial position, the IPD shown in Figure 11(a) is more than (b) because of the decreased signal wavelength. The IPD cue will also change due to differences in position only, as is shown in Figure 11(a) and (c). Although not shown, IETD is similar to IPD, meaning that it only varies with changes in spatial origin and modulation frequency.

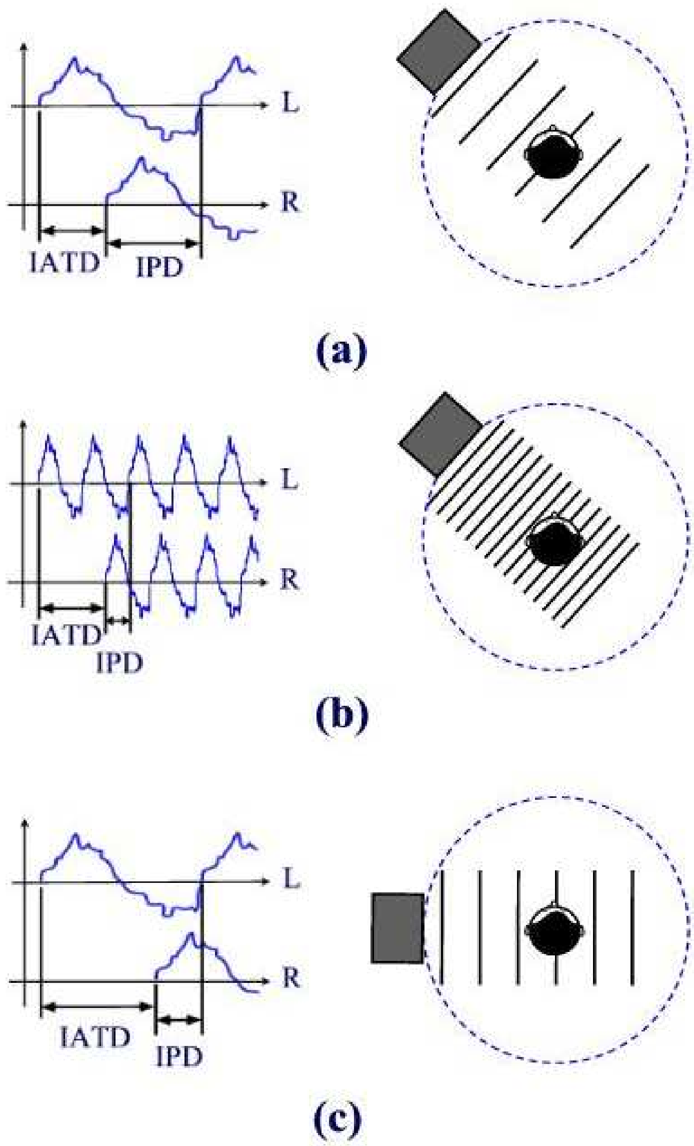


Figure 11: IATD varies only with position, while IPD varies with position and signal frequency. Shown for (a) low frequency source (b) high frequency source (c) high frequency source with new spatial location

The generic results shown in Figure 10 and Figure 11 are quite simplified from actual in-ear measurements (i.e. recordings taken with microphones placed at the entrance to the ear canals). Many frequency dependent influences such as variations from the pinnae and torso reflections are not represented. In fact, complex plots of both ILD and ITD occur with changes in azimuth and frequency. These patterns are shown from Blauert (1999) in Figure 12**.** Note that ILD (left) and ITD (right) are presented, where each plot shows the variations versus frequency for a fixed azimuth. The azimuth is varied from the top plot (φ=150º) to the bottom plot (φ=30º), where azimuth is measured counter-clockwise from directly in front.

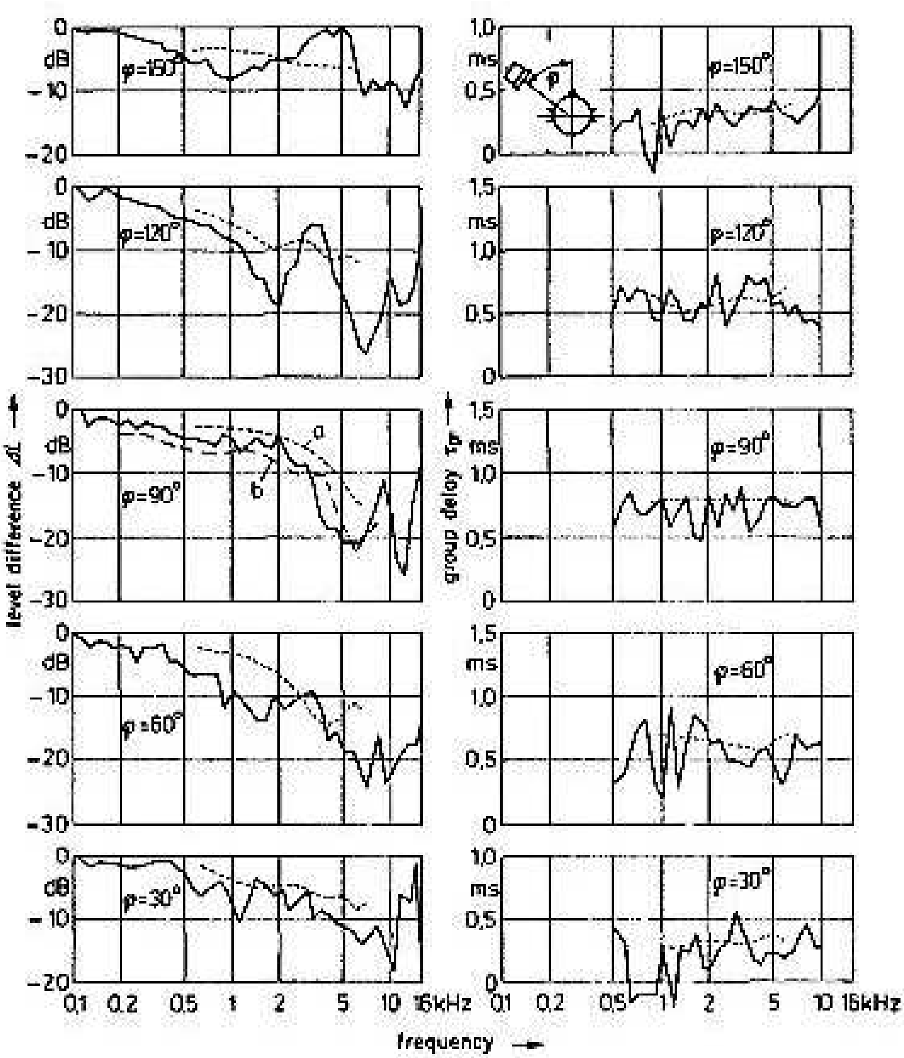


Figure 12: Complex patterns of ILD (left) and ITD (right) with varied horizontal plane positions (azimuth). Reprinted from Blauert (1999), with permission from the MIT press.

As for monaural cues, it is most relevant to consider the frequency range where they physically occur. Because of the wavelength of sound relative to the size of the pinnae folds, monaural cues exist in only a small portion of the audible spectrum. One of the most common ways to determine their range of occurrence is by analyzing in-ear recordings. Batteau (1967) appears to have been one of the first to measure in-ear recordings as he contemplates that more research of this type had not be done because "the extraordinary fidelity needed in all aspects of this system, microphones, amplifiers, headphones, acoustic isolation, perhaps has prevented construction of the requisite systems until now" (p. 161).

Butler and Belendiuk (1977) furthered the use of in-ear recordings by comparing the amplitudes of the signals at the two ears. With this analysis, they showed that as a sound source moves from above to below the interaural axis, a notch of frequencies moves continuously from high (approx. 7 kHz at 15° elevation) to low (approx. 5.5 kHz at -30° elevation) (p. 1267). Also, Musicant and Butler's (1984) experiments used various low pass and high pass noises to show that spectral cues above 4 kHz are originating from the pinnae and help avoid front/rear confusion. They also found spectral cues in the 1-4 kHz range, which are caused by the interaction of sound with the torso. The higher end of monaural cues reaches 10-12kHz, where Hebrank and Wright (1974) measured the effect of a small peak responsible for an upper-rear sense of direction.

Additionally, Middlebrooks (1997) calculated the "directional transfer function (DTF)," obtained by taking in-ear measurements and then subtracting a signal representing the average of 360° microphone measurements with no subject present. The DTFs showed that a spectral notch increases in center frequency from 7 kHz to 9.5 kHz for 0° to 160° azimuth and from 6 kHz to 10 kHz from -60° to 60° elevation.

Perceptual Aspects

The human auditory system is more sensitive to some localization cues than others, and at times will even ignore a physical cue (Buell & Trahiotis, 1994). What is known has been taken mostly from lateralization (headphone) tests, where the physical cues can be independently controlled. A specific kind of "trading experiment" is very popular, where the cues are put into conflicting conditions to see which will dominate.

Through these trading experiments, summarized by Blauert (1999, p. 172), it has been shown that signals with most content below 1.6 kHz are dominated by IATD/IPD. In this condition, 40 μs of time difference is equal to 1 dB of ILD. For signals above 1.6 kHz, IPD no longer has an impact. In fact, only IETD and ILD will create changes in localization, where up to 200 μs of IETD is needed to adjust for only 1 dB of ILD.

It is also insightful to discuss localization blur (MAA) because it essentially represents a summary of the localization cue perceptual salience. In other words, a smaller MAA suggests more dominant localization cues and vice versa for larger MAAs. Although, MAA is considerably smaller for complex signals with wide bandwidths than it is for simple tones or narrow band sounds. This is because as bandwidth increases, the number and agreement of the localization cues can also increase. This creates a stronger, more definite sense of where a sound is coming from.

With regards to frequency, it has already been discussed that MAA is the largest in the middle frequency range. In fact, for tones in the horizontal plane, Mills (1958) found the MAA to be maximum between 1000-3000 Hz while minimum from 250-1000 Hz and 3000-6000 Hz. Similarly, Stevens and Newman (1936) showed the middle frequency range of 2000-4000 Hz to be the hardest range to notice changes in azimuth.

Other experimenters have made similar comments regarding the degradation of localization performance in the middle frequency bands as opposed to higher or lower frequency ranges (Perrott, 1969; Perrott and Tucker, 1988; Pulkki and Karjalainen, 2001; Grantham, 1984).

Regardless, it is probably most important to keep perspective on the relative salience of all the available localization cues. Middlebrooks (1997) states that localization research has "led to a general acceptance of the notion that interaural difference cues provide the principal cues to the horizontal dimension and that spectral shape [monaural] cues provide the principal cues to vertical and front/back localization" (p. 78). Also, Fisher and Freedman (1968) mention that while pinnae cues may be useful for motionless (fixed head) experiments, they are of little importance in realistic conditions for binaural individuals. They showed that listeners who are free to move their heads, yet without pinnae cues (listening through tubes), can correctly localize free field sounds.

Wightman and Kistler (1997) also support the dominance of interaural cues. In particular, they have warned against performing monaural experiments using binaural individuals with an occluded ear. Their experiments show that even slight levels present in the occluded ear causes a dominance of interaural time difference. Thus, subjects tend to ignore both the monaural spectral cues under investigation and the unnaturally created ILD cues due to occlusion, and instead rely solely on ITDs.

Therefore, the three types of localization cues seem to perceptually rank with interaural time differences being the most dominant, followed by interaural level differences, and finally monaural spectral cues. With respect to localization versus frequency, the middle frequency range is the most difficult to localize with low frequencies being dominated by ITD cues and high frequencies by ILD cues. Also, monaural pinnae cues present in the 5-12 kHz range are useful in some situations, mostly for avoiding front/back confusion and estimating the distance of a sound.

All of the localization cues vary both in physical and perceptual significance over the audible frequency range (see Figure 13). While their variations do have complex patterns that change with frequency and azimuth, generalities can be made from the results of localization experiments and physical measurements. Specifically, physical ITDs sharply drop while ILDs sharply rise with increasing frequency. Perceptually, IPD's influence begins to decrease around 800 Hz, having no effect above 1.6 kHz. ILD has a lesser but relatively stable influence, with a slight peak around 2 kHz (Blauert, 1999, p. 158). Pinna cues are of minimal importance in most real world conditions, but occur in the range of 5-12 kHz.

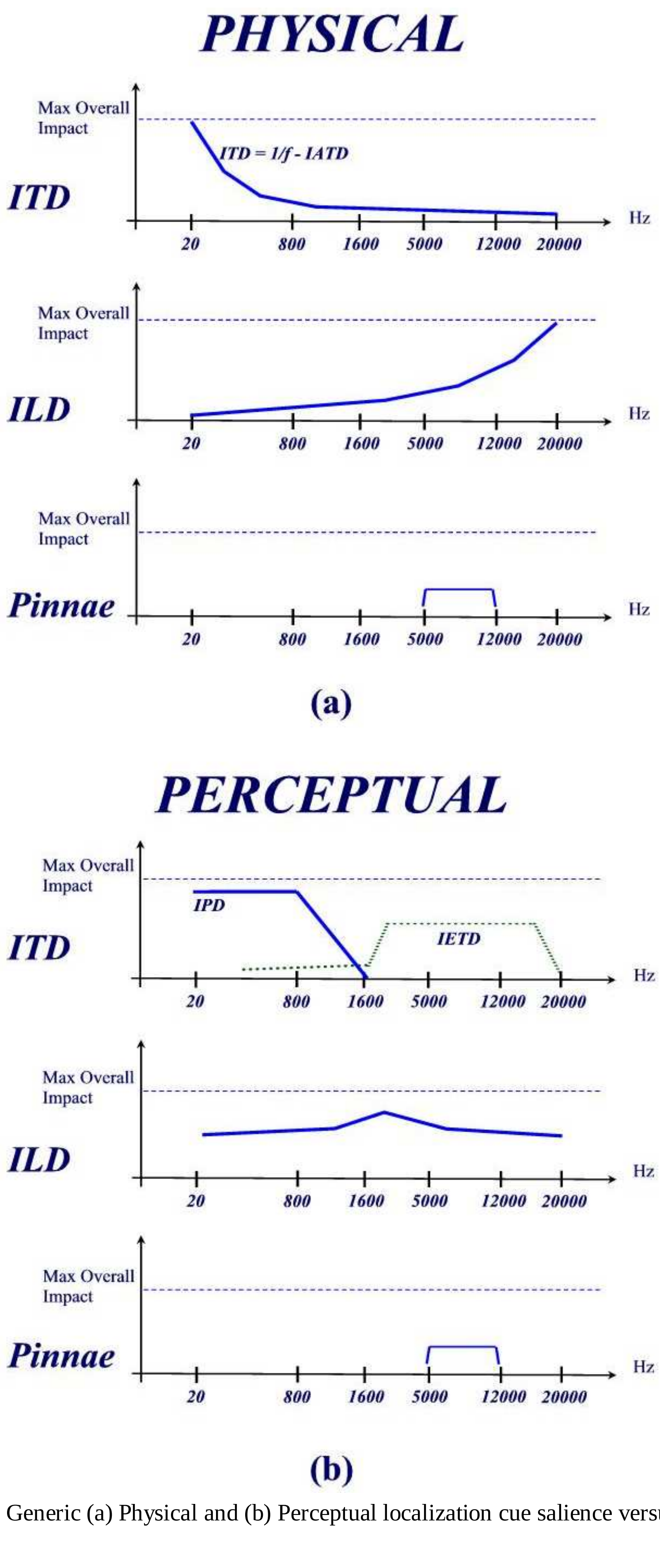


Figure 13: Generic (a) Physical and (b) Perceptual localization cue salience versus frequency

Chapter 4: Auditory Scene Analysis

In the presence of multiple sound sources, listeners can sometimes isolate their attention to one source while ignoring the others. At other times, sources might be perceptually integrated to create auditory illusions (images) that occur away from any of the actual source locations. This occurs often when listening to stereo loudspeakers. Bregman (1999) has studied these behaviors and offers the term "auditory scene analysis" to generically describe the perceptual mechanisms which allow us to interact with complex listening environments.

The most important of these mechanisms is probably the "precedence effect." This unique characteristic of our auditory system allows us to hear a sound while ignoring any close temporal replications. This most commonly occurs in the echoes of a reverberant acoustic space. An understanding of the precedence effect allows a higher-level discussion on the topic of auditory stream segregation. Each sound source is considered to be a stream of information to the listener. Depending on the characteristics of those sound streams, varied levels of integration or segregation will occur. Finally, an introduction to the influences of an acoustic space will be discussed.

The topic of auditory scene analysis applies to this research in that one must understand the consequences of spatially relocating frequencies. Moreover, studying the characteristics of this behavior will provide insight as to how far the sources might be separated, or which frequencies might be more easily relocated before segregation occurs.

Precedence Effect

Discussions of the "precedence effect" (Wallach, 1949), "Haas effect " (named after Helmut Haas' 1949 dissertation), or the "law of the first wavefront" specifically refers to the human tendency to perceive only one sound event when two (or more) sequentially reoccurring sounds occur. The most common example of this phenomenon is found when listening in a reverberant room. Here, sound reaches the listener both directly from the sound source as well as from varied directions due to wall reflections. Yet, only one sound is typically heard and in only one direction.

Blauert (1999) as well as Litovsky, Colburn, Yost, & Guzman (1999) have presented summaries of the theory and results of many precedence effect experiments. Here, "clicks" tend to be the most commonly used test signal because they provide a wide spectral bandwidth and definite temporal presentation to the listener. Litovsky et al. have further categorized precedence effect experiments as either auditory fusion, localization dominance, or lag discrimination. Lag discrimination is not particularly relevant to this research and will not be discussed.

Auditory fusion experiments present sequential clicks with a silent interval in-between them. The goal is to identify signal characteristics that will change the perceived number of sound events from one to two. Clicks with <5ms between them create only one auditory event, where the two-event threshold is around 8-10 ms (Litovsky et al., 1999). A more common term for this performance is called the "echo threshold," referring to the time difference necessary to create an audible echo.

With the two sequential sounds fused into one image, it is interesting to next consider the conditions that can affect the overall localization of the image. This is probably the most relevant to this research and is called "localization dominance." Scientists study localization dominance by varying the interaural differences (ITD and ILD) between two sequential sounds and asking listeners to comment on its perceived location. This has already been discussed in early sections, and will not be presented in detail here.

However, one particularly interesting aspect of localization dominance is the "Franssen effect." With a standard stereo loudspeaker setup, a short burst is played from the left speaker and a sequential longer burst plays from the right speaker. Under this condition, the listeners will hear a short burst at the left, followed by a diffuse sound coming from between the loudspeakers. In essence, the short leading sound has caused the longer lagging image to be perceived as spatially diffuse, even though it comes from only the right speaker. Litovsky et al (1999) additionally comment that this illusion will not occur with tones of high or low frequency, but rather requires sounds that are difficult to localize (suggesting mid frequencies) (p. 1638).

Auditory Stream Segregation

Conceptualize each sound source in a multi-source setup as a stream of information to the listener. Under certain conditions, the streams can be perceptually integrated into one overall auditory event. However, each stream contributes characteristics to the event and can influence the sound in many ways.

In general, streaming research is classified as either sequential or concurrent auditory streaming (Yost, 1994). Sequential streaming experiments present several nonoverlapping sound events whereas concurrent streaming presents simultaneous events. For this thesis, concurrent streaming is more applicable because the spatially relocated frequencies are presented at the same time as the other sounds. However, sequential streaming is a more severe condition and can better exemplify the signal characteristics that will cause stream segregation.

Through psychoacoustic experiments, it has been found that those characteristics most affecting auditory streaming include the temporal interrelationship, relative similarity of fundamental frequencies, spectral distribution (harmonics) and the perceptual location of the sound sources (Bregman, 1999). Obviously, those events that occur at the same point in time are likely to belong to the same stream source. Also, increasing the time between events is more likely to suppress the second sound (precedence effect) or eventually associate it with a separate auditory event.

In addition, those sounds with common spectral characteristics are more likely to be integrated. On the other hand, having different fundamental frequencies or timbres (harmonics) will most likely to lead to stream segregation. This point is particularly relevant to this thesis, because the greater the perceived spectral change of the signal, the greater the chance the auditory image will split. However, the spatially relocated sounds will still share the same basic harmonic structure of the original signal.

Finally, the influence of the perceived location of the sound sources should be considered. Bregman (1999) states that "one of the most powerful strategies for grouping spectral components it to group those that have come from the same spatial direction and to segregate those groups that have come from different directions" (p. 658). In a multi-source situation, the brain will use the physical interaural and monaural localization cues in an attempt to interpret a spatial location for each stream of sound. If the streams are the same due to physical proximity, or otherwise present a perceived image with a shared spatial origin (i.e. correlated left and right ear signals, as in stereo listening), the brain will integrate them into one event. However, strong differences in localization cues will cause the two streams to segregate. The term to describe these kinds of phenomena is called summing localization (Blauert, 1971).

Summing localization research performed by Gockel, Carlyon, and Micheyl (1999) used several sequential band-limited harmonically complex sounds over headphones. Their focus was to determine if perceived location (through changing interaural differences) would have an impact on auditory streaming. They found that presenting the second sound with interaural differences increased the subject's segregation tendencies. However, the high frequency region (3900-5400 Hz) showed the least amount of segregation as compared to the mid (1375-1875 Hz) and low (125-625 Hz) frequency regions. This suggests that interaural changes at higher frequencies are less likely to cause stream segregation than those same changes at middle or lower frequencies.

Along similar lines, Thurlow and Marten (1962) performed sequential streaming experiments on high pass noise (>2000 Hz) coming from two loudspeakers in free space. They continually increased speaker separation and found that listeners perceived one source of sound until approximately 6.4° of separation, where 50% of the listeners perceived one steady sound and one intermittent sound (suggesting partial fusion) (p. 1858). Further increase in speaker separation eventually caused the sound to split into two unique streams.

Finally, some interesting commentary on concurrent streaming experiments can be found in Gardner (1973). Here, he provides a technical review of auditory illusions caused by multiple sound sources radiating similar signals. He discusses how spatially separated loudspeakers radiating signals of similar quality will create fused images (auditory integration). He continues, stating that even if one speaker radiates low pass noise and the other radiates high pass noise, both will fuse into the perception of a single source of full bandwidth noise. This example illustrates that the "quality" and general agreement of the sound sources can result in auditory fusion despite differences in the source's spatial and spectral content. Gardner reasons that these streaming and fusion effects are largely due to the precedence effect. As will be seen later, this result has a direct implication to this research, where spatially relocated frequency bands of white noise are integrated into one defined image.

The Acoustic Space

Discussing all the nuances of acoustics is not necessary or relevant to this research. However, it is important to realize that an acoustic space can influence the results of listening experiments. Any reverberation or spectral variations caused by the acoustic space will essentially change the perceived listening material. This is why many of the discussed localization experiments have been performed in anechoic (relatively free from reverberation) environments. However, it is interesting to consider how the room itself affects localization.

Hartmann (1983) concedes that localization is predominately determined by the interaural and monaural characteristics of the incident waveform. In addition, any secondary waveforms such as echo reflections are most likely suppressed by the precedence effect. However, he points out that the historical body of research on precedence has been performed either in free field, with headphones, or via paired-loudspeakers in anechoic chambers. His research therefore contributes experimental data regarding the influence of room geometry and wall absorption on horizontal localization.

Using the Espace de Projection (ESPRO) variable-acoustic concert hall in Paris, Hartmann (1983) ran several test signals through different room conditions (absorbing, reflecting, and low ceiling). Of particular interest is that for impulsive sounds (50 ms sine bursts), localization showed no significant changes due to the different room types. However, for non-impulsive sinusoids (6-10 sec. rise times), significant localization blur did occur. With tones of 200, 500 and 5000 Hz, he showed that only the 5000 Hz signal could be localized with moderate accuracy. Both the 200 Hz and 500 Hz signals had errors suggesting random guesses. Other non-impulsive complex tones and broadband noises were also used in his experiments, which were localized significantly better than the steady-state tones.

Hartmann (1983) concludes that the localization of non-impulsive low frequencies appears to be deteriorated due to room acoustics. This is in agreement with other papers which have stated that non-impulsive sinusoids are in general the most difficult test signals to localize (Wagenaars, 1990; Rakerd & Hartmann, 1985; Hartmann, 1993). Thus, influences of room acoustics are probably the most likely explanation of why most people claim that low frequencies are "hard to localize."

Chapter 5: Experimentation

Having detailed the background theory on the various localization topics this thesis surrounds, it is now important to lay out the experiments used in this research. This includes an explanation of the listening tests and the thoughts that went into designing them. The following will present the experimental conditions as well as the test signals and variables used for this research. Also, the equipment and the test methodology will be discussed.

Experimental Conditions

The goal of the listening tests was to identify whether low or high frequencies have a greater impact on the localization of a stereo image. In essence, the listeners would be asked to comment on the change in position of the image between two different conditions. The two conditions were created by relocating different frequency bands (i.e. low vs. high) of the left stereo signal to a new spatial position.

With this in mind, the devised test setup consisted of a right speaker (R) and its complementary left speaker (L) at the same distance (2m) and symmetrical angle from midline. A third speaker, called the "Spatially Relocated" (SR) speaker was positioned at the same distance to the listener, but closer to midline than the left speaker by a small azimuth delta (see Figure 14).

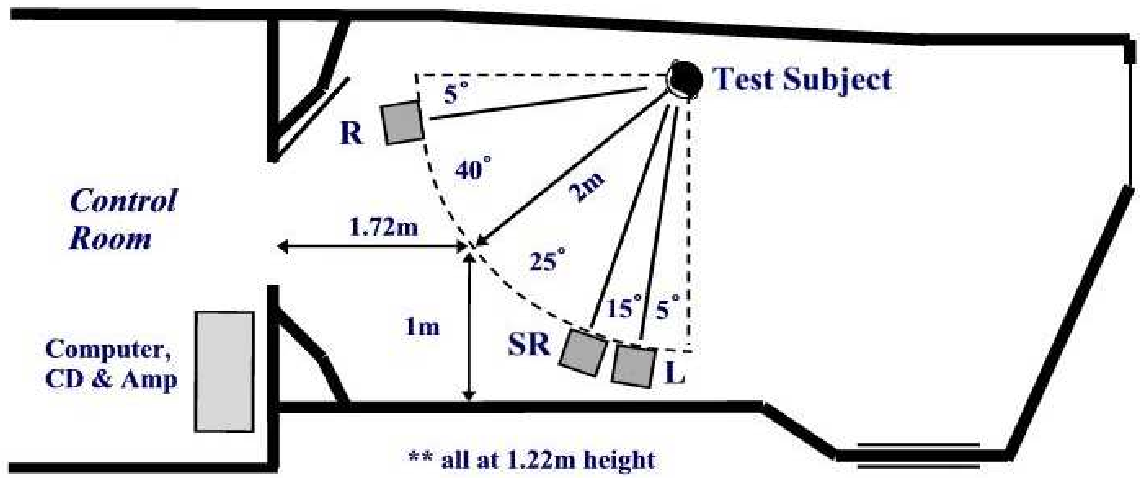


Figure 14: Physical setup of test

The stereo speaker angular separation was partially based on this thesis' ties to automotive applications. Thus, it was desirable to have them wider than a typical stereo system, which Blauert (1999) suggested as ± 30º from midline. With some additional constraints from room and positioning factors, the resulting speaker separation was selected at L/R of ±40º from midline.

The relative position of the SR speaker was also important. There needed to be enough angular separation to cause shifts of the stereo image; yet not too much distance such that the L and SR signals might segregate. Thus, the SR speaker must be definitely outside the typical MAA. Several papers on this topic showed a threshold to be in the area of 5-10º (Mills, 1958; Stevens & Newman, 1936). Thus it was felt that 15º would be far enough to be noticed without splitting the L and SR auditory streams.

Test Signals

With the setup decided, it was next important to choose test signals that would give useful data. The spectral content and distribution of the signal was the most crucial factor. It was felt that a wideband signal with a strong center image would make an ideal test track. However, this is difficult to find in popular music, which is why white noise was chosen as the primary test signal.

Because white noise is not particularly interesting to listen to, a music track was also used in an attempt to collect more data while keeping the listener's attention. Of course, it was expected that the results of the music track might be difficult to explain because of the many variables that music introduces, such as time-varying spectral content.

The listener was tested with the music passage first because it contained stereo images which were easier to conceptualize than those created by white noise. It was thus desirable to find a music track with a strong, consistent central stereo image and a reasonably wide bandwidth. A fairly simple, almost monophonic sound stage would also make it easier to notice shifts in the image's position. The female voice seemed to reasonably comply with all of these constraints, thus an 8 second clip (0:27-0:34) from Joan Baez's "Diamonds and Rust" was used.

In order to present a comprehensive description of this track, consider the spectrogram in Figure 15. The amplitude of the signal is represented by relative color intensity as shown in the color bar indicator, while time is shown on the x-axis and frequency on the y-axis. This was performed using Matlab (see Appendix C for code). The total energy of the L/R channels was also calculated for the entire signal (Figure 16) and the various subbands (Figure 17), including the grouped subband representations of CD and CDE (Figure 18).

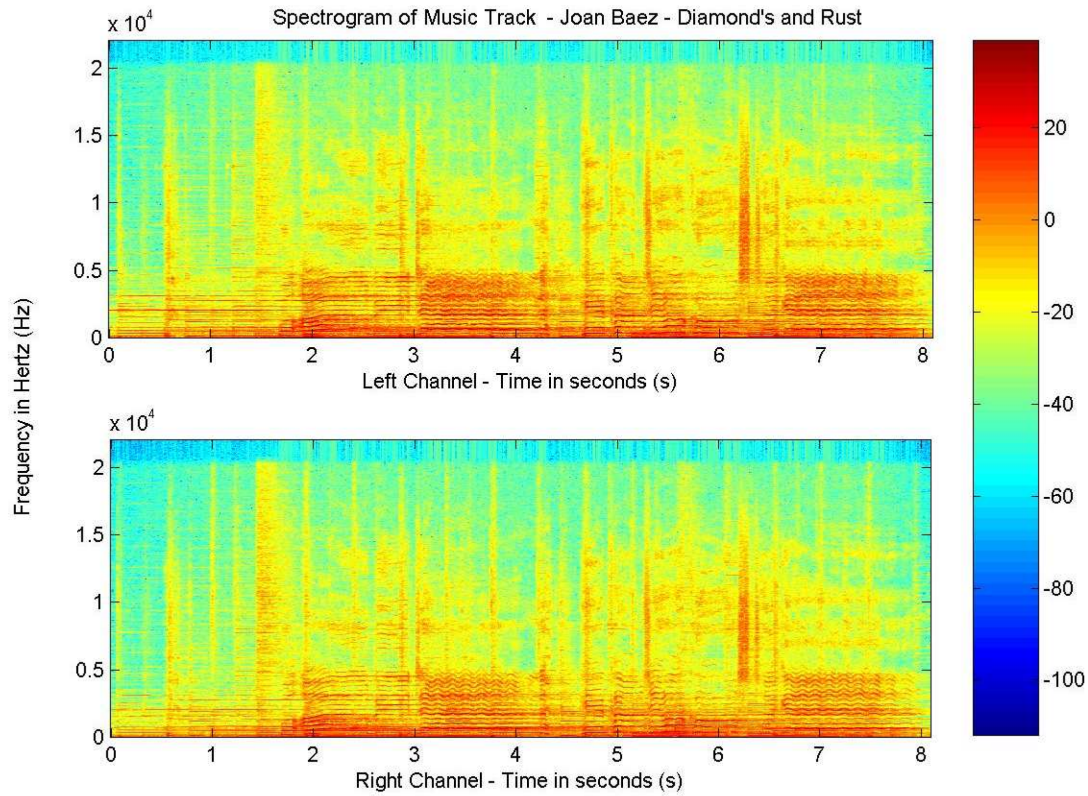


Figure 15: Spectrogram of music passage L (top) and R (bot)

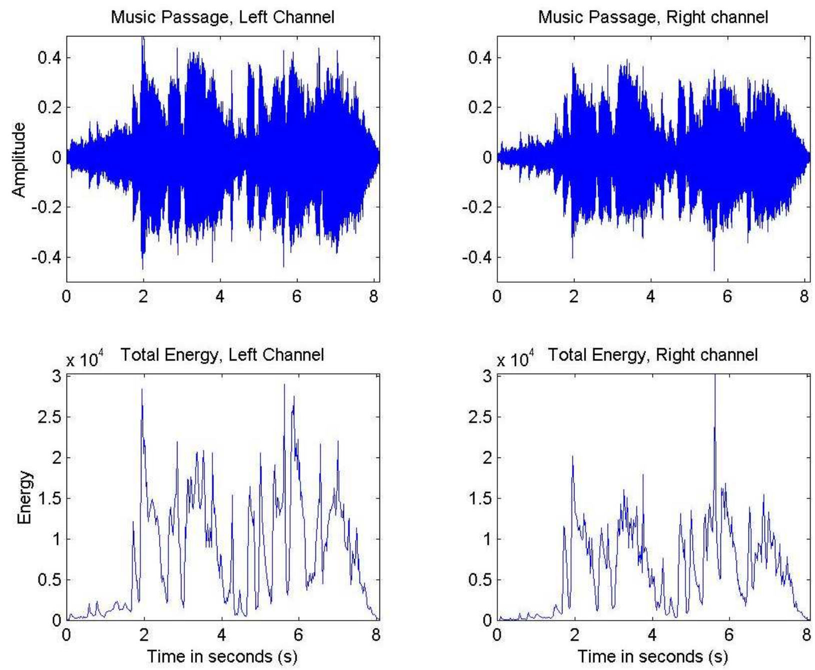


Figure 16: Music passage temporal (top) and total energy (bottom)

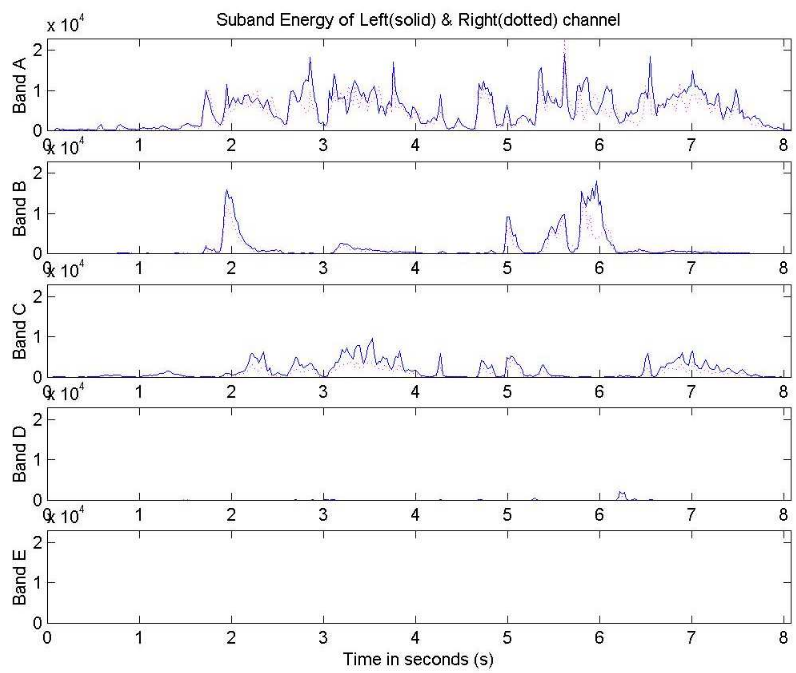


Figure 17: Music passage's subband energy

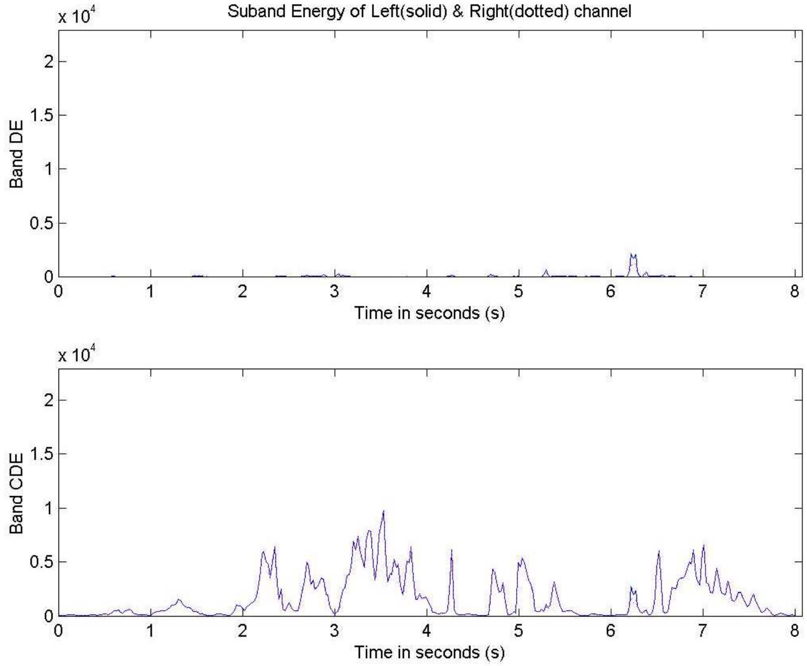


Figure 18: Music passage's combined subband energy

Moving on to the white noise test signal, it was decided that short noise bursts would be even easier to localize that a continuous noise segment. This is because of the additional transient localization cues that occur. Therefore, six 250 ms bursts of white noise, with 20 ms onset/offset ramps and 300 ms silent intervals, were used as the primary test signal. The spectrogram of the noise bursts can be seen in Figure 19.

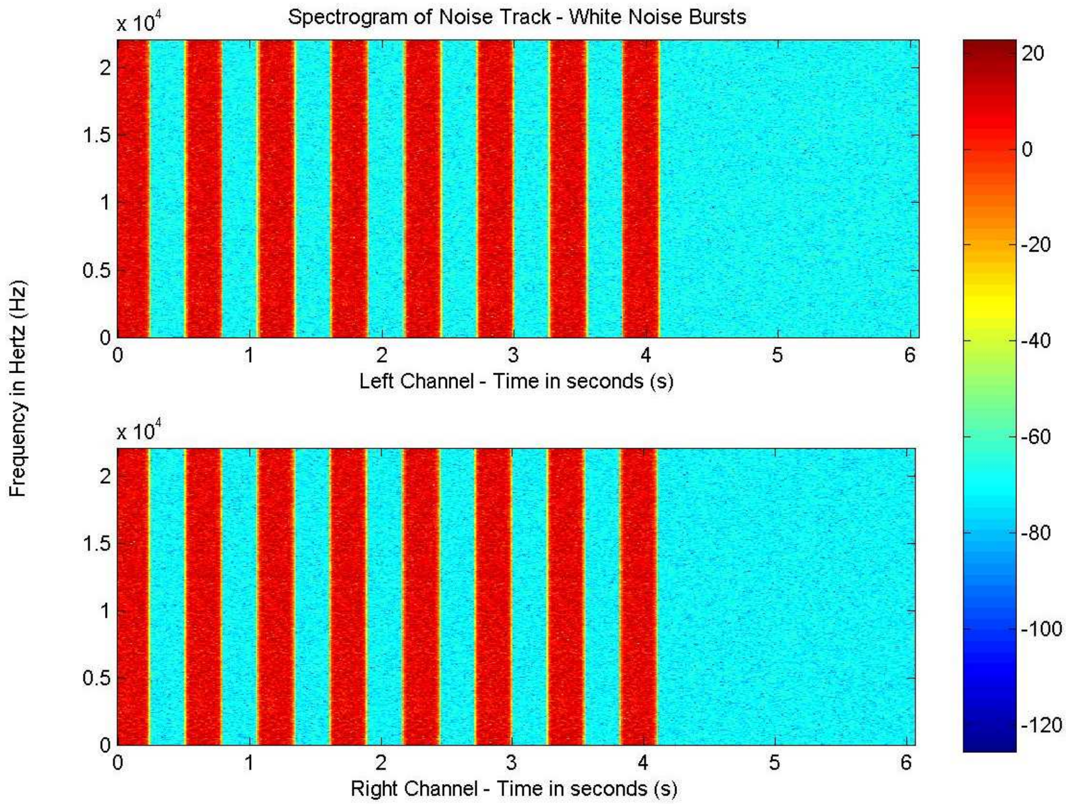


Figure 19: Spectrogram of white noise passage

The next decision involved choosing which portion of the test signals would be relocated to the SR channel. During the background localization research, it was noticed that the localization cues were often discussed in terms of the range of frequencies they were most effective in. Therefore, it seemed reasonable that these points would create bands of frequencies that were known to have a dominant localization cue. This resulted in defining the following bands:

- Band A = 20-800 Hz
- Band B = 800-1600 Hz
- Band C = 1600-5000 Hz
- Band D = 5000-12000 Hz
- Band E = 12000-20000 Hz

Recall that localization cues are generically categorized into ITDs, ILDs, and pinnae effects. Blauert (1999) states that the localization effects of IATD/IPD greatly decrease above 800 Hz, ultimately having no effect above 1600 Hz. This led to the development of bands A and B. Next, the pinnae effects are thought to have the most impact from 5-12 kHz, which is what band D represents. Band E represents the highest frequency band and is also a conveniently close to the range that Smyth (1999) discussed in his paper on DTS' Coherent Acoustics. Band C was a remnant of the other bands, but is also known to contain monaural cues caused by the torso.

Experimental Variables

It is necessary to provide some insight into the variables used in the experiments. Potential variables included loudspeaker position (horizontal vs. vertical), signal intensity, listening material, acoustic space, and loudspeakers. As mentioned, this thesis would focus on localization cues versus frequency. Thus, it was decided that the SR frequency bands would be the lone variables.

Yet, five frequency bands still create an excessively large number of combinations to potentially test in a span of thirty minutes. Again keeping with the theme of low vs. high frequencies, the following ten test conditions were implemented:

- E vs. Stereo (STR)
- E vs. A
- E vs. AB
- DE vs. Stereo (STR)
- DE vs. A
- DE vs. AB
- DE vs. ABC
- CDE vs. Stereo (STR)
- CDE vs. A
- CDE vs. AB

Test Methodology

For each of the conditions listed above, the listener would be sequentially presented with two versions of the signal and asked to comment on the movement of the stereo image. Thus "E vs. A" would first present "band E" playing from the SR speaker while "Left - E" played from the L speaker. This would be followed by a short silent interval. The listener would then be presented, in this case, with "band A" playing from the SR speaker while "Left - A" played from the left speaker. Again, the R speaker always played the original right stereo signal.

To reiterate this technique, consider Figure 20. This shows how certain bands of the L signal are relocated to the SR channel. In fact, the speaker configuration previously shown in Figure 14 is ideal because without the SR channel, a solid center image can also be presented to the listener. However, the test's primary function was to move spectral energy of the L channel to the SR speaker in order to move the stereo image. Of course, the presumption was that the stereo image would shift to varying degrees to the right, based on the competing L and SR localization cues.

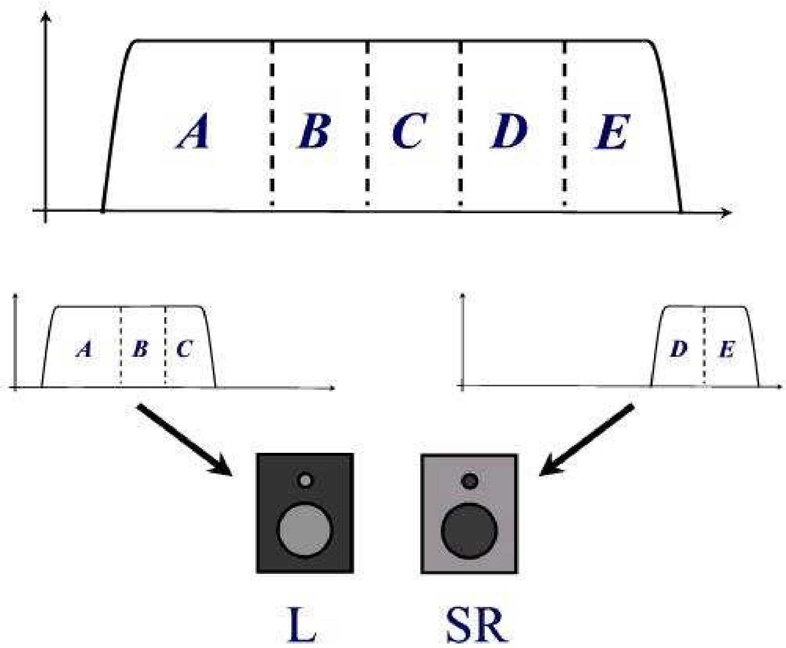


Figure 20: Division of Frequencies between L and SR speaker

With the experimental details and variables decided, it was necessary to also standardize the process of the experiment. Thus, before the test subject entered the room, the speaker configuration was hid behind a curtain. Upon entering the room, the listeners were asked to sit in a fixed chair and wear sunglasses with opaque lenses. A laser-pointing device was used to align the listener's ear canal with the top of the speaker's woofer. The curtain was then removed, placing the listener at the midline of the stereo (L/R) speaker array (see Figure 14). They were then read the following passage:

> I have put you in an ideal listening position. Please try to keep from rotating the chair, or moving your head or body. These listening tests are investigating the human ability to locate sounds. This is generically called localization. As you may have noticed in your casual music listening, a stereo system can recreate realistic audio "images" where the singer, or instrument, sounds as if it is coming from some point in front of you, but between the actual loudspeakers.

> For instance, I will play you a short passage of music to illustrate. Notice that the piano sound is directly in front of you, the triangle comes from your left and the cello comes from your right. [play track 1]
>
> Now listen to a modified version of that track. Notice that the piano is no longer directly in front of you, but has moved off to your right. These are the kinds of differences I am going to ask you to pay attention to [play track 2].
>
> To begin with, I will give you a couple of practice runs. I ask that you pay particular attention to the female voice in the center. I will play you two versions of this clip, with a short pause in between them. Then I will ask you to comment on the position of the singer in the second clip relative to the first.
>
> Realize that it is the RELATIVE position, left/right/up/down/far/near or any combination of these, which I am asking you for. Thus, if both images sound as if they are to the right of center, I am asking you to tell me if the second image you hear has moved to the left or right, or other direction, of the first image. If they don't change, simply tell me they are the same.
>
> You may also experience a feeling where part of the sound comes from one location and another part of that same image seems to come from a different location. This is called a "split" image, and I would ask that you tell me when you notice this occurring.

As mentioned, the musical passage was played first because it was easier to identify localization shifts. This was followed by the white noise bursts test signal. For each test subject, the variables (varied bands played from the SR channel) were randomized within each set of trials. The Arcade DSP Amplifier presets were used to switch between the variables while the CD player would repeat the same track twice per trial. For a brief description of the Arcade Amplifier's capabilities, see the following section.

However, because the amplifier was limited to 8 presets (and 10 variables were desired), the tests were further broken into two sub trials (see Table 1). The "DE vs." variables were repeated partially to even out the two sub trials, but also because it was felt this comparison was the crux of the thesis. Obviously more trials provide a higher statistical confidence to the results.

Subtrial 1	Subtrial 2
E vs None (STR)	* D+E vs None (STR)
E vs A	* D+E vs A
E vs A+B	* D+E vs A+B
* D+E vs None (STR)	* D+E vs A+B+C
* D+E vs A	C+D+E vs None (STR)
* D+E vs A+B	C+D+E vs A
	C+D+E vs A+B

Table 1: Trial subband variables

Test Equipment

The room used throughout these experiments was located in the top floor of the Gusman Concert Hall on the university's campus. It is affectionately called the "dead room," because of its pseudo-anechoic response created by acoustical treatment on the walls. The general dimensions and shape of the room can be seen in Figure 21, while Appendix D has a measured frequency response.

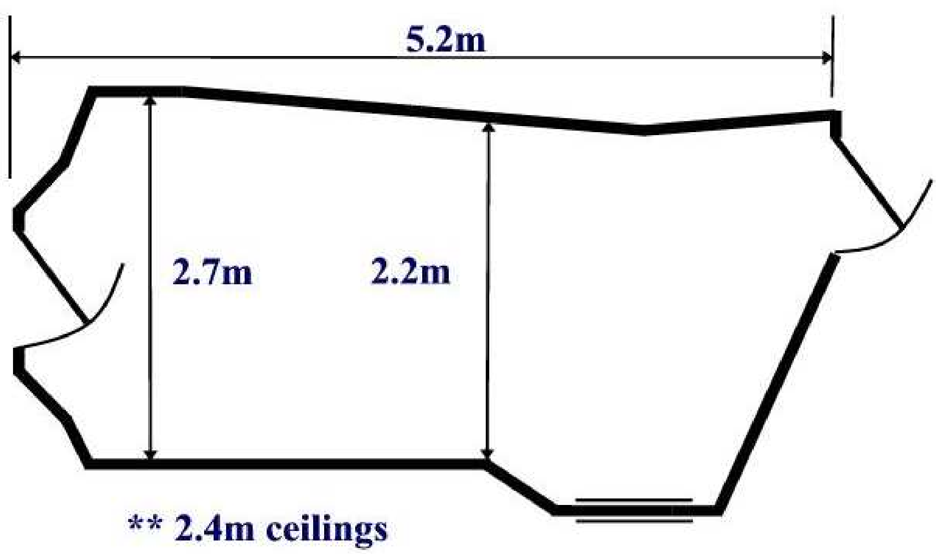


Figure 21: Gusman "dead" room basic dimensions

Next, the following equipment was used for the listening experiments:

- **Loudspeakers**: Miller & Kreisler (M&K) model MPS-1610 loudspeakers are a 2-way near/mid-field monitor with a 1" tweeter and 6.5" (4Ω) woofer. Their frequency response is 80Hz-20kHz ± 2dB, with a passive crossover at 1.2 kHz. MSRP is $650 each. See Appendix E for a measured frequency response.
- **Speaker Stands:** Studio Tech SN-A adjustable metal satellite speaker stands
- **CD Player:** "HP CD-Writer Plus" CD-ROM drive was used to play audio CDs directly to the Power Amp via line level output.
- **Amplifier:** Proprietary DSP-based power amplifier intended for multi-channel car audio. Accompanied by Windows-based "Arcade v1.2" software provided by Kevin Heber of Delphi Automotive Systems. Allowed using stereo input to distribute and manipulate multi-channel audio outputs.
- **Power Supply:** Hewlett Packard (HP) supply up to10A.
- **Windows Personal Computer (PC):** Windows-based PC to run Power Amp software and CD player.

The electrical setup of the experiment is shown in Figure 22.

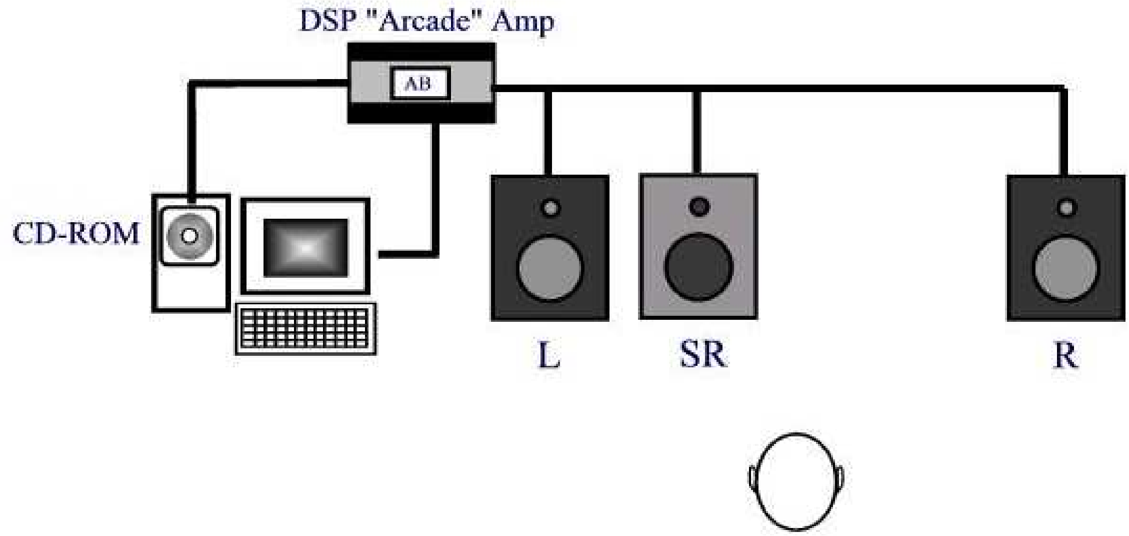


Figure 22: Electrical schematic of test setup

The DSP-based amplifier was a critical piece of equipment, used for both developing and testing ideas as well as the final listening experiments (see Figure 23 and Figure 24). It has the ability to perform multi-input and multi-output frequency and temporal signal processing of up to four stereo inputs and twelve stereo outputs. Besides signal processing, it is also a power amplifier with the capability of 12x35 watts into 8 ohms at < 1% THD+N.

The amplifier had a serial interface to a PC, which ran a software program called "Arcade." The user interface allowed the input/output and signal processing parameters to be varied and tested. It also allowed up to 8 different preset amplifier configurations to be instantly changed in real time. This was used during listening tests, to change the variables of the experiment between test tracks.

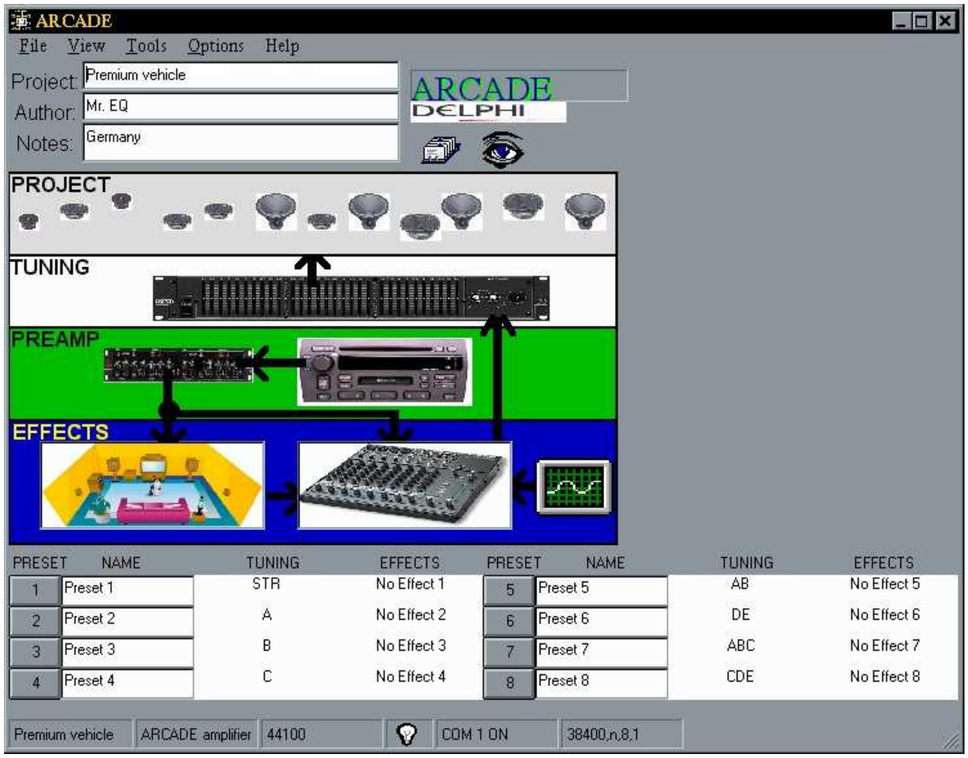


Figure 23: Arcade's Main Program Screen

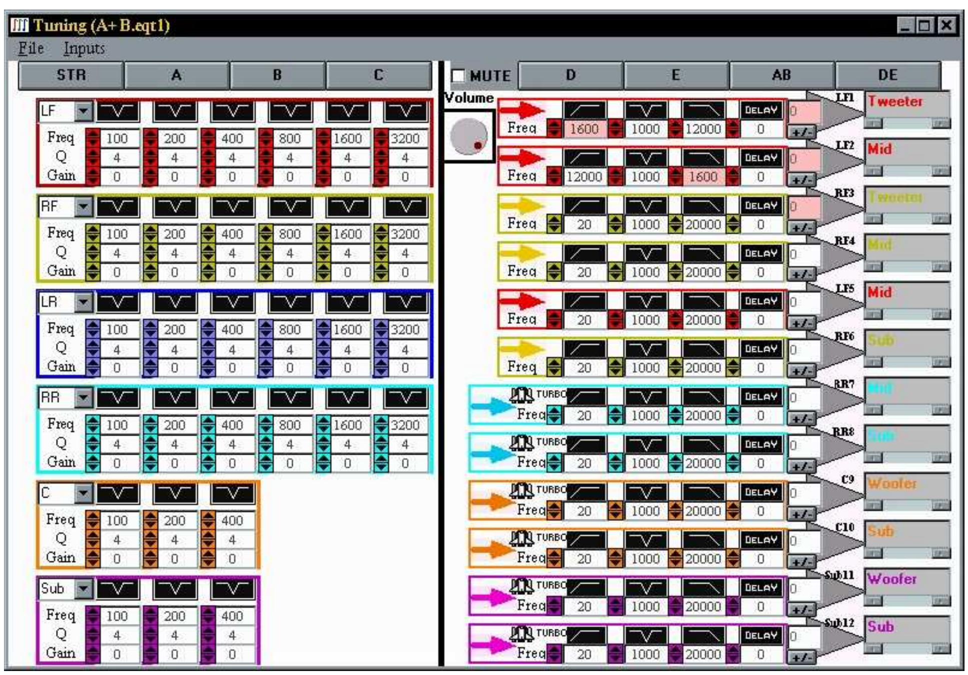


Figure 24: Arcade's Amplifier Configuration screen with 8 selector bars (at top)

Chapter 6: Results and Analysis

Fourteen test subjects completed the listening tests previously described. They were all students in the Music Engineering or Audio Engineering programs at the University of Miami. All were assumed to have normal hearing abilities, but should be considered untrained listeners. However, most were musicians and thus could be considered to have above average abilities from the typical untrained listener.

The test subjects were asked to comment on the relative movement of a centrally located audio image caused by spatially relocating various bands of the left stereo channel. They were not limited in the vocabulary of their response, however their answers were interpreted and entered into 8 different image movement categories (or combinations of): No Shift, Right, Left, Up, Down, Near, Far, Split. Occasionally, due to time constraints a few of the trials would be skipped, which caused some variation in the number of total test trials.

Experimental Results

The following plots include a graphical representation of each listening test (see Appendix A for complete results). They begin with a summary of the "music track" results, followed by a plot for each individual SR band that was tested with the music track. Then, the summary plot of the "noise burst track" is presented, similarly followed by the results of each band for the noise track. Obviously, the number of "no shift," "left," and "right" responses were of primary interest for this research. However, "up" and "down" responses were occasionally mentioned, which is why they were included in the two summary plots.

The x-axis groups are presented such that the responses shown represent the image shift caused by the second listed frequency band. For example, Figure 26 shows the results of moving band E to the SR channel first and then comparing that image position to one caused by relocating either: nothing (Stereo, STR), band A, or band AB. The first grouping in this plot shows the results of relocating E vs. STR for 14 test subjects. In this case, 3 subjects felt the stereo image was to the "left" of the E image, 1 subject said it was to the "right," while the majority (9) felt the image did not shift.

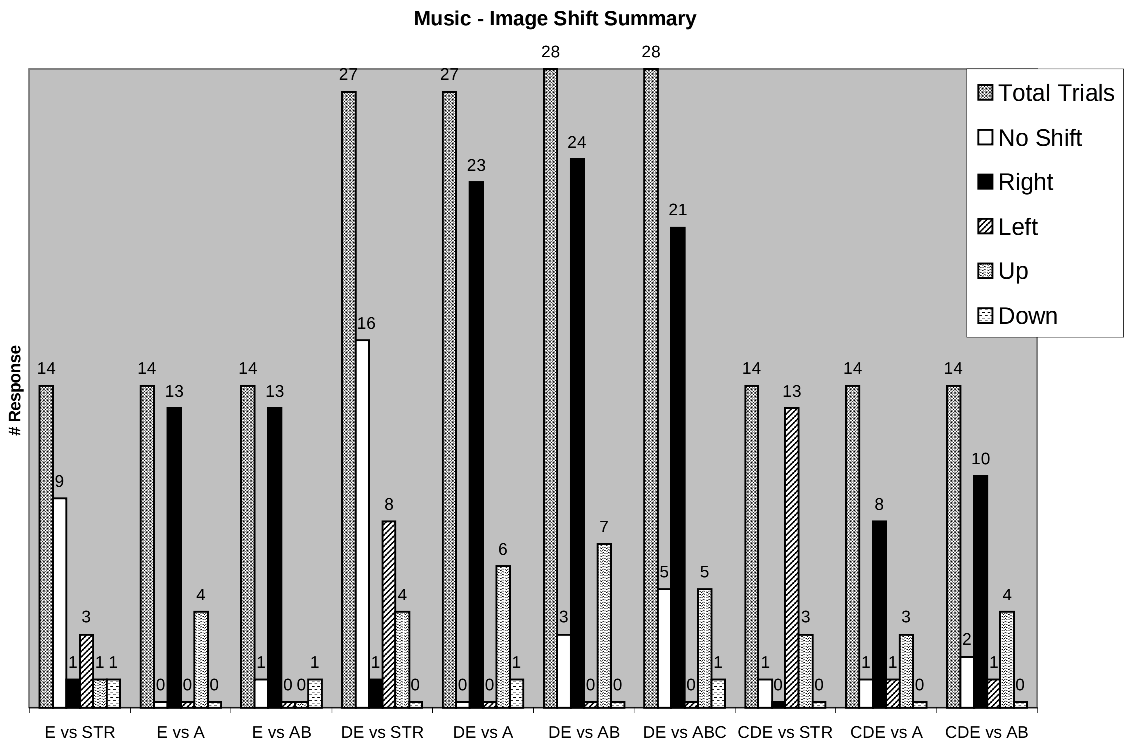


Figure 25: Music Image Summary

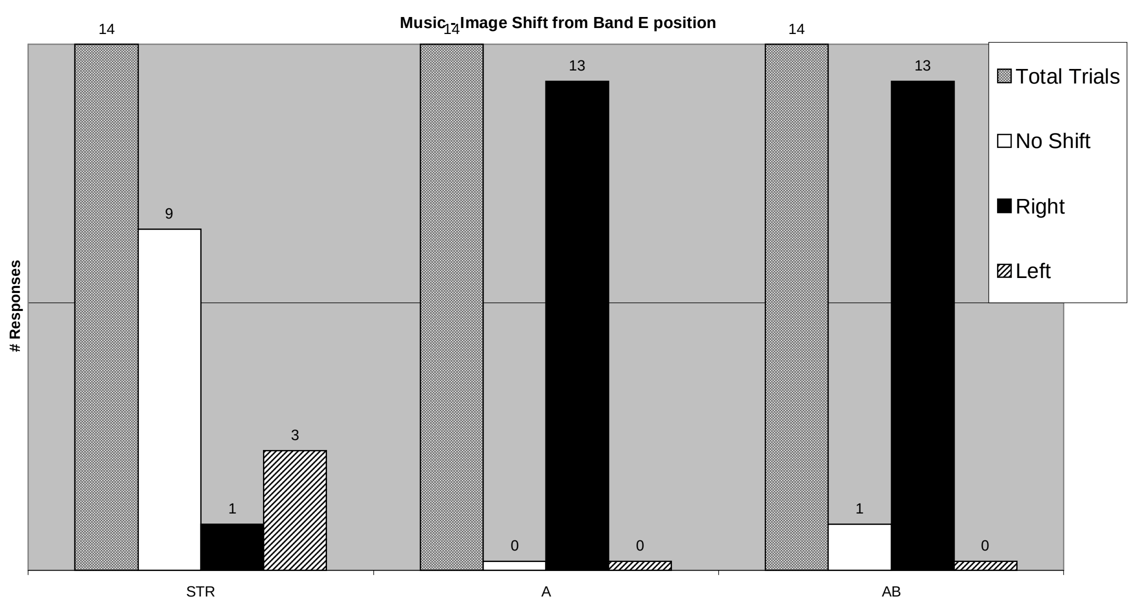


Figure 26: Band E vs. Music Image

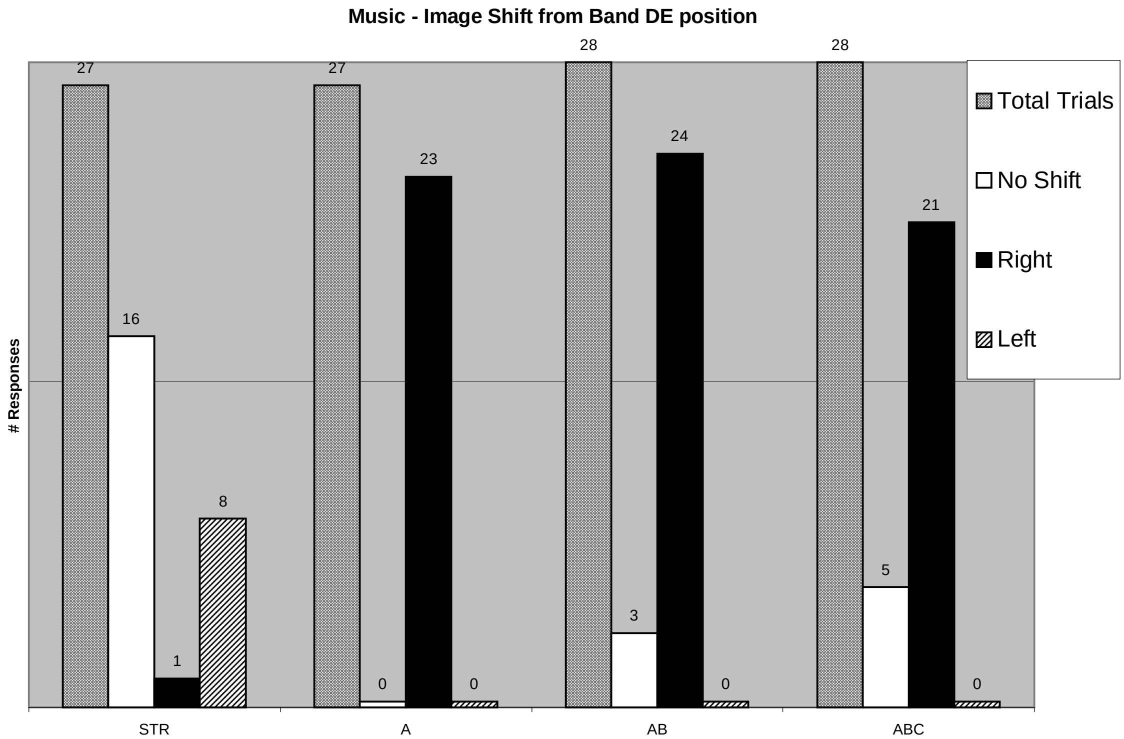


Figure 27: Band DE vs. Music Image

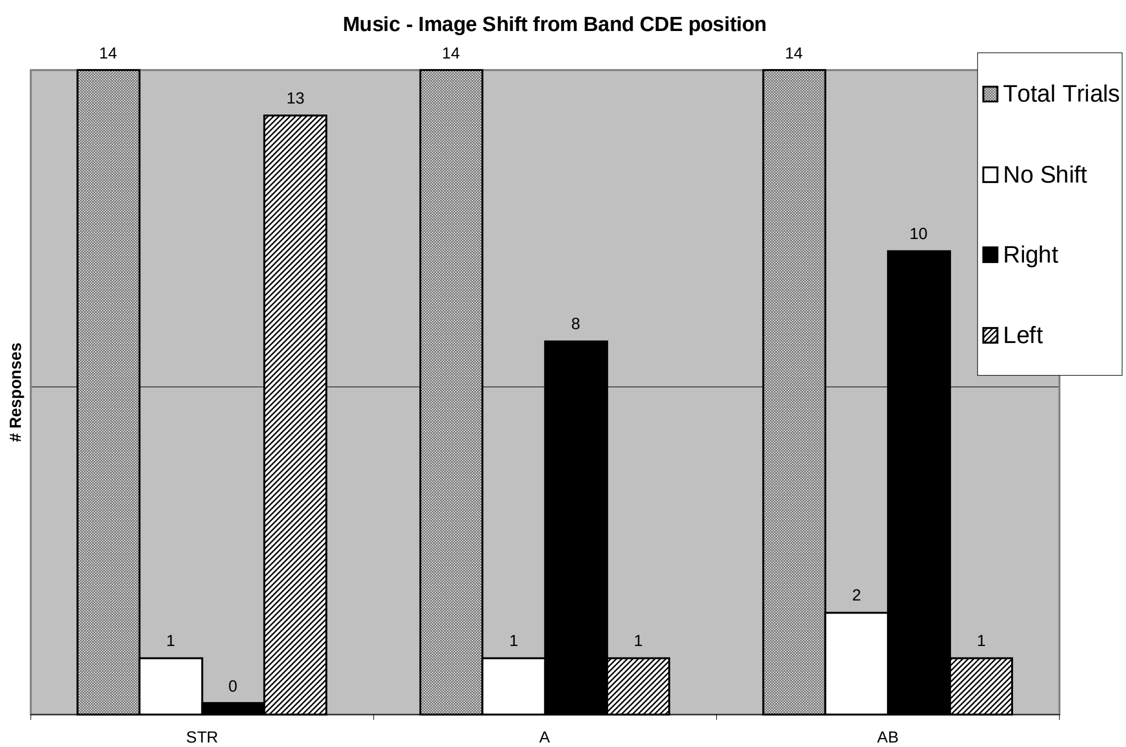


Figure 28: Band CDE vs. Music Image

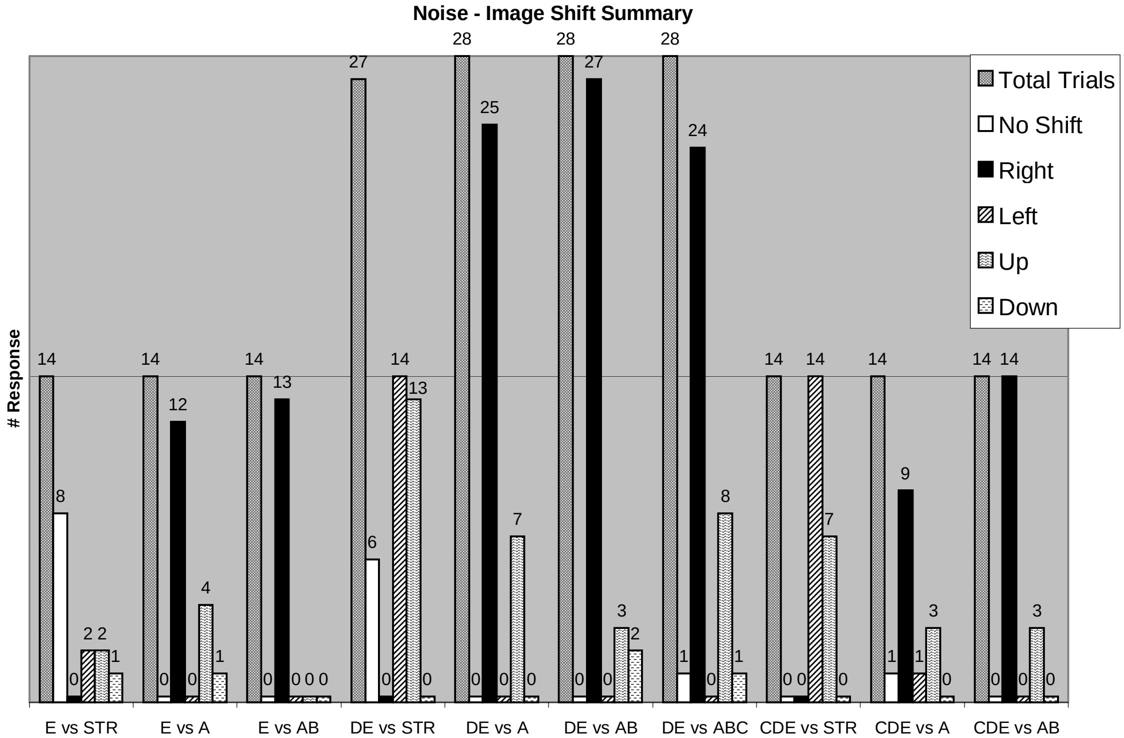


Figure 29: Noise Image Summary

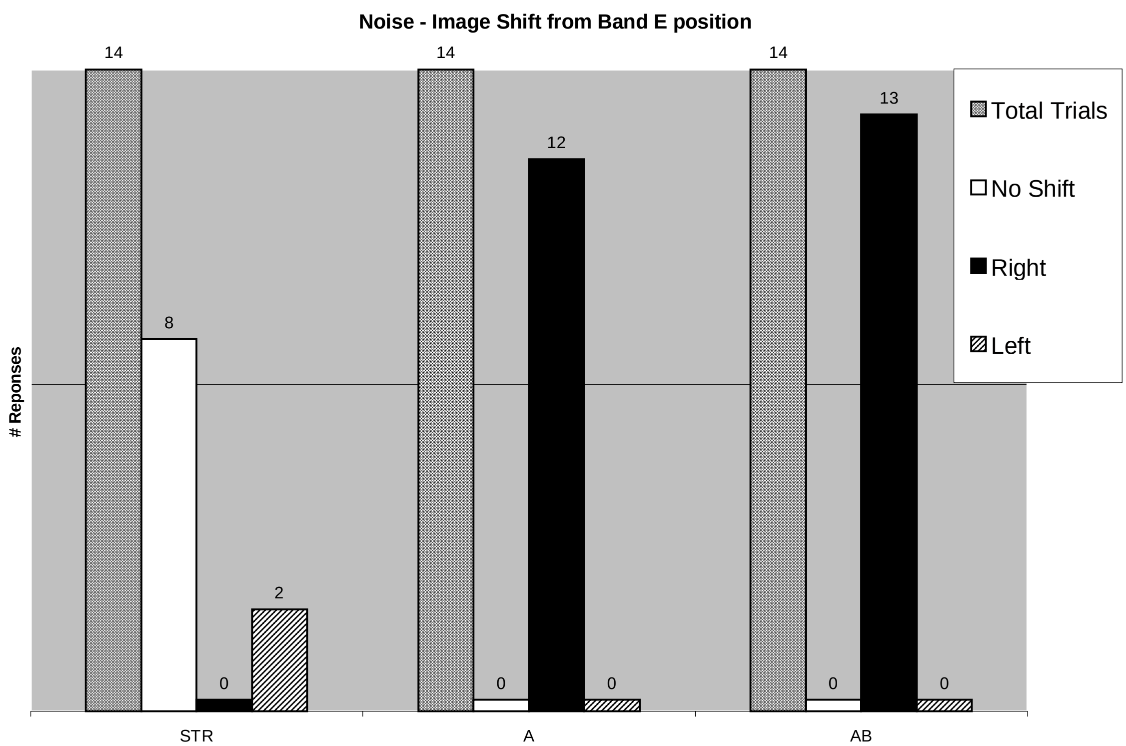


Figure 30: Band E vs. Noise Image

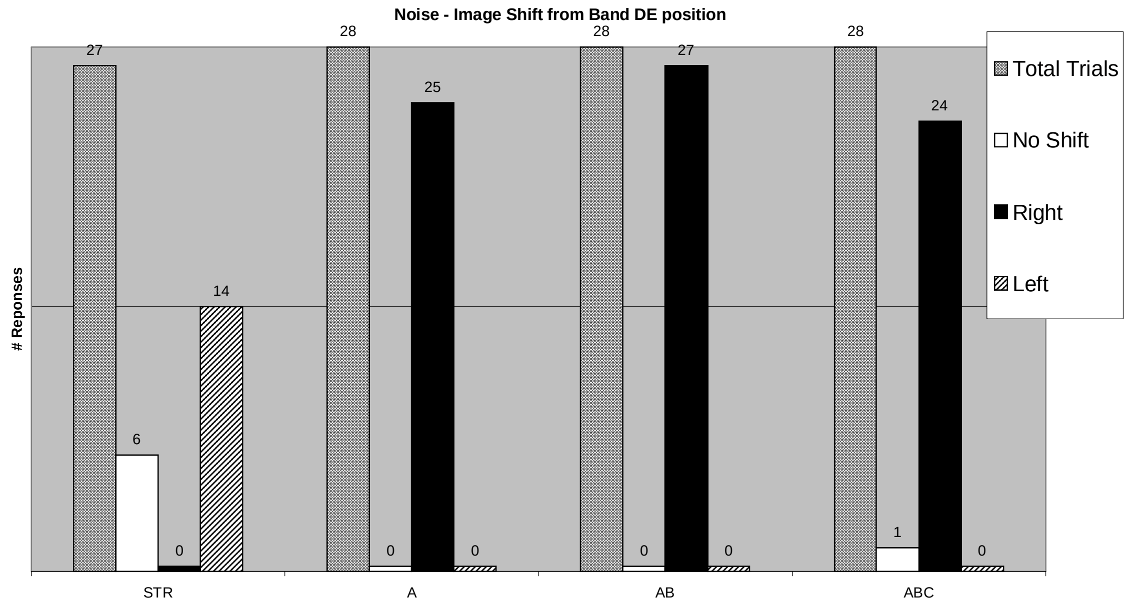


Figure 31: Band DE vs. Noise Image

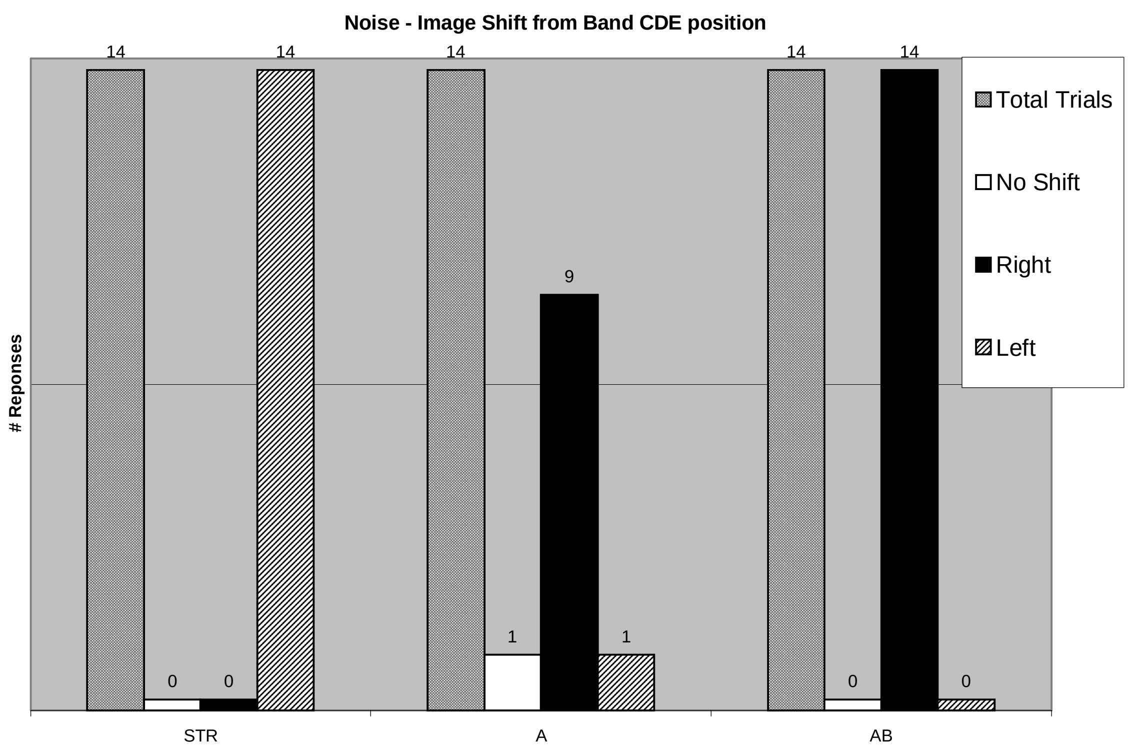


Figure 32: Band CDE vs. Noise Image

Analysis

The overall results of the listening tests suggest that noticeable shifts in the stereo image did occur when portions of the spectrum are relocated to the SR speaker. Also, the relative direction and amount of the image's movement seems to depend on which frequency band was moved to the SR channel.

It is first necessary to view the collection of results and try to eliminate correct answers due to chance. This is accomplished by calculating significance levels. The following analytical discussion is based upon an interpretation of the test subject's free-form verbal responses. Specifically, their answers were regrouped into three horizontal plane categories (no shift, right, or left). For example, a response of "up and to the left" was interpreted as "left," where a response of "closer" was interpreted as "no shift," because the subject did not mention any observed left/right movements.

Once the data was interpreted, the statistical analysis of the results was performed, as guided by Burstein (1990). Essentially, for any listening test of limited sample size, one must discuss the results in terms of probability, and not in terms of certainty. In fact when providing test results, it is informative to include the "criterion of significance" (α') along with the calculated significance level. For this thesis, two values for the criterion of significance were used; an α' of .05 was considered the predominant indicator of positive statistical significance, while an α' of .1 was considered to be lesser, yet still suggestive.

However, note that in this testing, an image that was shifting just right of the minimum audible angle, might produce an even number of responses for "no shift" and "right." In this case, both may fall short of the criterion of significance. Yet, despite two unconvincing level of criterion (for right and center) it is apparent from the results that the image did not shift left. Any cases like this will be explicitly pointed out during this analysis.

The significance level, *P(r)*, was determined by assuming the collected data follows a binomial distribution of random guessing. It can be directly calculated using:

$$P(r) = \sum_{r}^{N} \left(\frac{N!}{r! \cdot (N-r)!} \cdot x^{r} \cdot (1-x)^{(N-r)} \right)$$

where

N = number of trials
r = number of successes
x = probability of success for one trial based on guessing

For these results "x" was assumed to be one -third (1/3), allowing equal probability for each of the three categorized responses (no shift, right, left).

The music track will be analyzed first in its entirety, followed by the white noise analysis. The results for the music passage can be seen in Table 2, where the two criterion of significance (.05 and .1) are highlighted.

Music Significance Level	No Shift	Right	Left
E vs STR	**0.004**	0.997	0.895
E vs A	0.997	**0.000**	1.000
E vs AB	0.997	**0.000**	1.000
DE vs STR	**0.001**	1.000	0.712
DE vs A	1.000	**0.000**	1.000
DE vs AB	0.993	**0.000**	1.000
DE vs ABC	0.866	**0.000**	1.000
CDE vs STR	0.997	1.000	**0.000**
CDE vs A	0.739	0.058	0.997
CDE vs AB	0.895	**0.004**	0.997

Green < .05

Yellow <.1, >.05

Table 2: Significance levels for Music test results

Considering the results, it was initially surprising to see that no shift was observed for either band E vs. STR (stereo) or DE vs. STR (α'=.004 and .001). Notice however, that STR was perceived to the left of band CDE (α'=.000), suggesting that band C must contribute significant energy. Most of the other highly confident shift results involved the lower frequency bands of A, AB, and ABC. However, one borderline case should also be noted, where band CDE vs. A appears to be somewhere between "no shift" and "right." It could thus be considered that A is just right of band CDE. A visual summary of these results can be seen in Figure 33.

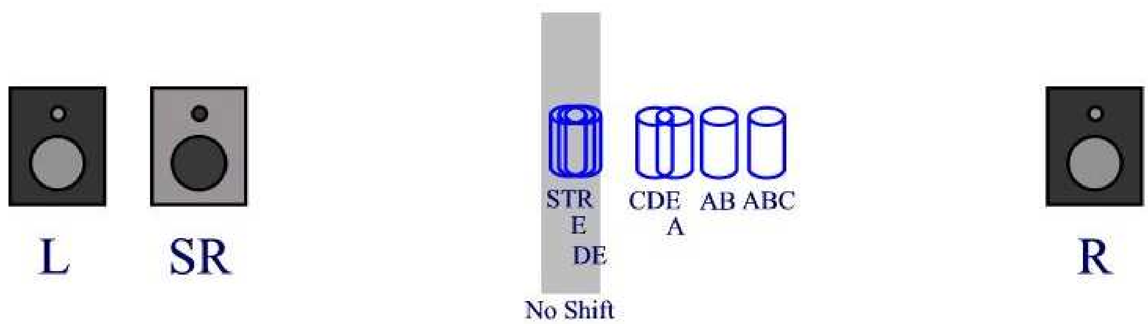


Figure 33: Shift hierarchy of SR bands for Music track

As was mentioned, a low significance level suggests that the change in localization cues was definite and noticeable. For the music test track, shifts were strong for the lower frequency bands. However, from previous analysis of the music track (see Figure 15-Figure 18), it is apparent that the signal's energy in the high frequency bands is fairly insignificant. This most likely explains why the listener did not perceive shifts for relocation of the higher frequency bands. While this is somewhat disappointing from the perspective of producing results to support this thesis, the results do support the validity of the listening tests. In other words, listeners seem to have reported only shifts that could have reasonably occurred.

On the other hand, the white noise test track will not have this lack-of-high-frequency problem because it has an even distribution of energy over the audible frequency range. The white noise track's test results are presented in Table 3. Similar to before, band E shows no shift compared to STR (α'=.001). Thus, even for the spectrally balanced white noise, moving band E creates unnoticeable shifts! However, STR is now noticeably to the left of bands DE and CDE (α' of .033 and .000), suggesting the high frequency bands do have some effect on the overall perceived location.

More importantly, moving band E versus the lower frequency bands of A and AB still produces noticeable shifts towards the right (α' of .000 and .000). Band DE, and CDE versus the low frequencies produces similar results, shifting noticeably further to the right with the low frequencies. These results alone provide strong support for this thesis, that high frequency cues are not as important as low frequency cues for the perception of the horizontal position of a stereo image. The resulting relative position of the images can also be seen in Figure 34.

Noise Signal's Significance Level	No Shift	Right	Left
E vs STR	**0.001**	1.000	0.973
E vs A	0.973	**0.000**	1.000
E vs AB	1.000	**0.000**	1.000
DE vs STR	0.074	1.000	**0.033**
DE vs A	1.000	**0.000**	1.000
DE vs AB	1.000	**0.000**	1.000
DE vs ABC	0.999	**0.000**	1.000
CDE vs STR	1.000	1.000	**0.000**
CDE vs A	0.973	**0.017**	0.997
CDE vs AB	1.000	**0.000**	1.000

Green < .05

Yellow <.1, >.05

Table 3: Significance levels for Noise test results

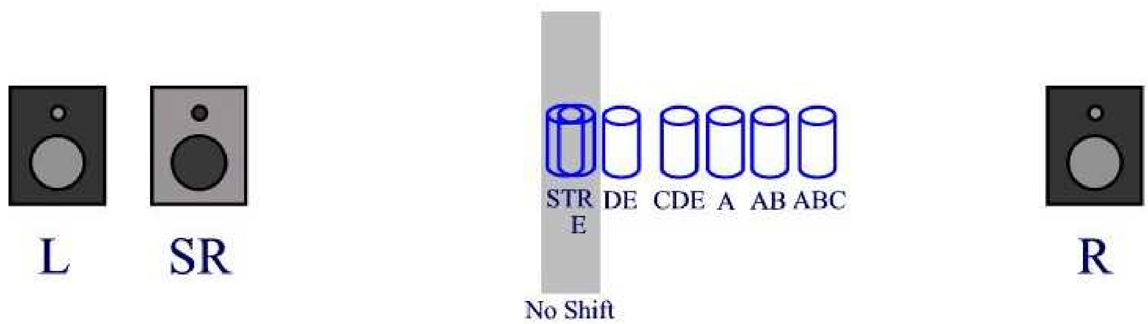


Figure 34: Shift hierarchy of SR bands for Noise track

One further point of discussion should surround the differences between the results from the music track and the white noise track. In fact, this was expected. The white noise test signal represents the ideal test condition, having a flat energy distribution across the audible spectrum. However, the music passage has a spectrogram that has a gross energy distribution amongst the frequency bands, left and right signals, and it also varies with time (see Figure 15). As was mentioned, this makes it difficult to draw conclusions from the results of the music tracks' trials.

Analysis - Loudness Calculations

With the results disclosed, one might argue that a "low frequency" band from 20 - 800 Hz (band A) is not a fair comparison versus a "high frequency" band from 12 -20 kHz (band E). The more abstract question might be, what is a fair comparison - considering that dominance is relative to the defined frequency range of each. What characteristics might have favored the low frequency dominance shown in this testing? One factor to consider is the loudness of the spatially relocated frequency bands.

Before calculating loudness, the basic concepts and terminology should be presented. From the fundamentals of acoustics, a vibrating body in space creates pressure variations in the medium, which radiates out from the source and are ultimately transferred to the listener's eardrums, thus creating a sensation of loudness. The magnitude of the pressure variations defines a physically measurable quantity called intensity. Loudness, a perceptual quality, is related to intensity but in a non-linear and frequency dependent fashion. Other important terms to consider include "loudness level" (measured in "Phon") and "loudness" (measured in "Sone"). The relationship of these terms are given below (also see ISO 131-1979 (E)):

$$\textit{Sound Pressure Level [dB SPL]} \quad = 20 \cdot \log(\frac{p}{p_0}) = 10 \cdot \log(\frac{I}{I_0}) = 10 \cdot \log(\frac{P}{P_0})$$

where $p_0 = 20\ \mu\text{Pa}$ and $p \equiv$ RMS pressure at a particular point

or $I_0 = 10^{-12}\ \text{w/m}^2$ and $I \equiv$ Average Intensity

or $P_0 = 10^{-12}$ w and $P \equiv$ Average Acoustical Power

$$\textit{Loudness Level}\ [\textit{Phon}] \equiv \textit{Perceptual equivalent loudness to a } 1000\ \textit{Hz tone, in dB SPL}$$

$$\textit{Loudness}\ [\textit{Sone}] = 2^{.1\cdot(\textit{PHON}-40)}$$

The equations above enable the db SPL of a sound to be determined from a measurable quantity. However, to determine the loudness level, a comparison must be made between the sound in question and the loudness of a 1000 Hz tone. For instance, if the test sound is determined to be equally loud as a 1000 Hz tone at 70 db SPL, then the sound is considered to have a loudness level of 70 Phon.

Unfortunately, loudness levels have no relative perceptual meaning, such that a sound at 60 Phon is not twice as loud as a sound at 30 Phon. In order to compare loudness levels, one must use the Sone scale; which has been determined through experimental evidence. This unit allows relative loudness comparisons to be made, such that 10 Sone is twice as loud as 5 Sone; and 4 Sone is half as loud as 8 Sone and etc. Also notice that 1 Sone is equivalent to a 1000 Hz tone at 40 dB SPL (i.e. 40 Phon).

Determining the perceived loudness of complex sounds is a well-published topic. The historical development of loudness models can be found in most audio engineering or psychoacoustics textbooks, such as Hartmann (1997) or Zwicker and Fastl (1990), which were used for this research.

From these texts, it is understood that Fletcher and Munson's (1933) work on loudness was the most popular origin for much of the later research. Continuing the tradition was two influential scientists including Stevens (1955, 1961) and Zwicker (1961) (also see Zwicker, Flottorp, & Stevens, 1957). Both published similar, but slightly different methods of calculating loudness, and both were ultimately included in the International Standards Organization (ISO) standard method for calculating loudness levels (see ISO 532-1975 (E)).

From the ISO standard, Stevens' Mark VI method (method A) is considered simpler and more effective for measurements taken in octave band increments, while Zwicker's (method B) is more accurate for one-third octave band measurements. Even more recent than the ISO standard, the two methods have been further improved. First Stevens (1971) himself improved his Mark VI method (to Mark VII), while Zwicker's method has suggested improvements from Moore and Glasberg (1996).

These loudness models are based on the fact that the ear has a non-linear frequency response, exhibits areas of masking (or inefficient excitation) called critical bands, and seems to follow the power law (Weber's or Steven's law) of perceptually measurable qualities (Hartmann, 1997). The power law suggests that there is a linear relationship between perceived loudness and the difference limen of intensity. Thus, increasing intensity by 10%, at any loudness level, will create a similar increase in perceived loudness.

For simplicity, only Steven's (1972) most recent Mark VII method will be used to calculate the loudness of the SR bands. It begins by dividing the sound's spectrum into octave bands. ISO recommends that the octaves are represented by their geometric mean, shown in Table 4 (also see ISO 266-1975 (E)).

Octave Band	1	2	3	4	5	6	7	8	9	10	11
Fc (geo. mean) [Hz]	16	32	63	125	250	500	1000	2000	4000	8000	16000

Table 4: ISO 266 Octave Band Frequency Centers

The power in each octave band is then used to calculate the sound's absolute level (in dB SPL). These level values are then correlated to a "loudness index" using lookup tables provided by Steven's (1972). Finally, the total loudness (in Sone) is calculated using the following formula:

$$S_{TOT} = S_{MAX} + F \cdot (\sum S - S_{MAX})$$

where

$$\begin{aligned} &S_{MAX} \equiv greatest\ loudness\ index, \\ &F \equiv \text{"}factor,\text{"} which\ depends\ on\ S_{MAX}\ and \\ &\quad band\ size\ used\ (octave,\ \tfrac{1}{3}\, octave,\ etc.);\ see\ tables \end{aligned}$$

Steven's method will be used to calculate the sound level for each of the six spatially relocated bands (A, AB, ABC, E, DE, CDE). As mentioned, the level of a sound can be calculated directly from the acoustical power (P) of the signal. Because the test signal was white noise, the total power of the signal is uniformly distributed over the entire spectrum and could thus be represented as:

$$dB\ SPL = 10 \cdot \log\left(\frac{P_{TOT} \cdot \frac{BW_{SRBAND}}{BW_{TOT}}}{P_0}\right)$$

where P_{TOT} = *Total Acoustical Power [Watts]*,

BW_{SRBAND} = *Bandwidth of Spatially Relocated band [Hz]*,

and BW_{TOT} = *Total Bandwidth [Hz]*

For the loudness experiments, the volume of the amplifier was adjusted so that the full bandwidth white noise playing from one speaker gave a level of 70 dB SPL on an A-weighted sound meter. The equation above yields a total acoustical power of .01 mW.

Then, assuming the reproduction equipment is ideal from 80-20,000 Hz (BW_{TOT} = 19920 Hz), the resulting sound levels were calculated and can be seen in Table 5.

	Freq. Range (Hz)	Band-width (Hz)	**Level (dB SPL)**
Band A	80 - 800	720	**55.6**
Band AB	80 - 1600	1520	**58.8**
Band ABC	80 - 5000	4920	**63.9**
Band E	12000 - 20000	8000	**66.0**
Band DE	5000 - 20000	15000	**68.8**
Band CDE	1600 - 20000	18400	**69.6**

Table 5: Level of Spatially Relocated Bands

However, in order to calculate the loudness of the SR frequency bands with Steven's method, it is necessary to further break down the loudness levels into octave bands. Thus the same approach using power (P) and the equation above was used to calculate the sound levels, in octaves, shown in Table 6.

Octave Band	Fc (geo. mean)	F-lower	F-upper	Band-width	Band A Band-width	**Band A level**	Band AB Band-width	**Band AB level**	Band ABC Band-width	**Band ABC level**	Band E Band-width	**Band E level**	Band DE Band-width	**Band DE level**	Band CDE Band-width	**Band CDE level**
1	16	11	22	11	0.0	**0.0**	0.0	**0.0**	0.0	**0.0**	0.0	0.0	0.0	0.0	0.0	0.0
2	32	22	43	22	0.0	**0.0**	0.0	**0.0**	0.0	**0.0**	0.0	0.0	0.0	0.0	0.0	0.0
3	63	43	86	43	6.0	**34.8**	6.0	**34.8**	6.0	**34.8**	0.0	0.0	0.0	0.0	0.0	0.0
4	125	85	170	85	85.0	**46.3**	85.0	**46.3**	85.0	**46.3**	0.0	0.0	0.0	0.0	0.0	0.0
5	250	170	340	170	170.0	**49.3**	170.0	**49.3**	170.0	**49.3**	0.0	0.0	0.0	0.0	0.0	0.0
6	500	340	680	340	340.0	**52.3**	340.0	**52.3**	340.0	**52.3**	0.0	0.0	0.0	0.0	0.0	0.0
7	1000	680	1360	680	120.0	**47.8**	680.0	**55.3**	680.0	**55.3**	0.0	0.0	0.0	0.0	0.0	0.0
8	2000	1359	2718	1359	0.0	0.0	241.0	**50.8**	1359.0	**58.3**	0.0	0.0	0.0	0.0	1118.0	**57.5**
9	4000	2718	5436	2718	0.0	0.0	0.0	0.0	2282.0	**60.6**	0.0	0.0	436.0	**53.4**	2718.0	**61.3**
10	8000	5437	10874	5437	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	5437.0	**64.3**	5437.0	**64.3**
11	16000	10873	21746	10873	0.0	0.0	0.0	0.0	0.0	0.0	8000.0	**66.0**	9127.0	**66.6**	9127.0	**66.6**

Table 6: Calculated Octave Band Levels

These level values were then matched with Stevens' (1972) tabular loudness indexes, which essentially compensate for the non-linear nature of human hearing and other experimental findings. The resulting loudness index values can be seen in Table 7.

Octave Band	Fc (geo. mean)	Band A level	Loud-ness Index Value	Band AB level	Loud-ness Index Value	Band ABC level	Loud-ness Index Value	Band E level	Loud-ness Index Value	Band DE level	Loud-ness Index Value	Band CDE level	Loud-ness Index Value
1	16	0.0	0.00	0.0	0.00	0.0	0.00	0.0	0.00	0.0	0.00	0.0	0.00
2	32	0.0	0.00	0.0	0.00	0.0	0.00	0.0	0.00	0.0	0.00	0.0	0.00
3	63	34.8	**0.06**	34.8	**0.06**	34.8	**0.06**	0.0	0.00	0.0	0.00	0.0	0.00
4	125	46.3	**0.46**	46.3	**0.46**	46.3	**0.46**	0.0	0.00	0.0	0.00	0.0	0.00
5	250	49.3	**1.34**	49.3	**1.34**	49.3	**1.34**	0.0	0.00	0.0	0.00	0.0	0.00
6	500	52.3	**2.58**	52.3	**2.58**	52.3	**2.58**	0.0	0.00	0.0	0.00	0.0	0.00
7	1000	47.8	**1.80**	55.3	**1.80**	55.3	**1.80**	0.0	0.00	0.0	0.00	0.0	0.00
8	2000	0.0	0.00	50.8	**3.15**	58.3	**3.15**	0.0	0.00	0.0	0.00	57.5	**5.24**
9	4000	0.0	0.00	0.0	0.00	60.6	**9.00**	0.0	0.00	53.4	**5.22**	61.3	**5.22**
10	8000	0.0	0.00	0.0	0.00	0.0	0.00	0.0	0.00	64.3	**12.00**	64.3	**12.00**
11	16000 (*12500)	0.0	0.00	0.0	0.00	0.0	0.00	66.0	**7.41**	66.6	**7.41**	66.6	**7.41**

Table 7: Loudness Index values from table lookup

With the loudness indexes obtained, the next step is to calculate the factor (F). Each of the six spatially relocated bands will have their own F value. In fact, F is determined by subtracting 4.9 dB from the level of the loudest octave in the spatially relocated band and then looking up a new index value in the tables. This new index is then correlated to an F value in a different factor tables in Stevens (1972). Also, because octave bands were used in the analysis, the final F value is double that listed in the table. All of this is per Steven's Mark VII instructions, the results of which are the "F factor" values shown in Table 8.

Finally, the loudness levels (in Sone) can be calculated using the previously mentioned equation. These results are also shown in Table 8, which provides some insight to the relative loudness of the SR bands. Note that the low frequency bands of A and AB are both considered to be softer than the high frequency energy of band E. In fact, band A is only two-thirds the loudness of band E. This fact in itself should dismiss loudness as the cause of the low frequency SR bands localization dominance.

	Freq. Range (Hz)	Band-width (Hz)	F (factor)	**Loud-ness (Sone)**
Band A	80 - 800	720	0.632	**4.9**
Band AB	80 - 1600	1520	0.618	**7.0**
Band ABC	80 - 5000	4920	0.500	**13.7**
Band E	12000 - 20000	8000	0.524	**7.4**
Band DE	5000 - 20000	15000	0.466	**17.9**
Band CDE	1600 - 20000	18400	0.466	**20.3**

Table 8: Final calculated loudness levels

Analysis - Loudness Experiments

It would have been interesting to compare each SR band to a reference 1000 Hz tone. This method provides a way to directly measure the relative loudness of each band, and would have also produced a Phon level for each band. However, it would not have helped determine what SR bandwidths would have made them equally loud. In other words, knowing that band A was 1.2 times softer (or louder) than band E would not have helped determine what band E's lower cutoff frequency should have been in order to equal the loudness of the two.

This question of relative bandwidth is also a relevant and interesting exercise, and prompted a second round of listening tests (see Appendix B for complete results). A different group of sixteen test subjects participated, all Music Engineering or Audio Engineering students at the University of Miami. The test was physically configured to be the same as the first tests, repeated in Figure 35.

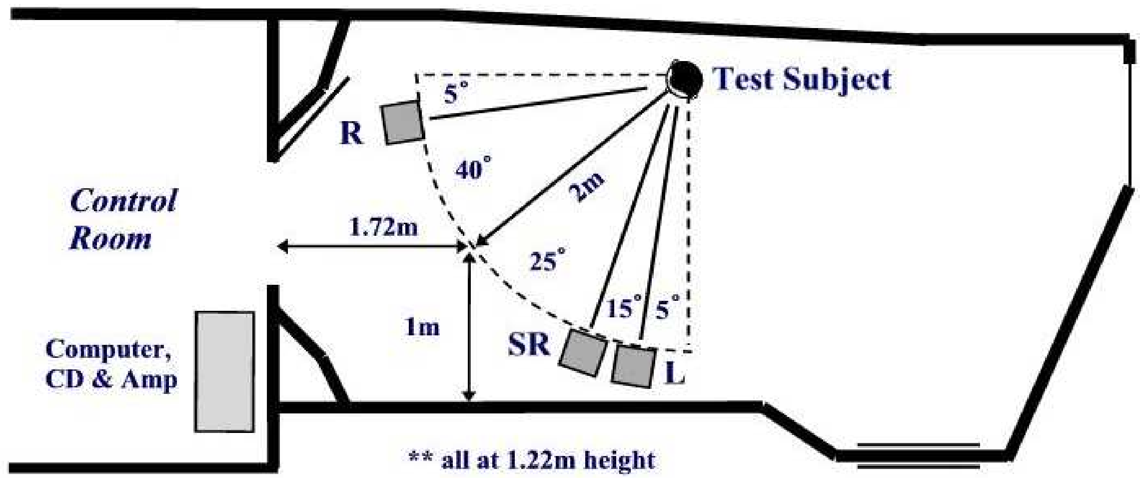


Figure 35: Listening Test II Setup

For these experiments, the test subjects were asked to compare the loudness of the same white noise bursts used in the first listening test. However this time, the bursts were only sent to the SR channel, since it was the main variable of the first round of experiments and directly caused shifting of the stereo image. The L and R channel reproduced no sound. The goal of the experiment was to compare the loudness of the SR bands with an adjustable low/high pass version of the filter. The cutoff frequency (signal bandwidth) would thus be adjusted to match the perceived loudness of the SR band (see Figure 36).

For example, the noise bursts representing band A (80-800 Hz) were first played for the test subjects. This was followed by a high pass filtered version of the noise bursts with an arbitrary lower frequency cutoff point. The subject was then asked if the second sound was "louder, softer, or about the same" as the first. Subsequent trials would ask the same question, while using a different lower cutoff frequency for the high pass noise. Obviously, the only variable for these tests is the frequency cutoff of the second filtered noise.

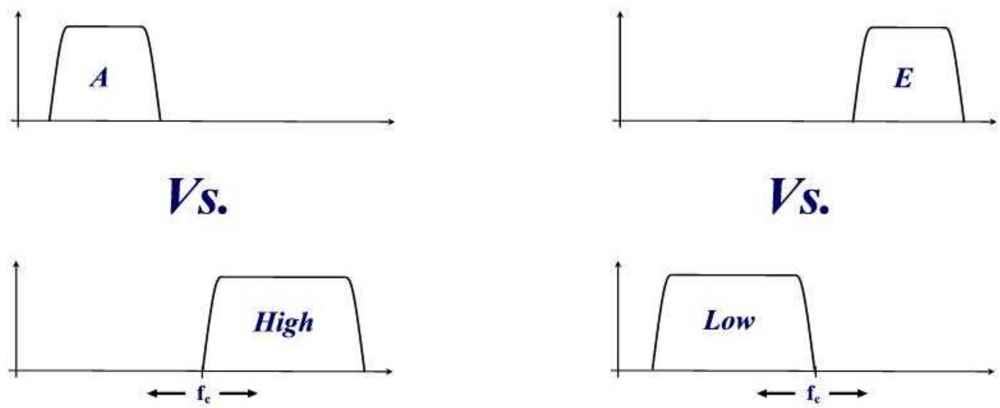


Figure 36: Loudness Comparisons of Low Frequency Bands with High pass (left) and High Frequency Bands with Low pass (right)

In other words, the low pass filtered spatially relocated bands (A, AB, and ABC) were compared to high pass noise of a varied lower frequency cutoff point. Similarly, the high pass spatially relocated bands (E, DE, and CDE) were compared to low pass noise with a varied higher frequency cutoff point (see Figure 36 for further clarification). The results of this experiment can be seen for each SR band in Table 9-Table 13.

A close analysis of the results suggests that the higher frequency bands are louder than the lower bands. Table 9 shows the results of the band A (80-800 Hz) versus high pass filtered noise comparison. Of course, as the lower frequency cutoff of the high pass noise is raised from 6 kHz to 15 kHz, it gets softer. Thus, the results suggest that the high pass noise is significantly louder than band A until around 14 kHz. Above this point, the high pass noise seems to approach the same loudness as A. However, there is no predominant indication that it is ever considered softer than band A. Particular notice should be given to the 12 kHz point (representing band E), which is considered by most to be louder than band A. Being louder than band E, this implies that bands DE and CDE should also be considered louder than band A.

Noise vs Bandwidth	Fc	SFTR		SAME		LDR		
A vs High	6k					2		
	8k					2		
	9k					1		
	10k	1				4		
	11k	1		1		5		
	12k	2		1		5		
	13k					3		
	14k	2		3		2		
	15k	1		3		2	Total	41

Table 9: Band A vs. Bandwidth Loudness

Similarly comparing band AB (80-1600 Hz) to the high pass noise in Table 10 shows that the subjects considered the high pass noise to be significantly louder until about 10 kHz. Around this point, the noise seems closer in loudness to band AB. Yet, above this point, most seemed to consider the high pass noise louder again. In particular, at 12kHz (band E), four listeners considered it to be louder than band AB. The results for 10kHz could thus simply be an anomaly. Regardless, these results also suggest that bands E, DE and CDE were louder than bands AB and A.

Noise vs Bandwidth	Fc	SFTR		SAME		LDR		
AB vs High	2k	1				5		
	6k			1		2		
	8k			2		6		
	9k	1				2		
	10k	2		3		3		
	11k			1				
	12k			1		4		
	13k			1		1		
	14k			4			Total	40

Table 10: Band AB vs. Bandwidth Loudness

Moving on to the high frequency versus low pass noise comparisons shows similar indications (Table 11). When comparing SR band E, most subjects reported the low pass noise was softer even up to 2,000 Hz (which contains both bands A and AB). Only at around 3 kHz did some listeners suggest the low pass noise became louder than band E.

Noise vs Bandwidth	Fc	SFTR		SAME		LDR		
E vs Low	300	2						
	500	4		1				
	600	1						
	700	2		3		1		
	800	3				1		
	900			1		1		
	1k	6		3		2		
	2k	2		1		1		
	3k			1		3	Total	39

Table 11: Band E vs. Bandwidth Loudness

For band DE (5,000-20,000), listeners did not describe the low pass noise as being significantly louder - even up to a cutoff of 7 kHz (see Table 12). Instead, they considered the low pass noise softer, or sometimes equally loud. One discrepancy seems to be their sense of equal loudness in the 2000-4000 Hz range, yet the low pas noise at 7 kHz was considered softer. Regardless, these results suggest that band DE is louder than bands A or AB, but about as loud as band ABC (80-5000).

Noise vs Bandwidth	Fc	SFTR		SAME		LDR		
DE vs Low	900	1						
	1k	3						
	2k	3		4				
	3k	3		1		2		
	4k			5				
	5k	1		2		1		
	6k			1		1		
	7k	3		1			Total	32

Table 12: Band DE vs. Bandwidth Loudness

Finally, band CDE can also be reviewed despite having limited data points. The results seem to again suggest that listeners did not consider the low pass noise to be louder than CDE, even at 7 kHz. Instead, they seemed to consider the low pass noise softer up to around 4 kHz, above which it appeared equally loud. This, as well as the results from the other tests, suggests that band CDE is most likely the loudest of all the spatially relocated bands.

Noise vs Bandwidth	Fc	SFTR		SAME		LDR		
CDE vs Low	2k	2						
	3k					1		
	4k	2						
	5k			2				
	6k			3				
	7k			1				
	9k					1	Total	12

Table 13: Band CDE vs. Bandwidth Loudness

Analysis - ABX Testing

It has been established that the lower frequency SR bands did dominate the localization of a stereo image and not because they are louder. However, the results are primarily based on white noise bursts instead of a more realistic music track. Therefore, during the second round of listening tests, ABX tests were additionally performed using two different music tracks. The same 16 test subjects of the second round of tests also performed a set of ABX tests for one of the two music tracks (alternatively selected).

ABX testing is often used to study the perceivable differences between two items. The test subject is played three versions of a test track: A, B, and X. Sounds A and B are different, but will stay the same in each trial. However, for each trial, track X is a random selection of either A or B. The subjects are repeatedly asked to identify whether X is A or B. The results produce a level of statistical confidence, which suggests whether the listener can reliably tell the difference between A and B. Obviously more correct answers of X produces a higher confidence. However, low confidence does not necessarily prove the sounds are the same, but usually implies the sounds are more similar than dissimilar.

In these experiments, the listeners were asked to discern between regular stereo music (condition A) and the SR frequency setup shown previously in Figure 14 (condition B). During condition B, the SR channel would play the left stereo signal's frequencies above 10 kHz. Thus, the R channel played the same right stereo signal for both trials A and B. The L channel played the full bandwidth left signal on trial A, and a low pass filtered (at 10 kHz) version for trial B. The SR channel played nothing for trial A and played the high pass filtered (at 10 kHz) version of the left stereo signal for trial B.

One of the chosen music tracks was a 10 second clip (0:10-0:20) of the introduction to Madonna's "Candy Perfume Girl." It was used because it contains a very unique wide bandwidth, noise-like sound that moves from center to right sound stage. Note that for the listening tests, the left and right stereo signals were flipped from the original recording in order to maximize the influence of the SR channel. Thus, listeners should have experienced the noise moving from center to left stage during trial A (stereo) of these experiments. The spectrogram of the test signal can be seen in Figure 37. Recall that the amplitude is represented by the intensity, with time as the x-axis and frequency as the y-axis.

The other test track was an eleven second clip of the Tower of Power's "What is Hip?" The section was chosen because it contained a high-note (squealing) riff from a right-panned trumpet section, while the other band members (sax, brass, percussions, singer, etc.) played simultaneously. It had a greater amount of high frequency energy than most of the other test tracks available. This will be discussed later in more detail. Again, for the ABX test, the left and right stereo signals were flipped from the original recording. The spectrogram of this test track can be seen in Figure 38.

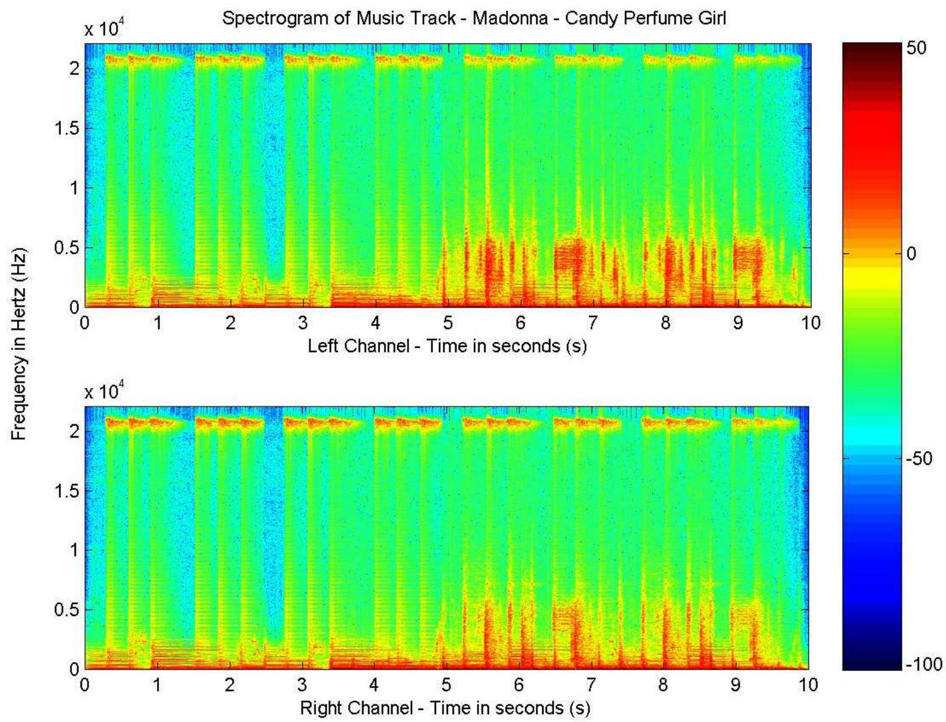


Figure 37: Spectrogram of Track One - Madonna

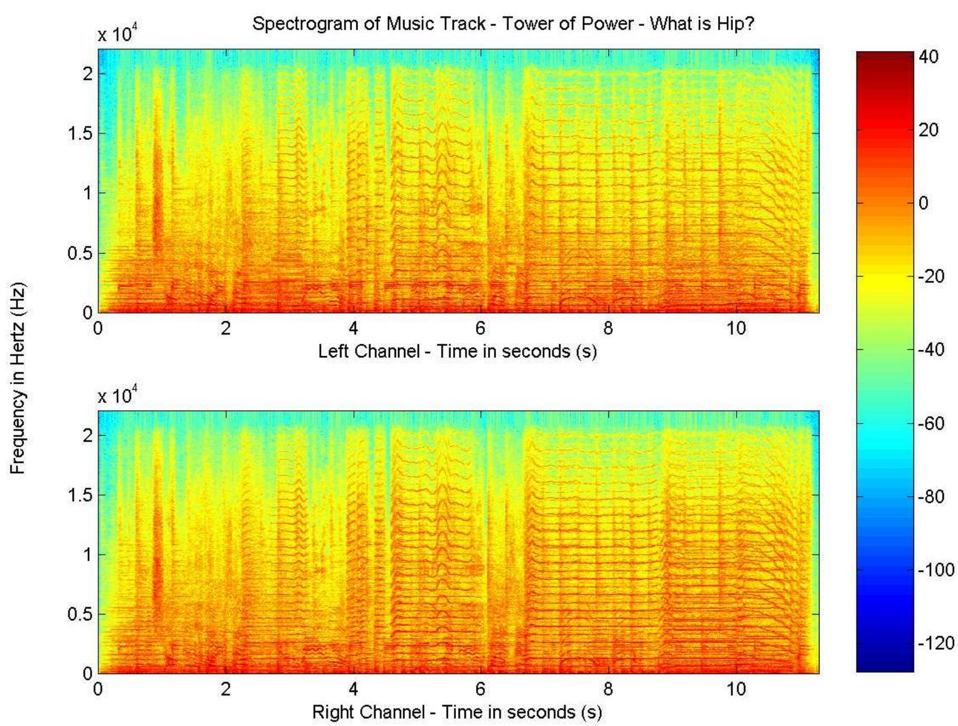


Figure 38: Spectrogram of Track Two - What is Hip?

Each subject performed 5-10 trials of the ABX tests for one of the two test tracks. If all subjects and trials are assumed to be equivalent, the tests produced 62 trials for track one and 83 trials for track two. The results can be seen in Table 14.

ABX Test, Track 1 - Madonna																	Grand Total
Sub	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	
Total	7	6	8	5	7	0	8	0	0	3	9	0	0	5	0	4	62
# Correct	3	4	2	1	1	0	4	0	0	1	5	0	0	1	0	2	24

ABX Test, Track 2 - What is Hip?																	Grand Total
Sub	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	
Total	6	6	0	0	8	8	0	9	9	10	0	10	9	0	8	0	83
# Correct	5	3	0	0	3	6	0	6	4	4	0	2	5	0	3	0	41

Table 14: ABX Test Results

Considering the high frequency content of the tracks, one might predict that trial B should have resulted in a noticeable shift of the high frequency stereo images towards the right. Also some changes in the spectral balance of the signal should have been noticeable, caused by the spatial relocation of the left signal's high frequencies.

However, the results suggest that the test subjects could not reliably tell the difference between the stereo and SR versions. In fact, calculating the results (as previously discussed, using x = 50%) gives a fairly high significance level for both trials (α' =.972 for trial 1, α' =.587 for trial 2). This is greatly above the typically acceptable value of α =.05. A high significance level yields low confidence levels, which ultimately suggests that the tracks were hard to differentiate.

However, note that some subjects could tell the difference between the two trials (see Table 14). Subjects 2 and 11 on track one, and 1, 6, and 8 on track two performed above average. Also notice that the listeners seemed to perform better on track two than they did on track one. Of course, varied results should be expected. After all, the perceptual significance of the SR channel depends heavily on the amount of high frequency information, and its proportion to concurrent lower frequency information.

A closer look at the two test tracks might provide further insight as to why they were difficult to discern from stereo, and even why track two was easier than track one. When considering these tracks, it is important to focus on the spectral content above 10 kHz of the left stereo signal (right signal of the original recording). After all, this was the only difference between trial A (stereo) and B (the SR version).

The amount of relocated energy is most likely a major factor. More energy suggests a louder SR speaker, which is the most measurable perceptual impact. Loudness however, is not necessarily the most important factor as has been previously shown. Thus, the energy distribution above and below 10 kHz was calculated and plotted in the lower right plot of Figure 39 and Figure 40. Recall that the left and right channel were flipped during testing. Thus, notice the portion of energy that is represented by the "right" signal of the two tracks.

Realize that the second music track has a greater amount of high frequency energy than the first. However, these energy plots assume an ideal brick-wall filter at 10 kHz; as opposed to the actual testing condition, which used a second order high, pass filter. This means that a slightly greater amount of energy could be expected from the SR channel than what is shown in the figures.

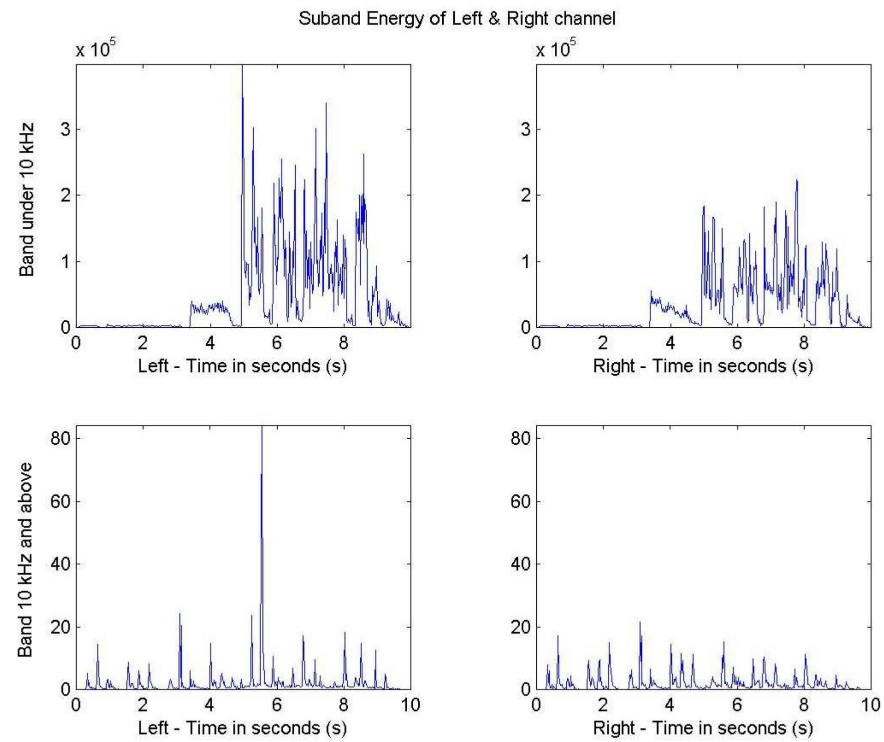


Figure 39: Spectral Energy for Track One - Madonna

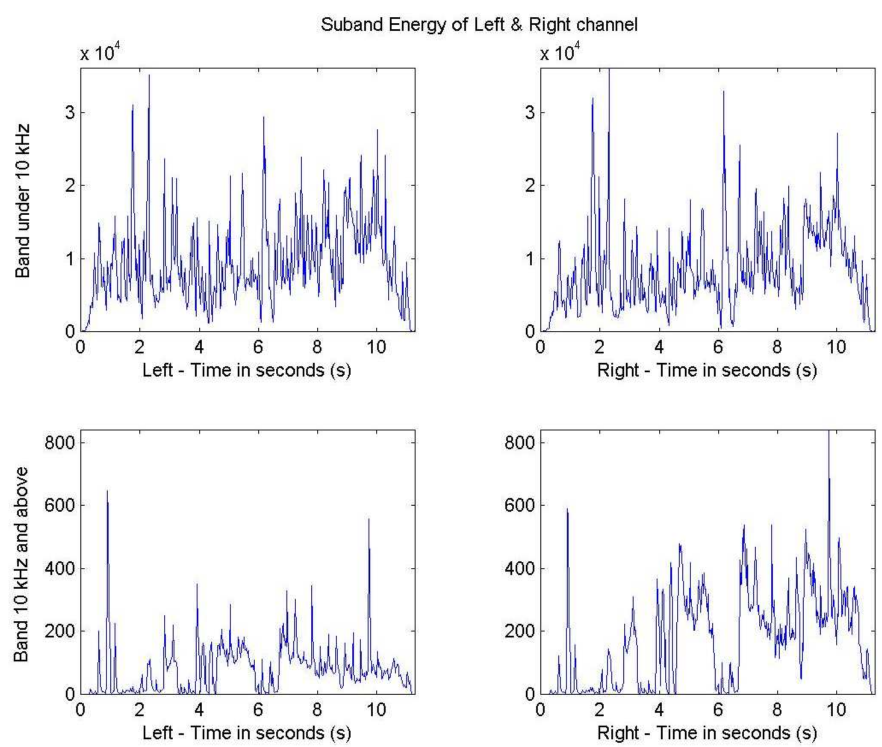


Figure 40: Spectral Energy for Track Two - What is Hip?

A logical follow-up question might be to ask what typical amount of energy is above 10 kHz in music. This is obviously difficult to say for certain, however evaluating a sample of varied music tracks should be insightful. Actually, this analysis was performed prior to the second round of listening tests in order to choose a music track that would hopefully create a significant image shift during testing.

Fifty-two tracks from a critical listening music CD were evaluated. The disc contained a wide selection of genres. For each of the tracks, a ten second segment containing the phrase with the most high frequency energy was chosen. This was determined by monitoring Sound Forge's spectrum analysis during playback. Then, the shorter sound clips were each analyzed using short time frequency analysis techniques in Matlab (see Appendix C for code). This produced a short-time calculation of total energy, energy below 10 kHz, and energy above 10 kHz. Ultimately these numbers were used to find an average value of energy above 10 kHz.

Surprisingly, only seven of the fifty-two clips had more than 3% average energy above 10 kHz (see Table 15). The final music tracks for ABX testing were chosen because they had a high frequency image that was stationary on the sound stage during playback and was non-impulsive. Track eight was also used in the testing , despite having only 1% average energy above 10 kHz.

Of course, it was assumed that "average energy" is representative metric for determining which track would exhibit shifts due to SR high frequencies. In actuality, the short-time ratio of energy above and below 10 kHz is probably more representative; and also more difficult to calculate.

Track #	Track Title	**Avg % Energy Above 10 kHz**	Highest Instantaneous Energy Above 10 kHz	Comments
1	Harry James - More Splutie Please	**4.00**	26	Brass/Big Band; good balance
2	**Tower of Power - What is Hip**	**3.11**	**33**	**Brass/Big Band; good balance**
3	Yello - Oh Yeah	**8.21**	58	Percussion/voice; very transient
4	Talking Heads - Flowers	**9.11**	62	percussion only; very transient
5	Don Dosey - Ascent	**5.77**	97	Percussion/Synth sounds; a lot of stage movement
6	Fuzzion	**4.09**	64	Drums; very transient
7	George Duke - Miss Wiggle	**4.26**	76	Synth/keyboard; a lot of movement on sound stage
8	**Madonna - Candy Perfurme Girl**	**1.00**	**2**	**Sent by K. Heber**

Table 15: Energy Analysis of misc. music tracks at 10 kHz

Another influential difference between the tracks could have been that track two had a stationary image, whereas track one was moving during playback. Thus, during ABX testing, the listener would have had to identify the panning movement and then notice that it became more stationary in the SR version. This is perhaps a lot to expect for an unfamiliar piece of music.

Chapter 7: Conclusions and Recommendations

Popular opinion on the topic of consumer audio systems is that low frequencies are "hard to localize," and thus can be reproduced by a mono sub woofer in some non-critical location. However, a paper discussing DTS' technique for encoding surround sound (Smyth, 1999) stated that "experimental evidence suggests that it is difficult to localize mid-to-high frequency signals above about 2.5 kHz" (p. 18). Smyth continues, commenting that when a listener is presented with concurrent low and high frequency information, that high frequencies are relatively unimportant for proper image localization. His statement seemed to oppose what is popularly said about localization, and prompted the preliminary research for this thesis.

What does the statement "hard to localize," encompass? There seems to be several ways to interpret this. One might say that this describes sounds that are least accurately localized (i.e. larger minimum audible angle). The easier a sound is to localize, the more accurate one should be able to locate it. However, others might argue this describes those sounds that are "confusing" to localize (especially front from back). For instance, sounds without monaural pinna cues (i.e. not containing the 5-12 kHz range) are often difficult to discern front from back. Even further, this might describe a listener's confidence in the location of the event. Most listeners are not confident in the location of narrow band continuous sounds, even though they may be able to determine the correct location. Realize that this claim is open to interpretation, and most likely is a collective representation of all of these items.

During this investigation, several interesting discoveries were made. First, it seems logical that subwoofers are also fairly difficult to localize. However, this is not necessarily because they reproduce low frequencies. In fact, from an absolute sense, the middle frequency range (1-3 kHz) is exhibits the largest minimum audible angle. This was determined specifically by Stevens and Newman (1936) and later by Mills (1958) (among others). They concluded that low frequency sounds exhibit definite interaural time differences, whereas high frequencies have strong interaural level differences. Both cues exist in the middle frequency range, yet neither seems to dominate - leading to larger MAAs for middle frequencies than either lower or higher frequencies.

Subwoofers are difficult to localize because they typically reproduce continuous, narrow band (20-100 Hz) sounds. Much of the localization research (see Blauert, 1999) has shown that sounds get easier to localize with increasing bandwidth. This is because the number and type of localization cues increase with bandwidth, thus providing more cues for the brain to compare. For instance, to localize a low frequency tone, a listener must rely only on interaural time differences; no level differences or pinna cues are present. On the other hand, white noise exhibits low frequency ITDs, pinnae cues and high frequency ILDs. Each of these cues will be in agreement to clearly indicate the location of the sound event. This results in a more confident sense of a source's location. Also, impulsive sounds produce transients in the localization cues, which allows the brain to better interpret the location.

Although having a smaller MAA than middle frequencies, low frequencies are probably the "most confusing" to localize. Frequencies un der about 5 kHz are devoid of monaural pinnae cues, which help avoid front/back confusion. In addition, Hartmann (1983) has suggested that reverberation due to room acoustics has the most impact on the localization of low frequency signals. Yet, listeners should be able to discern left-from-right, because of dominant low frequency ITD cues.

In an attempt to further investigate this topic of localization versus frequency, this thesis studied the effect that spatially relocating portions of the audible spectrum have on the localization of a stereo image. This "spatial relocation" commonly occurs in consumer electronics, where tweeters (high frequencies) are often physically separated from a woofer (low frequencies), sometimes by a significant distance (see Figure 1).

The experiments for this thesis compared the relative shift of a stereo image caused by horizontally relocating various low and high frequency bands. Specifically, in a stereo speaker setup (± 40°), frequency bands were relocated from the left speaker to an offset speaker 15° closer to the median plane (see Figure 14). The subjects were asked to comment on the relative shift of a centrally located image created by white noise bursts and music.

These back-to-back comparisons would include two of seven conditions including stereo (no relocation) or six different relocated frequency bands: A (80-800 Hz), AB (80-1,600 Hz), ABC (80-5,000 Hz), E (12,000-20,000 Hz), DE (5,000-20,000 Hz), or CDE (1,600-20,000 Hz). The frequency points which define these bands were chosen because they compare low and high frequencies and are known to contain localization cues of relative dominance. ITDs are known to dominate the 20-800 Hz range, while their effect diminishes from 800-1,600 Hz, having no effect above this range. Pinnae cues occur in the 5-12 kHz range (see Blauert, 1999).

Moving these frequency bands from the left channel to the SR channel (see Figure 20) would essentially alter the localization cues of the overall auditory event. The theory behind the experiments was that the most dominant localization cues would create the most noticeable shift towards the right. This is because ultimately, there are several localization cues that have a relative salience across the audible spectrum. The most important of these is the low frequency (< 800 Hz) interaural phase differences (IPD). This is followed by high frequency (2 -20 kHz) interaural level difference (ILD) and lastly, the monaural spectral cues of the pinnae (5-12 kHz).

From experimental listening tests performed for this thesis, it has been shown that relocating the lower frequency bands (A, AB, ABC) caused more noticeable horizontal shifts to the stereo image than those caused by relocating the high frequency bands (E, DE, CDE). While music was used for portions of the experiment, the most reliable test signal was a set of white noise bursts. Not only does white noise represent an even distribution of spectral energy, but is also known to be one of the easiest types of sounds to localize (Stevens & Newman, 1936). Music has a time-varying amount of spectral energy, which makes it more difficult for the listener to notice the spatial relocation.

Results of the noise track showed that relocating band E typically produced no noticeable shift as compared to stereo (see Figure 29 and Table 3). However, moving bands A, AB, or ABC produced significant shifts towards the right. Comparably, moving bands DE and CDE also shifted the image to the right of stereo, but not as far as those created by the lower frequency bands. Essentially, the results suggest that the stereo image created by relocating the low frequency bands was generally shifted further to the right than with relocating the high frequency bands.

The specific reason for this apparent low frequency dominance is difficult to determine. Loudness could be a possible factor, because left/right panning is typically associated with the balance of the stereo channels. For instance, if band A was much louder than band E, this could explain why it was more influential. Therefore, loudness was both calculated (see Table 8) and experimentally determined (Table 9 - Table 15) for the SR bands. The results found the high frequency bands to be louder. In fact, band E is almost one-third louder than band A. Therefore, loudness is probably not the cause of the low frequency localization dominance seen here.

It seems more likely that the 15º change in azimuth creates different localization cues for the SR band, and that changes in low frequency ITD cues produce a more noticeable image shift. Thus, moving band E mainly changes ILD cues, whereas moving band A causes changes in ITD. It is well established that ITDs tend to dominate overall perception, which this research supports.

Also, a large concentration of high frequency energy does not seem to commonly occur in music. An analysis of fifty-two mixed-genre music tracks produced only seven with more than 3% of their average energy above 10 kHz. Therefore, it seems reasonable that in most music tracks, high frequency energy is a fairly insignificant portion of the overall energy being reproduced.

An additional factor could be that listeners may not pay much attention to the stereo sound stage, especially for the upper audible spectrum. This was supported during ABX testing, where listeners were not able to differentiate between regular stereo and a setup that shifted the left speaker's high frequency (>10 kHz) signals by a 15º azimuth towards midline. These short music clips contained a slightly greater average (~3%) of high frequency energy than the "typical" music track; determined from the above - mentioned sampling of mixed-genre tracks.

Having shown that low frequency energy dominates the localization of a stereo image for this particular test setup and variables, there are several directions future research could take. The most obvious has practical applications, where one could develop the "mono-ized" tweeter system discussed in the introduction. However, simply moving the high frequency information to a central tweeter with no pre-processing will create a system with an odd sound stage. Recall that this setup creates localization shifts towards the tweeter for any instrument/image that has a large concentration of high frequency energy relative to the chosen crossover frequency (i.e. a cymbal).

The most noticeable differences for the mono tweeter system will be the change in high frequency image position and spectral balance. The new image position is formed because without a tweeter on both sides of the listener, the system no longer reproduces the intended interaural level differences. Instead, the level differences are dictated purely by the spatial position of the mono tweeter and the amount of high frequency energy it is reproducing. This is difficult, if not impossible, to compensate for.

The spectral balance will also be different, because only one tweeter (instead of two) is replicating high frequencies. This will potentially reduce the high frequency loudness. Also, the sound is no longer coming from an off-center azimuth (i.e. 30º), but instead from a location near center (i.e. 0º). Each spatial position has a different path to the ears, which exhibits a characteristic "filtering" effect described by the Head Related Transfer Function (HRTF). For instance, if the mono tweeter is located at 0º azimuth, and that position is known to attenuate 5 kHz signals compared to a typical position of 30º, this should be compensated for by boosting the 5 kHz range during preprocessing. HRTFs have been extensively researched and applied to audio systems offering "simulated" surround sound with two speakers or "3D" headphone listening systems (see Gardner, 1998).

This preprocessing relies on the ability to predict the intended spatial origin of the recorded high frequency information. Because high frequency images are localized using interaural level differences, the cross correlation of high frequency left/right amplitude information should indicate the intended spatial position of the recording. Equal energy suggests the image is towards the middle, while unbalanced energy suggests an image on one side. Knowing the intended spatial position, along with the actual location of the mono tweeter, one could process the signal using HRTFs to better camouflage the missing tweeter.

However, the image's location will still be incorrect due to the inability to control high frequency interaural level differences (having only one tweeter). One idea to explore would be to introduce a time-delayed version of the mono signal, while encoding the original and time delayed version with different temporal envelope modulation. At high frequencies, only envelop time differences (and level differences, which are fixed in this case) will have an effect on the resulting stereo image position.

As to more simple investigations, one might change certain parameters used in these experiments in an attempt to support or disprove these findings. Perhaps, comparing results between a horizontal and vertical SR channel would be interesting. It is expected that the vertical channel would be even more difficult to notice. This is because a change in vertical position only alters the monaural spectral cues, while level and time differences will stay the same. Strybel & Fujimoto (2000) showed the vertical minimum audible angle to be 4-5° larger than the horizontal MAA at 0° azimuth. Others might investigate the effect of relocating "equally loud" frequency bands, different loudspeakers, a reverberant vs. "typical" acoustic space, or varied speaker locations and configurations.

The incorporation of in-ear recordings could also lead to a more analytical approach to understanding the listening test results. These recordings are obtained by playing the test signals while monitoring the ear canal microphones of a dummy head or by placing probe microphones at the ear canals of actual listening subjects (see Blauert, 1999, p. 31). The recordings provide spectral and temporal representations of the ear canal signals, which could be used to support the listening subjects' responses.

In summary, this research suggests that the localization of a stereo image is most affected by the spatial origin of low-to-mid frequencies as opposed to higher frequencies. This is not due to any absolute localization abilities, considering that the middle frequency range (1-3 kHz) is known by scientists to have the largest minimum audible angle (i.e. the least accurately localized). Instead, it is probably due to the perceptual dominance of low frequency interaural time differences.

Regardless, a large amount of high frequency information is not typically present in music. This was additionally supported when a group of average listeners did not notice that the high frequency sound stage had been altered. Therefore, modifying high frequency localization cues seems to have a minimal impact on the perceived performance of the sound system. This technique could thus be applied to perceptual coders, as is implemented in DTS' technique (Smyth, 1999), or audio system designers via the mono tweeter system. The mono tweeter system will obviously reduce the cost of the system while having a minimal effect on fidelity, especially if some preprocessing is performed.

References

Angell, J.R., & Fite, W. (1901). The monaural localization of sound. *Psychol. Rev., 8,* 225-243.

Ballman, M. (1990). *Auditory localization research tool.* Unpublished Master's Thesis. University of Miami, Florida.

Batteau, D.W. (1967). The role of the pinna in human localization. *Proc. Roy. Soc. London Series B, 168 (1011),* 158-180.

Begault, D. (1994). *3-D sound for virtual reality and multimedia.* Boston: AP Professional.

Blauert, J. (1969). Sound localization in the median plane. *Acustica, 22,* 205-213.

Blauert, J. (1971). Localization and the law of the first wavefront in the median plane. *J. Acoust. Soc. Am., 50 (2),* 466-470.

Blauert, J. (1999). *Spatial hearing: the psychophysics of human sound localization.* Cambridge, Mass.: The MIT Press. (Original work published in 1974)

Bregman, A.S. (1999). *Auditory scene analysis: the perceptual organization of sound.* Cambridge, Mass: MIT Press.

Buell, T.N., & Trahiotis, C. (1994). Recent experiments concerning the relative potency and interaction of interaural cues. In R.H. Gilkey & T.R. Anderson (Eds.), *Binaural and spatial hearing in real and virtual environments.* (pp. 139-149). Mahwah, NJ: Lawrence Erlbaum Associates.

Burstein, H. (1990). Statistics in A/B testing: by the numbers. *Audio, 74,* 43-48.

Butler, R.A., & Belendiuk, K. (1977). Spectral cues utilized in the localization of sound in the median sagittal plane. *J. Acoust. Soc. Am., 61 (5),* 1264-1269.

Cherry, E.C. (1953). Some experiments on the recognition of speech, with one and two ears. *J. Acoust Soc. Am., 25 (5),* 975-979.

Durrant, J., & Lovrinic, J. (1984). *Bases of hearing science.* Baltimore: Williams & Wilkins.

Fisher, H., & Freedman, S.J. (1968). The role of the pinna in auditory localization. *J. Audi. Res., 8,* 15-26.

Fletcher, H., & Munson, W.A. (1933). Loudness, its definition, measurement and calculation. *J. Acoust. Soc. Am., 5,* 82-108.

Gardner, M.B. (1973). Some single- and multiple-source localization effects. *J. Audio Eng. Soc., 21 (6),* 430-437.

Gardner, W. (1998). *3-D audio using loudspeakers.* Boston: Kluwer Academic

Gelfand, S.A. (1998). *Hearing: an introduction to psychological and physiological acoustics.* New York: Marcel Dekker.

Gockel, H., Carlyon, R., & Micheyl, C. (1999). Context dependence of fundamental-frequency discrimination: Lateralized temporal fringes. *J. Acoust. Soc. Am., 106,* 3553-3563.

Grantham, D.W. (1984). Interaural intensity discrimination: insensitivity at 1000 Hz. *J. Acoust. Soc. Am., 75 (4),* 1191-1194.

Hartmann, W.M. (1983). Localization of sound in rooms. *J. Acoust. Soc. Am., 74 (5),* 1380-1392.

Hartmann, W.M. (1993). Auditory localization in rooms. In S. Bech (Ed.), *Proc. Audio Engr. Soc. 12th International Conference,* New York: AES.

Hartmann, W.M. (1997). *Signals, sound, and sensation.* Woodbury, New York: American Institute of Physics.

Hartmann, W.M. (1999). How we localize sound. *Physics Today, 11, 24-29.*

Hebrank, J., & Wright, D. (1974). Spectral cues used in the localization of sound sources on the median plane. *J. Acoust. Soc. Amer., 56,* 1829-1834.

International Organization for Standardization. (1980). *ISO Standards Handbook 4: Acoustics, Vibration and Shock.* Switzerland: International Organization for Standardization.

Kendall, G. (1995). A 3D sound primer: directional hearing and stereo reproduction. *Computer Music Journal, 19 (4),* 23-46.

Litovsky, R.Y., Colburn, H.S., Yost, W.A., & Guzman, S.J. (1999). The precedence effect. *J. Acoust. Soc. Am., 106 (4),* 1633-1654.

Middlebrooks, J.C. (1997). Spectral shape cues for sound localization. In R.H. Gilkey & T.R. Anderson (Eds.), *Binaural and spatial hearing in real and virtual environments.* (pp. 77-97). Mahwah, NJ: Lawrence Erlbaum Associates.

Mills, A.W. (1958). On the minimum audible angle. *J. Acoust. Soc. Am., 30,* 237-246.

Moore, B.C. (Ed.). (1986). *Frequency selectivity in hearing*. London: Academic Press.

Moore, B. C., & Glasberg, B. R. (1996). A revision of Zwicker's loudness model. *Acustica, 82*, 335-345.

Musicant, A.D., & Butler, R.A. (1984). The influence of pinnae-based spectral cues on sound localization. *J. Acoust. Soc. Am., 75 (4),* 1195-1200.

Perrott, D.R. (1969). Role of signal onset in sound localization. *J. Acoust. Soc. Am., 45 (2),* 436-445.

Perrott, D.R., & Tucker, J. (1988). Minimum audible movement angle as function of signal frequency and the velocity of the source. *J. Acoust. Soc. Am., 83 (4),* 1522-1527.

Pulkki, V., & Karjalainen, M. (2001). Localization of amplitude-panned virtual sources I: stereophonic panning. *J. Audio Eng. Soc., 49 (9),* 739-752.

Rakerd, B., & Hartmann, W.M. (1985). Localization of sound in rooms, II: The effects of a single reflecting surface. *J. Acoust. Soc. Am., 78 (2),* 524-533.

Schmidt, P.H., Van Gemert, A.H., De Fries, R.J, & Duyff, J.W. (1953). Binaural threshold for azimuth differences. *Acta Physiol. Pharmacol. Nederl., 3,* 2-18.

Smyth, M. (1999). *White paper: an overview of the coherent acoustics coding system.* Retrieved October 1, 2001, from http://www.dtsonline.com/whitepaper.pdf

Stevens, S.S. (1955). The measurement of loudness. *J. Acoust. Soc. Am., 27 (5),* 815-829.

Stevens, S.S. (1957). Calculation of loudness of complex noise. *J. Acoust. Soc. Am., 28,* 807-832.

Stevens, S.S. (1961). Procedure for calculating loudness: Mark VI. *J. Acoust. Soc. Am., 33*, 1577-1585.

Stevens, S.S. (1972). Perceived level of noise by mark VII and decibels (E). *J. Acoust. Soc. Am., 51*, 575-601.

Stevens, S.S., & Newman, E.B. (1936). Localization of actual sources of sound. *Amer. J. Psychol., 48,* 297-306.

Strutt, J.W. [Rayleigh, L.] (1907). On our perception of sound direction. *Philosophical Magazine, 6 (13)*, 214-232.

Strybel, T., & Fujimoto, K. (2000). Minimum audible angles in the horizontal and vertical planes: effects of stimulus onset asynchrony and burst duration. *J. Acoust. Soc. Am., 108 (6),* 3092-3095.

Thurlow, W.R., & Marten, A.E. (1962). Perception of steady and intermittent sounds with alternating noise-burst stimuli. *J. Acoust. Soc. Am., 34 (12),* 1853-1858.

Wagenaars, W.M. (1990). Localization of sound in room with reflecting walls. *J. Audio Eng. Soc., 38 (3),* 99-109.

Wallach, A. (1939). On sound localization. *J. Acoust. Soc. Am., 10,* 270-274.

West, J. (1998) *Five-channel panning laws: an analytical experimental comparison.* Unpublished Master's thesis. University of Miami, Florida.

Wightman, F.L., & Kistler, D.J. (1997). Monaural sound localization revisited. *J. Acoust. Soc. Am., 101(2),* 1050-1063.

Wright, D., Hebrank, J.H., & Wilson, B. (1974). Pinna reflections as cues for localization. *J. Acoust. Soc. Am., 56 (3),* 957-962.

Yost, W.A. (1994). The cocktail party problem: Forty years later. In R.H. Gilkey & T.R. Anderson (Eds.), *Binaural and spatial hearing in real and virtual environments.* Mahwah, NJ: Lawrence Erlbaum Associates.

Yost, W.A., & Dye, R.H. (1989). Cross-spectrum effects in localization: parameters of localizing multiple sources. *Sound localization by human observers symposium proceedings.* Washington, D.C.: National Academy Press.

Yost, W.A., & Gourevitch, G. (Eds.). (1987). *Directional hearing.* New York: Springer-Verlag.

Zwicker, E. (1961). Subdivision of the audible frequency range into critical bands. *J. Acoust. Soc. Am., 33 (2),* 248.

Zwicker, E., & Fastl, H. (1990). *Psychoacoustics - facts and models.* Berlin: Springer-Verlag.

Zwicker, E., Flottorp, G., & Stevens, S.S. (1957). Critical band width in loudness summation. *J. Acoust. Soc. Am., 29,* 548-557.

Appendix

Appendix A- Full Data Set - Listening Tests I

MUSIC - E vs ALL - increased image shift right

Test Subject	STR	A	B	C	D	AB
1	0	1	0	1	1	1
2	0	1	0	0	0	0
3	0	1	0	1	0	1
4	0	1	1	1	0	1
5	0	0	0	1	0	1
6	0	1	0	1	0	1
7	0	1	1	0	1	1
8	0	1	1	0	0	1
9	0	1	1	0	0	1
10	0	1	1	0	0	1
11	0	1	0	1	0	1
12	1	1	0	1	1	1
13	0	1	0	1	0	1
14	0	1	1	0	0	1

E vs	STR	A	B	C	D	AB
Total Trials	14	14	14	12	13	14
No Shift	9	0	6	3	5	1
Right	1	13	6	8	3	13
Left	3	0	0	1	3	0
Up	1	4	2	1	2	0
Down	1	0	1	0	1	1
Split	0	0	0	0	0	0
Far	0	0	0	0	0	0
Close	0	0	0	0	0	0
Non-L/R	1	1	2	0	2	0

E vs %	STR	A	B	C	D	AB
Overall % None	64%	0%	43%	25%	38%	7%
Overall % Non-L/R	7%	7%	14%	0%	15%	0%
Sumed Nones	71%	7%	57%	25%	54%	7%
Overall % R	7%	93%	43%	67%	23%	93%
Overall % L	21%	0%	0%	8%	23%	0%

E vs ALL - other shifts

Test Subject	STR	A	B	C	D	AB
1	L	0	S	0	0	0
2	S	0	S	X	X	S
3	S	0	S	0	D	0
4	U,L	0	U	0	S	0
5	S	U	S	0	L	0
6	U	U	U	0	U	0
7	S	0	0	X	0	0
8	L	0	0	S	L	0
9	S	0	0	S	S	0
10	S	0	0	S	S	0
11	S	U	S	0	S	0
12	D	U	D	U	U	D
13	S	0	S	0	L	0
14	S	0	0	L	S	0

E vs	STR	A	B	C	D	AB
Total Trials	14	14	14	12	13	14
None + Non-L/R	10	1	8	3	7	1
Right	1	13	6	8	3	13
Left	3	0	0	1	3	0

*** L = Left, R = Right, U = Up, D = Down, S = Same, SPLT = Split, F = Farther, C = Closer, X = skipped**

MUSIC - DE vs ALL - increased image shift right

Test Subject	STR	A	B	C	AB
1	0	1	0	0	1
2	0	0	0	0	0
3	0	0	0	0	1
4	0	1	0	1	1
5	0	1	1	0	1
6	0	1	0	0	1
7	0	1	1	1	1
8	0	1	0	0	1
9	0	1	1	1	1
10	0	1	1	0	1
11	0	1	0	0	1
12	0	1	1	0	1
13	0	0	0	0	0
14	0	1	0	0	1

Total Trials	13	13	14	14	14
# shift right	0	11	5	3	12
# shift left	3	0	1	0	0
# shift up	1	2	4	5	5
# shift down	0	1	0	0	0
# split	0	2	0	0	0
# far	0	0	0	0	0
# close	0	0	0	0	0
# SAME	9	0	6	6	1
Non-L/R	1	0	2	5	1

DE vs ALL - other shifts

Test Subject	STR	A	B	C	AB
1	S	0	S	S	0
2	X	X	S	U	U
3	S	SPLT	U	U	U
4	L	D	S	0	U
5	S	U	0	S	U
6	U	U	U	U	U
7	L	0	0	0	0
8	S	0	S	S	0
9	S	0	0	0	0
10	S	0	U	U	0
11	S	0	S	S	0
12	L	0	U	S	0
13	S	SPLT	L	U	S
14	S	0	S	S	0

DE vs	STR	A	B	C	AB	ABC
Total Trials	27	27	28	28	28	28
No Shift	16	0	11	11	3	5
Right	1	23	10	10	24	21
Left	8	0	1	0	0	0
Up	4	6	8	8	7	5
Down	0	1	0	1	0	1
Split	0	3	0	0	0	0
Far	0	0	0	1	0	1
Close	0	0	0	0	0	0
Non-L/R	1	1	6	7	1	2

DE vs	STR	A	B	C	AB	ABC
GRAND Total Trials	27	27	28	28	28	28
None + Non-L/R	17	1	17	18	4	7
Right	1	23	10	10	24	21
Left	8	0	1	0	0	0

DE vs	STR	A	B	C	AB	ABC
ADJUSTED GRAND Total Trials	13.5	13.5	14	14	14	14
None	8	0	5.5	5.5	1.5	2.5
Right	0.5	11.5	5	5	12	10.5
Left	4	0	0.5	0	0	0
Up	2	3	4	4	3.5	2.5
Down	0	0.5	0	0.5	0	0.5
Split	0	1.5	0	0	0	0
Far	0	0	0	0.5	0	0.5
Close	0	0	0	0	0	0

DE vs %	STR	A	B	C	AB	ABC
Overall % None	59%	0%	39%	39%	11%	18%
Overall % Non-L/R	4%	4%	21%	25%	4%	7%
Sumed Nones	63%	4%	61%	64%	14%	25%
Overall % R	4%	85%	36%	36%	86%	75%
Overall % L	30%	0%	4%	0%	0%	0%

MUSIC - DE vs ALL - increased image shift right

Test Subject	STR	A	B	C	AB	ABC	ABC
1	0	1	1	0	0	1	1
2	0	1	0	0	0	0	0
3	0	1	0	0	1	1	1
4	0	0	0	0	1	1	1
5	0	1	0	1	1	1	1
6	0	1	0	1	1	1	1
7	1	1	1	1	1	1	1
8	0	1	0	1	1	1	1
9	0	1	1	1	1	1	0
10	0	1	0	0	1	1	1
11	0	1	0	0	1	0	0
12	0	1	1	1	1	1	1
13	0	0	1	1	1	1	0
14	0	1	0	0	1	1	0

Total Trials	14	14	14	14	14	28
# shift right	1	12	5	7	12	21
# shift left	5	0	0	0	0	0
# shift up	3	4	4	3	2	5
# shift down	0	0	0	1	0	1
# split	0	1	0	0	0	0
# far	0	0	0	1	0	1
# close	0	0	0	0	0	0
# SAME	7	0	5	5	2	5
Non-L/R	0	1	4	2	0	2

DE vs ALL - other shifts

Test Subject	STR	A	B	C	AB	ABC	ABC
1	L	0	0	S	S	0	0
2	S	U	S	S	S	S	S
3	U,L	0	U	U	0	0	0
4	L	SPLT	U	S	0	0	0
5	S	0	S	0	0	U	0
6	U,L	U	S	U	U	U	U
7	0	0	0	0	0	0	0
8	S	0	S	0	0	0	0
9	S	0	0	0	0	0	S
10	L	0	U	FAR	0	0	U
11	S	U	U	S	0	S	FAR
12	U,L	0	0	D	U	0	D
13	S	U	0	U	0	0	U
14	S	0	S	S	0	0	S

*** L = Left, R = Right, U = Up, D = Down, S = Same, SPLT = Split, F = Farther, C = Closer,**

MUSIC - CDE vs ALL - increased image shift right

Test Subject	STR	A	B	AB
1	0	1	1	1
2	0	0	0	1
3	0	1	0	1
4	0	0	0	1
5	0	1	0	0
6	0	0	0	1
7	0	1	1	1
8	0	1	0	1
9	0	1	0	0
10	0	1	0	1
11	0	0	0	1
12	0	0	0	0
13	0	0	0	0
14	0	1	0	1

CDE vs	STR	A	B	AB
Total Trials	14	14	14	14
No Shift	1	1	6	2
Right	0	8	2	10
Left	13	1	5	1
Up	3	3	1	4
Down	0	0	0	0
Split	0	1	0	0
Far	0	0	0	0
Close	0	0	0	1
Non-L/R	0	3	1	1

CDE vs %	STR	A	B	AB
Overall % None	7%	7%	43%	14%
Overall % Non-L/R	0%	21%	7%	7%
Sumed Nones	7%	29%	50%	21%
Overall % R	0%	57%	14%	71%
Overall % L	93%	7%	36%	7%

CDE vs ALL - other shifts

Test Subject	STR	A	B	AB
1	L	0	0	0
2	L	U	S	U
3	L	0	S	U
4	L,U	SPLT	L	0
5	L	0	L	S
6	L,U	U	L	0
7	L	0	0	0
8	L	0	S	0
9	L	0	L	S
10	L,U	0	U	C
11	S	U	S	0
12	L	S	S	U
13	L	L	L	L,U
14	L	0	S	0

CDE vs	STR	A	B	AB
Total Trials	14	14	14	14
None + Non-L/R	1	4	7	3
Right	0	8	2	10
Left	13	1	5	1

*** L = Left, R = Right, U = Up, D = Down, S = Same, SPLT = Split, F = Farther, C = Closer, X = skipped**

NOISE - E vs ALL - increased image shift right

Test Subject	STR	A	B	C	D	AB
1	0	1	1	1	0	1
2	0	1	0	0	0	1
3	0	0	0	1	0	1
4	0	1	0	1	0	1
5	0	1	1	1	0	1
6	0	1	0	1	1	1
7	0	1	1	1	1	1
8	0	1	1	1	1	1
9	0	1	1	0	0	1
10	0	0	0	0	1	1
11	0	1	0	1	0	1
12	0	1	1	0	0	1
13	0	1	1	0	0	0
14	0	1	1	1	0	1

E vs	STR	A	B	C	D	AB
Total Trials	14	14	13	14	13	14
No Shift	8	0	1	0	5	0
Right	0	12	8	9	4	13
Left	2	0	0	0	0	0
Up	2	4	4	4	4	0
Down	1	1	1	4	2	0
Split	1	0	0	0	0	1
Far	0	0	0	0	0	0
Close	0	0	0	0	0	0
Non-L/R	3	2	4	5	4	0

E vs %	STR	A	B	C	D	AB
Overall % None	57%	0%	8%	0%	38%	0%
Overall % Non-L/R	21%	14%	31%	36%	31%	0%
Sumed Nones	79%	14%	38%	36%	69%	0%
Overall % R	0%	86%	62%	64%	31%	93%
Overall % L	14%	0%	0%	0%	0%	0%

E vs ALL - other shifts

Test Subject	STR	A	B	C	D	AB
1	S	0	0	D	S	0
2	S	0	X	U	X	0
3	S	U	U	0	S	0
4	L	0	U	U	U	0
5	D	D	0	D	S	0
6	U	0	U	0	U	0
7	S	0	0	0	0	0
8	S	0	0	0	0	0
9	S	0	0	U	S	0
10	U	U	U	U	U	0
11	S	U	S	0	D	0
12	L	U	D	D	D	0
13	SPLT	0	0	D	U	SPLT
14	S	0	0	0	S	0

E vs	STR	A	B	C	D	AB
Total Trials	14	14	13	14	13	14
None + Non-L/R	11	2	5	5	9	0
Right	0	12	8	9	4	13
Left	2	0	0	0	0	0

*** L = Left, R = Right, U = Up, D = Down, S = Same, SPLT = Split, F = Farther, C = Closer, X = skipped**

NOISE - DE vs ALL - increased image shift right

Test Subject	STR	A	B	C	AB
1	0	1	0	0	1
2	0	1	1	0	1
3	0	1	0	0	1
4	0	1	0	1	1
5	0	1	0	0	1
6	0	1	0	0	1
7	0	1	1	1	1
8	0	1	0	1	1
9	0	1	0	0	1
10	0	1	1	1	1
11	0	1	1	0	1
12	0	1	0	1	1
13	0	0	1	0	1
14	0	1	1	1	1

Total Trials	14	14	14	14	14
Right	0	13	6	6	14
Left	8	0	2	0	0
Up	8	4	5	5	0
Down	0	0	1	1	1
Split	0	1	0	0	0
Far	0	0	0	0	0
Close	0	0	0	0	0
None	2	0	2	3	0
Non-L/R	4	0	4	5	0

DE vs ALL - other shifts

Test Subject	STR	A	B	C	AB
1	L,U	0	U	U	0
2	L	U	0	U	0
3	U	U	U	U	0
4	L,U	0	L	U	0
5	U	0	L	U	0
6	U,L	0	U	S	0
7	L	0	0	0	0
8	U	0	S	0	0
9	L	0	S	S	0
10	L	0	U	0	0
11	U	U	0	S	0
12	U,L	U	U	0	D
13	S	SPLT	D	D	0
14	S	0	0	0	0

DE vs	STR	A	B	C	AB	ABC
Total Trials	27	28	27	26	28	28
No Shift	6	0	3	6	0	1
Right	0	25	13	12	27	24
Left	14	0	2	0	0	0
Up	13	7	12	10	3	8
Down	0	0	2	1	2	1
Split	0	2	1	1	0	1
Far	0	0	0	0	0	0
Close	0	0	0	0	0	0
Non-L/R	7	1	8	7	1	2

DE vs	STR	A	B	C	AB	ABC
ADJUSTED GRAND Total Trials	13.5	14	13.5	13	14	14
None	3	0	1.5	3	0	0.5
Right	0	12.5	6.5	6	13.5	12
Left	7	0	1	0	0	0
Up	6.5	3.5	6	5	1.5	4
Down	0	0	1	0.5	1	0.5
Split	0	1	0.5	0.5	0	0.5
Far	0	0	0	0	0	0
Close	0	0	0	0	0	0

DE vs %	STR	A	B	C	AB	ABC
Overall % None	22%	0%	11%	23%	0%	4%
Overall % Non-L/R	26%	4%	30%	27%	4%	7%
Sumed Nones	48%	4%	41%	50%	4%	11%
Overall % R	0%	89%	48%	46%	96%	86%
Overall % L	52%	0%	7%	0%	0%	0%

NOISE - DE vs ALL - increased image shift right

Test Subject	STR	A	B	C	AB	ABC	ABC
1	0	1	1	1	1	1	1
2	0	1	1	0	1	0	0
3	0	1	0	0	0	1	1
4	0	0	1	0	1	1	1
5	0	1	0	0	1	1	1
6	0	1	0	0	1	1	1
7	0	1	1	1	1	1	1
8	0	1	1	1	1	1	1
9	0	1	0	1	1	1	1
10	0	0	0	1	1	1	1
11	0	1	0	0	1	1	1
12	0	1	1	1	1	0	1
13	0	1	0	0	1	1	0
14	0	1	1	0	1	1	1

Total Trials	13	14	13	12	14	28
Right	0	12	7	6	13	24
Left	6	0	0	0	0	0
Up	5	3	7	5	3	8
Down	0	0	1	0	1	1
Split	0	1	1	1	0	1
Far	0	0	0	0	0	0
Close	0	0	0	0	0	0
None	4	0	1	3	0	1
Non-L/R	3	1	4	2	1	2

DE vs	STR	A	B	C	AB	ABC
GRAND Total Trials	27	28	27	26	28	28
None + Non-L/R	13	1	11	13	1	3
Right	0	25	13	12	27	24
Left	14	0	2	0	0	0

DE vs ALL - other shifts

Test Subject	STR	A	B	C	AB	ABC	ABC
1	S	0	0	U	0	0	0
2	X	0	U	X	0	S	U
3	U	U	U	S	U	U	U
4	L	SPLT	U	U	0	U	U
5	U	0	X	X	0	0	0
6	U,L	U	U	U	U	0	D
7	L	0	0	0	0	0	0
8	U,L	0	U	U	0	0	U
9	S	0	D	0	0	0	0
10	L	U	U	0	0	0	0
11	S	0	S	S	0	0	0
12	U	0	U	U	U	U	U
13	L	0	SPLT	SPLT	D	0	SPLT
14	S	0	0	S	0	0	0

*** L = Left, R = Right, U = Up, D = Down, S = Same, SPLT = Split, F = Farther, C = Closer, X = skipped**

NOISE - CDE vs ALL - increased image shift right

Test Subject	STR	A	B	AB
1	0	0	0	1
2	0	0	0	1
3	0	0	0	1
4	0	1	0	1
5	0	1	0	1
6	0	1	0	1
7	0	1	0	1
8	0	1	1	1
9	0	1	1	1
10	0	1	1	1
11	0	1	0	1
12	0	0	0	1
13	0	0	0	1
14	0	1	0	1

CDE vs	STR	A	B	AB
Total Trials	14	14	14	14
No Shift	0	1	2	0
Right	0	9	3	14
Left	14	1	3	0
Up	7	3	9	3
Down	0	0	0	0
Split	0	2	0	0
Far	0	0	0	0
Close	0	0	0	0
Non-L/R	0	1	6	0

CDE vs %	STR	A	B	AB
Overall % None	0%	7%	14%	0%
Overall % Non-L/R	0%	7%	43%	0%
Sumed Nones	0%	14%	57%	0%
Overall % R	0%	64%	21%	100%
Overall % L	100%	7%	21%	0%

CDE vs ALL - other shifts

Test Subject	STR	A	B	AB
1	L	S	U	0
2	L	U	S	0
3	U,L	SPLT	U	U
4	L	0	U	0
5	L	0	U	0
6	U,L	0	U	0
7	L	0	L	0
8	L,U	0	U	0
9	L	0	0	0
10	U,L	0	U	0
11	U,L	U	S	U
12	U,L	U,L	U	0
13	L,U	SPLT	L,U	U
14	L	0	L	0

CDE vs	STR	A	B	AB
Total Trials	14	14	14	14
None + Non-L/R	0	2	8	0
Right	0	9	3	14
Left	14	1	3	0

*** L = Left, R = Right, U = Up, D = Down, S = Same, SPLT = Split, F = Farther, C = Closer, X = skipped**

Appendix B- Full Data Set - Listening Tests II

Madonna track		Sub 1
Trial	X	Ans
1	A	B
2	A	B
3	A	A
4	B	B
5	B	A
6	A	B
7	B	B
8	N/A	N/A
9	N/A	N/A
10	N/A	N/A

Madonna track		Sub 2
Trial	X	Ans
1	A	A
2	B	B
3	A	B
4	A	A
5	A	A
6	B	A
7	N/A	N/A
8	N/A	N/A
9	N/A	N/A
10	N/A	N/A

Madonna track		Sub 3
Trial	X	Ans
1	A	B
2	A	B
3	B	B
4	A	A
5	A	A
6	A	B
7	A	B
8	A	B
9	N/A	N/A
10	N/A	N/A

Madonna track		Sub 4
Trial	X	Ans
1	B	A
2	B	A
3	B	A
4	B	B
5	B	A
6	N/A	N/A
7	N/A	N/A
8	N/A	N/A
9	N/A	N/A
10	N/A	N/A

Madonna track		Sub 5
Trial	X	Ans
1	A	B
2	A	A
3	A	B
4	A	B
5	B	A
6	A	B
7	B	A
8	N/A	N/A
9	N/A	N/A
10	N/A	N/A

Madonna track		Sub 6
Trial	X	Ans
1	N/A	N/A
2	N/A	N/A
3	N/A	N/A
4	N/A	N/A
5	N/A	N/A
6	N/A	N/A
7	N/A	N/A
8	N/A	N/A
9	N/A	N/A
10	N/A	N/A

Madonna track		Sub 7
Trial	X	Ans
1	A	A
2	A	B
3	B	A
4	B	A
5	B	A
6	B	B
7	A	A
8	A	A
9	N/A	N/A
10	N/A	N/A

Madonna track		Sub 8
Trial	X	Ans
1	N/A	N/A
2	N/A	N/A
3	N/A	N/A
4	N/A	N/A
5	N/A	N/A
6	N/A	N/A
7	N/A	N/A
8	N/A	N/A
9	N/A	N/A
10	N/A	N/A

Madonna track		Sub 9
Trial	X	Ans
1	N/A	N/A
2	N/A	N/A
3	N/A	N/A
4	N/A	N/A
5	N/A	N/A
6	N/A	N/A
7	N/A	N/A
8	N/A	N/A
9	N/A	N/A
10	N/A	N/A

Madonna track		Sub 10
Trial	X	Ans
1	A	A
2	A	B
3	B	A
4	N/A	N/A
5	N/A	N/A
6	N/A	N/A
7	N/A	N/A
8	N/A	N/A
9	N/A	N/A
10	N/A	N/A

Madonna track		Sub 11
Trial	X	Ans
1	A	A
2	A	B
3	B	B
4	A	A
5	A	B
6	B	B
7	A	A
8	B	A
9	A	A
10	N/A	N/A

Madonna track		Sub 12
Trial	X	Ans
1	N/A	N/A
2	N/A	N/A
3	N/A	N/A
4	N/A	N/A
5	N/A	N/A
6	N/A	N/A
7	N/A	N/A
8	N/A	N/A
9	N/A	N/A
10	N/A	N/A

Madonna track		Sub 13
Trial	X	Ans
1	N/A	N/A
2	N/A	N/A
3	N/A	N/A
4	N/A	N/A
5	N/A	N/A
6	N/A	N/A
7	N/A	N/A
8	N/A	N/A
9	N/A	N/A
10	N/A	N/A

Madonna track		Sub 14
Trial	X	Ans
1	B	B
2	B	A
3	B	A
4	B	A
5	A	B
6	N/A	N/A
7	N/A	N/A
8	N/A	N/A
9	N/A	N/A
10	N/A	N/A

Madonna track		Sub 15
Trial	X	Ans
1	N/A	N/A
2	N/A	N/A
3	N/A	N/A
4	N/A	N/A
5	N/A	N/A
6	N/A	N/A
7	N/A	N/A
8	N/A	N/A
9	N/A	N/A
10	N/A	N/A

Madonna track		Sub 16
Trial	X	Ans
1	B	B
2	B	A
3	A	B
4	B	B
5	N/A	N/A
6	N/A	N/A
7	N/A	N/A
8	N/A	N/A
9	N/A	N/A
10	N/A	N/A

ABX Test, Track 1 - Madonna																	Grand Total
Sub	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	
Total	7	6	8	5	7	0	8	0	0	3	9	0	0	5	0	4	62
# Correct	3	4	2	1	1	0	4	0	0	1	5	0	0	1	0	2	24

What's Hip?		Sub 1
Trial	X	Ans
1	A	A
2	A	A
3	A	A
4	B	A
5	A	A
6	B	B
7	N/A	N/A
8	N/A	N/A
9	N/A	N/A
10	N/A	N/A

What's Hip?		Sub 2
Trial	X	Ans
1	B	A
2	A	B
3	A	A
4	B	B
5	B	B
6	A	B
7	N/A	N/A
8	N/A	N/A
9	N/A	N/A
10	N/A	N/A

What's Hip?		Sub 3
Trial	X	Ans
1	N/A	N/A
2	N/A	N/A
3	N/A	N/A
4	N/A	N/A
5	N/A	N/A
6	N/A	N/A
7	N/A	N/A
8	N/A	N/A
9	N/A	N/A
10	N/A	N/A

What's Hip?		Sub 4
Trial	X	Ans
1	N/A	N/A
2	N/A	N/A
3	N/A	N/A
4	N/A	N/A
5	N/A	N/A
6	N/A	N/A
7	N/A	N/A
8	N/A	N/A
9	N/A	N/A
10	N/A	N/A

What's Hip?		Sub 5
Trial	X	Ans
1	B	A
2	B	A
3	A	B
4	A	A
5	A	A
6	B	B
7	A	B
8	A	B
9	N/A	N/A
10	N/A	N/A

What's Hip?		Sub 6
Trial	X	Ans
1	A	A
2	A	A
3	A	B
4	B	B
5	B	A
6	A	A
7	B	B
8	A	A
9	N/A	N/A
10	N/A	N/A

What's Hip?		Sub 7
Trial	X	Ans
1	N/A	N/A
2	N/A	N/A
3	N/A	N/A
4	N/A	N/A
5	N/A	N/A
6	N/A	N/A
7	N/A	N/A
8	N/A	N/A
9	N/A	N/A
10	N/A	N/A

What's Hip?		Sub 8
Trial	X	Ans
1	A	A
2	A	B
3	B	B
4	A	A
5	B	B
6	A	B
7	B	A
8	B	B
9	A	A
10	N/A	N/A

What's Hip?		Sub 9
Trial	X	Ans
1	B	A
2	B	A
3	B	B
4	A	A
5	A	B
6	A	A
7	A	A
8	B	B
9	A	B
10	N/A	N/A

What's Hip?		Sub 10
Trial	X	Ans
1	B	B
2	A	B
3	A	A
4	A	B
5	A	A
6	B	A
7	B	A
8	A	B
9	A	B
10	A	A

What's Hip?		Sub 11
Trial	X	Ans
1	N/A	N/A
2	N/A	N/A
3	N/A	N/A
4	N/A	N/A
5	N/A	N/A
6	N/A	N/A
7	N/A	N/A
8	N/A	N/A
9	N/A	N/A
10	N/A	N/A

What's Hip?		Sub 12
Trial	X	Ans
1	B	A
2	B	A
3	B	A
4	B	A
5	B	B
6	A	B
7	A	A
8	B	A
9	A	B
10	B	A

What's Hip?		Sub 13
Trial	X	Ans
1	A	B
2	B	B
3	B	B
4	A	A
5	A	A
6	A	B
7	A	B
8	B	A
9	B	B
10	N/A	N/A

What's Hip?		Sub 14
Trial	X	Ans
1	N/A	N/A
2	N/A	N/A
3	N/A	N/A
4	N/A	N/A
5	N/A	N/A
6	N/A	N/A
7	N/A	N/A
8	N/A	N/A
9	N/A	N/A
10	N/A	N/A

What's Hip?		Sub 15
Trial	X	Ans
1	A	A
2	A	B
3	B	A
4	B	B
5	A	A
6	B	B
7	B	A
8	B	A
9	N/A	N/A
10	N/A	N/A

What's Hip?		Sub 16
Trial	X	Ans
1	N/A	N/A
2	N/A	N/A
3	N/A	N/A
4	N/A	N/A
5	N/A	N/A
6	N/A	N/A
7	N/A	N/A
8	N/A	N/A
9	N/A	N/A
10	N/A	N/A

ABX Test, Track 2 - What is Hip?																	**Grand Total**
Sub	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	
Total	6	6	0	0	8	8	0	9	9	10	0	10	9	0	8	0	83
# Correct	5	3	0	0	3	6	0	6	4	4	0	2	5	0	3	0	41

Noise vs Bandwidth		Sub 1							
Trial		Fc 1	Ans	Fc 2	Ans	Fc 3	Ans	Fc 4	Ans
1	A vs High	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
2	AB vs High	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
4	E vs Low	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
5	DE vs Low	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
6	CDE vs Low	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A

Noise vs Bandwidth		Sub 2							
Trial		Fc 1	Ans	Fc 2	Ans	Fc 3	Ans	Fc 4	Ans
1	A vs High	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
2	AB vs High	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
4	E vs Low	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
5	DE vs Low	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
6	CDE vs Low	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A

Noise vs Bandwidth		Sub 3							
Trial		Fc 1	Ans	Fc 2	Ans	Fc 3	Ans	Fc 4	Ans
1	A vs High	10k	H LDR	8k	H LDR	12k	H SFTR	N/A	N/A
2	AB vs High	10k	H SFTR	8k	SAME	6k	H LDR	N/A	N/A
4	E vs Low	200	L SFTR	500	SAME	700	L LDR	N/A	N/A
5	DE vs Low	1k	L SFTR	3k	L LDR	2k	SAME	N/A	N/A
6	CDE vs Low	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A

Noise vs Bandwidth		Sub 4							
Trial		Fc 1	Ans	Fc 2	Ans	Fc 3	Ans	Fc 4	Ans
1	A vs High	10k	H LDR	11k	H LDR	12.5k	SAME	N/A	N/A
2	AB vs High	10k	SAME	8k	H LDR	11k	SAME	N/A	N/A
4	E vs Low	500	L SFTR	700	L SFTR	1.5k	L LDR	N/A	N/A
5	DE vs Low	5k	SAME	4k	SAME	7k	SAME	2k	L SFTR
6	CDE vs Low	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A

Noise vs Bandwidth		Sub 5							
Trial		Fc 1	Ans	Fc 2	Ans	Fc 3	Ans	Fc 4	Ans
1	A vs High	10k	H LDR	6k	H LDR	12k	H SFTR	11k	H SFTR
2	AB vs High	800	H LDR	1.5k	H SFTR	1.3k	H LDR	N/A	N/A
4	E vs Low	800	L LDR	600	L SFTR	700	SAME	N/A	N/A
5	DE vs Low	2k	SAME	4k	L LDR	1k	L SFTR	N/A	N/A
6	CDE vs Low	5k	SAME	3k	L LDR	1.5k	L SFTR	N/A	N/A

Noise vs Bandwidth		Sub 6							
Trial		Fc 1	Ans	Fc 2	Ans	Fc 3	Ans	Fc 4	Ans
1	A vs High	11k	H LDR	13k	H LDR	15k	H LDR	N/A	N/A
2	AB vs High	1.3k	H LDR	3k	H LDR	7k	H LDR	9k	H LDR
4	E vs Low	700	SAME	300	L SFTR	500	L SFTR	N/A	N/A
5	DE vs Low	4k	SAME	6k	L LDR	2k	L SFTR	N/A	N/A
6	CDE vs Low	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A

Noise vs Bandwidth		Sub 7							
Trial		Fc 1	Ans	Fc 2	Ans	Fc 3	Ans	Fc 4	Ans
1	A vs High	11k	H LDR	13k	H LDR	15k	SAME	14.5k	SAME
2	AB vs High	8k	H LDR	10k	H LDR	12k	H LDR	15k	SAME
4	E vs Low	500	L SFTR	900	SAME	1.5k	SAME	N/A	N/A
5	DE vs Low	4k	SAME	6k	SAME	2k	L SFTR	N/A	N/A
6	CDE vs Low	2k	L SFTR	4k	L SFTR	7k	SAME	N/A	N/A

Noise vs Bandwidth		Sub 8							
Trial		Fc 1	Ans	Fc 2	Ans	Fc 3	Ans	Fc 4	Ans
1	A vs High	14K	H SFTR	12k	H LDR	11k	SAME	9k	H LDR
2	AB vs High	12k	SAME	10k	SAME	8k	H LDR	N/A	N/A
4	E vs Low	900	L LDR	500	L SFTR	700	SAME	1k	L LDR
5	DE vs Low	2k	SAME	900	L SFTR	5k	H LDR	N/A	N/A
6	CDE vs Low	6k	SAME	9k	L LDR	6k	SAME	N/A	N/A

Noise vs Bandwidth		Sub 9							
Trial		Fc 1	Ans	Fc 2	Ans	Fc 3	Ans	Fc 4	Ans
1	A vs High	12k	H LDR	14k	H SFTR	13k	H LDR	11k	H LDR
2	AB vs High	10k	H SFTR	8k	H LDR	9k	H LDR	N/A	N/A
4	E vs Low	800	L SFTR	1.2k	SAME	3k	L LDR	1.3k	L LDR
5	DE vs Low	3k	L SFTR	4k	SAME	5k	SAME	N/A	N/A
6	CDE vs Low	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A

Noise vs Bandwidth		Sub 10							
Trial		Fc 1	Ans	Fc 2	Ans	Fc 3	Ans	Fc 4	Ans
1	A vs High	12k	H LDR	10k	H LDR	15k	H SFTR	N/A	N/A
2	AB vs High	10k	H LDR	13k	H LDR	14k	SAME	N/A	N/A
4	E vs Low	800	L SFTR	1k	L SFTR	N/A	N/A	N/A	N/A
5	DE vs Low	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
6	CDE vs Low	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A

Noise vs Bandwidth		Sub 11							
Trial		Fc 1	Ans	Fc 2	Ans	Fc 3	Ans	Fc 4	Ans
1	A vs High	12k	H LDR	14k	SAME	15k	SAME	N/A	N/A
2	AB vs High	10k	SAME	8k	H LDR	9k	H SFTR	N/A	N/A
4	E vs Low	800	L SFT	1k	SAME	2k	L LDR	N/A	N/A
5	DE vs Low	3k	L LDR	1k	L SFTR	2k	SAME	N/A	N/A
6	CDE vs Low	6k	SAME	4k	L SFTR	5k	SAME	N/A	N/A

Noise vs Bandwidth		Sub 12							
Trial		Fc 1	Ans	Fc 2	Ans	Fc 3	Ans	Fc 4	Ans
1	A vs High	8k	H LDR	5k	H LDR	10k	H SFTR	N/A	N/A
2	AB vs High	8k	SAME	6k	SAME	2k	H LDR	N/A	N/A
4	E vs Low	10k	L LDR	2k	L SFTR	N/A	N/A	N/A	N/A
5	DE vs Low	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
6	CDE vs Low	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A

Noise vs Bandwidth		Sub 13							
Trial		Fc 1	Ans	Fc 2	Ans	Fc 3	Ans	Fc 4	Ans
1	A vs High	12k	H LDR	14k	H LDR	15k	SAME	N/A	N/A
2	AB vs High	10k	H LDR	8k	H LDR	12k	H LDR	14k	SAME
4	E vs Low	700	L SFTR	1k	L SFTR	1.3k	L SFTR	3k	SAME
5	DE vs Low	3k	L SFTR	5k	LSFTR	6.5k	L SFTR	7.5k	L SFTR
6	CDE vs Low	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A

Noise vs Bandwidth		Sub 14							
Trial		Fc 1	Ans	Fc 2	Ans	Fc 3	Ans	Fc 4	Ans
1	A vs High	14k	H LDR	15k	H LDR	N/A	N/A	N/A	N/A
2	AB vs High	12k	H LDR	14k	SAME	N/A	N/A	N/A	N/A
4	E vs Low	1k	L SFTR	3k	L LDR	2k	SAME	N/A	N/A
5	DE vs Low	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
6	CDE vs Low	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A

Noise vs Bandwidth		Sub 15							
Trial		Fc 1	Ans	Fc 2	Ans	Fc 3	Ans	Fc 4	Ans
1	A vs High	14k	SAME	11k	H LDR	N/A	N/A	N/A	N/A
2	AB vs High	12k	H LDR	13k	SAME	N/A	N/A	N/A	N/A
4	E vs Low	1k	L SFTR	3k	L LDR	2k	L SFTR	N/A	N/A
5	DE vs Low	3k	L SFTR	3k	SAME	N/A	N/A	N/A	N/A
6	CDE vs Low	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A

Appendix C- Matlab Files

1) Energy Distribution Matlab File

```
% Rob Hartman
% Thesis - Energy Distribution around 10000 Hz
% Last Modified: 9/21/02
%
% This program calculates the total energy of the signal, and finds the percentage of energy above and
% below 10000 Hz
%

clear
clc
close all

screen = get(0,'ScreenSize');
size = [(screen(3)/2)-275, (screen(4)/2)-20, 450, 100];
% Uncomment the following lines to be prompted by an acutal window, instead of just the title
parent = figure('MenuBar','none','Name','Hartman Thesis','NumberTitle','off','Position', size);
% Create text string in parent window
position = [1, size(4)-50, size(3), 25];
text = uicontrol('FontSize',15,'String','Please choose a *.WAV file to ANALYZE...','Parent', parent,'Style',
'text','Position', position,'HorizontalAlignment','center');
figure(parent);

% Prompt the user to select the desired WAV file for analysis

[filename, pathname] = uigetfile('*.wav','Choose WAV file to ANALYZE',screen(4)/2,screen(3)/2);
addpath(pathname);
[y,fs,trash]=wavread(filename);  % stereo WAVs are read into 2 columns. col 1 is left, col 2 is right.

% Set the FFT window size, overlap, and time vectors

N=2048; % size of FFT window
overlap=floor(.5*N);
win=hamming(N);
t=0:overlap/fs:(length(y)-N)/fs;
t2=0:1/fs:(length(y)-1)/fs;

% Calculate the index values corresponding to the various frequency bands low and high cutoffs
% A = 80-800 Hz, B = 800-1600 Hz, C = 1600-5000 Hz, D = 5000-12000 Hz, E = 12000-20000 Hz
% Also add F for 10 kHz results

N_A1=ceil(80/(fs/N));
N_E2=ceil(20000/(fs/N));
N_F=ceil(10000/(fs/N));


% Calculate FFT and energy for each band of Left/Right CHANNEL using windows and overlap of the
% WAV file

n=0;
for i=1:overlap:(length(y)-N),
  n=n+1;
```

```
  data=y(i:i+N-1,1).*win; % Left Channel
  fftdata=abs(fft(data,N));
  fftmag=fftdata(1:end/2).^2;
  totenergyLT(n)=sum(fftmag(N_A1:N_E2));
  bandFLenergyLT(n)=sum(fftmag(N_A1:N_F-1));
  bandFHenergyLT(n)=sum(fftmag(N_F:N_E2));

  data=y(i:i+N-1,2).*win; % Right Channel
  fftdata=abs(fft(data,N));
  fftmag=fftdata(1:end/2).^2;
  totenergyRT(n)=sum(fftmag(N_A1:N_E2));
  bandFLenergyRT(n)=sum(fftmag(N_A1:N_F-1));
  bandFHenergyRT(n)=sum(fftmag(N_F:N_E2));
end

% Find the max value, in order to standardize the plots

lngthmax=max(t);

htmax2=max(y);
htmax2=max(htmax2); % max of temporal signal plot
lngthmax2=lngthmax;

htmax3(1)=max(totenergyLT); % max total energy
htmax3(2)=max(totenergyRT);
htmax3=max(htmax3);

% Plot the temporal left stereo signal and total energy results

figure
subplot(2,2,1),
plot(t2,y(1:end,1))
title(' Music Passage, Left Channel' )
ylabel(' Amplitude' )
ax2=axis;
ax2(4)=htmax2+.1;
ax2(3)=-htmax2-.1;
ax2(2)=lngthmax2;
axis([ax2]);

subplot(2,2,3),
plot(t,totenergyLT)
title(' Total Energy, Left Channel' )
ylabel(' Energy' )
xlabel(' Time in seconds (s)' )
ax3=axis;
ax3(4)=htmax3;
ax3(2)=lngthmax;
axis([ax3]);

subplot(2,2,2),
plot(t2,y(1:end,2))
title(' Music Passage, Right channel' )
axis([ax2]);

subplot(2,2,4),
```

```
plot(t,totenergyRT)
title('Total Energy, Right channel')
xlabel('Time in seconds (s)')
axis([ax3]);

maxFL(1)=max(bandFLenergyLT);
maxFL(2)=max(bandFLenergyRT);
maxFH(1)=max(bandFHenergyLT);
maxFH(2)=max(bandFHenergyRT);
maxFL=max(maxFL);
maxFH=max(maxFH);

figure
subplot(2,2,1),
plot(t,bandFLenergyLT)
title('Sub-band Energy of Left & Right channel')
ax=axis;
ax(4)=maxFL;
ax(2)=lngthmax;
axis([ax]);
xlabel('Left-Time in seconds (s)')
ylabel('Band under 10 kHz')
subplot(2,2,2),
plot(t,bandFLenergyRT)
axis([ax]);
xlabel('Right-Time in seconds (s)')

subplot(2,2,3),
plot(t,bandFHenergyLT)
ax=axis;
ax(4)=maxFH;
ax(2)=lngthmax;
axis([ax]);
xlabel('Left-Time in seconds (s)')
ylabel('Band 10 kHz and above')
subplot(2,2,4),
plot(t,bandFHenergyRT)
axis([ax]);
xlabel('Right-Time in seconds (s)')

% Now calculate percentage of energy above 100000 Hz

for a=1:n,
  if totenergyLT(a)>0
    prcntenrgyblw10kLT(a)=ceil(100*bandFLenergyLT(a)/totenergyLT(a));
    prcntenrgyabov10kLT(a)=ceil(100*bandFHenergyLT(a)/totenergyLT(a));
  else
    prcntenrgyabov10kLT(a)=0;
    prcntenrgyabov10kLT(a)=0;
  end

  if totenergyRT(a)>0
       prcntenrgyblw10kRT(a)=ceil(100*bandFLenergyRT(a)/totenergyRT(a));
    prcntenrgyabov10kRT(a)=ceil(100*bandFHenergyRT(a)/totenergyRT(a));
  else
    prcntenrgyblw10kRT(a)=0;
```

```
    prcntenrgyabov10kRT(a)=0;
  end
end


avgeabv10k=(mean(prcntenrgyabov10kLT)+ mean(prcntenrgyabov10kRT))/2
maxLT=max(prcntenrgyabov10kLT)
maxrt=max(prcntenrgyabov10kRT)

figure
subplot(2,2,1)
stairs(prcntenrgyabov10kLT)
xlabel('Energy above 10 kHz, Left')
subplot(2,2,2)
stairs(prcntenrgyabov10kRT)
xlabel('Energy above 10 kHz, Right')
subplot(2,2,3)
stairs(prcntenrgyblw10kLT)
xlabel('Energy below 10 kHz, Left')
subplot(2,2,4)
stairs(prcntenrgyblw10kRT)
xlabel('Energy below 10 kHz, Right')
```

2) Power Spectral Density Matlab File

```
% Rob Hartman
% Thesis - Power Spectral Density and Energy Distribution
% Last Modified: 9/21/02
%
%   This program prompts the user for a *.WAV file.  It then calculates the FFT and the index
%  of the frequency points of the 5 spatially relocated Subbands (A,B,C,D,E).  It then calculates the energy
%  in each subband (and total energy) over windows of time, using Parseval's (sum of FFT values) for each
%  value.  It normalizes the results and plots two figures. One figure is a 4-window subplot showing the left
%  and right temporal signals, followed by the left and right total energy values vs. time.  The second figure
is
%  the energy, vs. time, of each subband.
%
%   At the end, it calculates the energy in bands DE, CDE, AB, and ABC.  It plots results in two different
%   figures - one for high freq. DE/CDE and one for low freq. AB/ABC.
%
%  Also, similar calculations for the energy above and below 10 kHz were later added for List Test II
%   Results and Analysis
%

clear
clc
close all

screen = get(0,'ScreenSize');
size = [(screen(3)/2)-275, (screen(4)/2)-20, 450, 100];
% Uncomment the following lines to be prompted by an acutal window, instead of just the title
parent = figure('MenuBar','none','Name','Hartman Thesis','NumberTitle','off','Position', size);
% Create text string in parent window
```

```
position = [1, size(4)-50, size(3), 25];
text = uicontrol(' FontSize' ,15,' String' , ' Please choose a *.WAV file to ANALYZE...' , ' Parent' , parent, ' Style' ,
' text' , ' Position' , position, ' HorizontalAlignment' , ' center' );
figure(parent);

% Prompt the user to select the desired WAV file for analysis

[filename, pathname] = uigetfile(' *.wav' , ' Choose WAV file to ANALYZE' ,screen(4)/2,screen(3)/2);
addpath(pathname);
[y,fs,trash]=wavread(filename);  % stereo WAVs are read into 2 columns. col 1 is left, col 2 is right.

% Set the FFT window size, overlap, and time vectors

N=2048; % size of FFT window
overlap=floor(.5*N);
win=hamming(N);
t=0:overlap/fs:(length(y)-N)/fs;
t2=0:1/fs:(length(y)-1)/fs;

% Calculate the index values corresponding to the various frequency bands low and high cutoffs
% A = 80-800 Hz, B = 800-1600 Hz, C = 1600-5000 Hz, D = 5000-12000 Hz, E = 12000-20000 Hz
% Also add F for 10 kHz results

N_A1=ceil(80/(fs/N));
N_A2=ceil(800/(fs/N));
N_B1=N_A2+1;
N_B2=ceil(1600/(fs/N));
N_C1=N_B2+1;
N_C2=ceil(5000/(fs/N));
N_D1=N_C2+1;
N_D2=ceil(12000/(fs/N));
N_E1=N_D2+1;
N_E2=ceil(20000/(fs/N));

N_F=ceil(10000/(fs/N));


% Calculate FFT and energy for each band of Left/Right CHANNEL using windows and overlap

n=0;
for i=1:overlap:(length(y)-N),
  n=n+1;
  data=y(i:i+N-1,1).*win; % Left Channel
  fftdata=abs(fft(data,N));
  fftmag=fftdata(1:end/2).^2;
  totenergyLT(n)=sum(fftmag(N_A1:N_E2));
  bandAenergyLT(n)=sum(fftmag(N_A1:N_A2));
  bandBenergyLT(n)=sum(fftmag(N_B1:N_B2));
  bandCenergyLT(n)=sum(fftmag(N_C1:N_C2));
  bandDenergyLT(n)=sum(fftmag(N_D1:N_D2));
  bandEenergyLT(n)=sum(fftmag(N_E1:N_E2));

  bandFLenergyLT(n)=sum(fftmag(N_A1:N_F-1));
  bandFHenergyLT(n)=sum(fftmag(N_F:N_E2));

  data=y(i:i+N-1,2).*win; % Right Channel
```

```
  fftdata=abs(fft(data,N));
  fftmag=fftdata(1:end/2).^2;
  totenergyRT(n)=sum(fftmag(N_A1:N_E2));
  bandAenergyRT(n)=sum(fftmag(N_A1:N_A2));
  bandBenergyRT(n)=sum(fftmag(N_B1:N_B2));
  bandCenergyRT(n)=sum(fftmag(N_C1:N_C2));
  bandDenergyRT(n)=sum(fftmag(N_D1:N_D2));
  bandEenergyRT(n)=sum(fftmag(N_E1:N_E2));

  bandFLenergyRT(n)=sum(fftmag(N_A1:N_F-1));
  bandFHenergyRT(n)=sum(fftmag(N_F:N_E2));
end

% Find the max value, in order to standardize the plots

htmax(1)=max(bandAenergyLT);
htmax(2)=max(bandBenergyLT);
htmax(3)=max(bandCenergyLT);
htmax(4)=max(bandDenergyLT);
htmax(5)=max(bandEenergyLT);
htmax(6)=max(bandAenergyRT);
htmax(7)=max(bandBenergyRT);
htmax(8)=max(bandCenergyRT);
htmax(9)=max(bandDenergyRT);
htmax(10)=max(bandEenergyRT);

htmax=max(htmax); % max of left/right band energy
lngthmax=max(t);

htmax2=max(y);
htmax2=max(htmax2); % max of temporal signal plot
lngthmax2=max(t2);

htmax3(1)=max(totenergyLT); % max total energy
htmax3(2)=max(totenergyRT);
htmax3=max(htmax3);

% Plot the temporal left stereo signal and total energy results

figure
subplot(2,2,1),
plot(t2,y(1:end,1))
title('Music Passage, Left Channel')
ylabel('Amplitude')
ax2=axis;
ax2(4)=htmax2+.1;
ax2(3)=-htmax2-.1;
ax2(2)=lngthmax2;
axis([ax2]);

subplot(2,2,3),
plot(t,totenergyLT)
title('Total Energy, Left Channel')
ylabel('Energy')
xlabel('Time in seconds (s)')
ax3=axis;
```

```
ax3(4)=htmax3;
ax3(2)=lngthmax;
axis([ax3]);

subplot(2,2,2),
plot(t2,y(1:end,2))
title('Music Passage, Right channel')
axis([ax2]);

subplot(2,2,4),
plot(t,totenergyRT)
title('Total Energy, Right channel')
xlabel('Time in seconds (s)')
axis([ax3]);

% Plot the energy of each of the five subbands

figure
subplot(5,1,1),
plot(t,bandAenergyLT)
%ax = axis; % finds the current axis settings
%ax(4)=htmax; % replaces the current height w/ the max. height
%ax(2)=lngthmax;
%axis([ax]);
ylabel('Band A')
title('Suband Energy of Left(solid) & Right(dotted) channel')
hold
plot(t,bandAenergyRT,'m:')
hold

subplot(5,1,2),
plot(t,bandBenergyLT)
%axis([ax]);
ylabel('Band B')
hold
plot(t,bandBenergyRT,'m:')
hold

subplot(5,1,3),
plot(t,bandCenergyLT)
%axis([ax]);
ylabel('Band C')
hold
plot(t,bandCenergyRT,'m:')
hold

subplot(5,1,4),
plot(t,bandDenergyLT)
%axis([ax]);
ylabel('Band D')
hold
plot(t,bandDenergyRT,'m:')
hold

subplot(5,1,5),
plot(t,bandEenergyLT)
```

```
%axis([ax]);
xlabel('   Time in seconds (s)'   )
ylabel('   Band E'   )
hold
plot(t,bandEenergyRT,'   m:'   )


bandABenergyLT=bandAenergyLT+bandBenergyLT;
bandABenergyRT=bandAenergyRT+bandBenergyRT;

bandABCenergyLT=bandABenergyLT+bandCenergyLT;
bandABCenergyRT=bandABenergyRT+bandCenergyRT;

maxAB(1)=max(bandABenergyLT);
maxAB(2)=max(bandABenergyRT);
maxAB(3)=max(bandABCenergyLT);
maxAB(4)=max(bandABCenergyRT);
maxAB=max(maxAB);

figure
subplot(2,1,1),
plot(t,bandABenergyLT)
title('   Suband Energy of Left(solid) & Right(dotted) channel'   )
%ax = axis;
%ax(4)= maxAB;
%axis([ax]);
xlabel('   Time in seconds (s)'   )
ylabel('   Band AB'   )
hold
plot(t,bandABenergyRT,'   m:'   )
hold

subplot(2,1,2),
plot(t,bandABCenergyLT)
%axis([ax]);
xlabel('   Time in seconds (s)'   )
ylabel('   Band ABC'   )
hold
plot(t,bandABCenergyRT,'   m:'   )

bandDEenergyLT=bandDenergyLT+bandEenergyLT;
bandDEenergyRT=bandDenergyRT+bandEenergyRT;

bandCDEenergyLT=bandCenergyLT+bandDenergyLT+bandEenergyLT;
bandCDEenergyRT=bandCenergyLT+bandDenergyRT+bandEenergyRT;

maxDE(1)=max(bandDEenergyLT);
maxDE(2)=max(bandDEenergyRT);
maxDE(3)=max(bandCDEenergyLT);
maxDE(4)=max(bandCDEenergyRT);
maxDE=max(maxDE);

figure
subplot(2,1,1),
plot(t,bandDEenergyLT)
title('   Suband Energy of Left(solid) & Right(dotted) channel'   )
```

```
%ax = axis;
%ax(4)= maxDE;
%axis([ax]);
xlabel('Time in seconds (s)')
ylabel('Band DE')
hold
plot(t,bandDEenergyRT,'m:')
hold

subplot(2,1,2),
plot(t,bandCDEenergyLT)
%axis([ax]);
xlabel('Time in seconds (s)')
ylabel('Band CDE')
hold
plot(t,bandCDEenergyRT,'m:')

maxFL(1)=max(bandFLenergyLT);
maxFL(2)=max(bandFLenergyRT);
maxFH(1)=max(bandFHenergyLT);
maxFH(2)=max(bandFHenergyRT);
maxFL=max(maxFL);
maxFH=max(maxFH);

figure
subplot(2,2,1),
plot(t,bandFLenergyLT)
title('Suband Energy of Left & Right channel')
ax=axis;
ax(4)=maxFL;
ax(2)=lngthmax;
axis([ax]);
xlabel('Left - Time in seconds (s)')
ylabel('Band under 10 kHz')
subplot(2,2,2),
plot(t,bandFLenergyRT)
axis([ax]);
xlabel('Right - Time in seconds (s)')

subplot(2,2,3),
plot(t,bandFHenergyLT)
ax=axis;
ax(4)=maxFH;
ax(2)=lngthmax;
axis([ax]);
xlabel('Left - Time in seconds (s)')
ylabel('Band 10 kHz and above')
subplot(2,2,4),
plot(t,bandFHenergyRT)
axis([ax]);
xlabel('Right - Time in seconds (s)')
```

<u>3) Impulse Response Matlab File</u>

```
% Rob Hartman
% Thesis - Impulse Response Analysis of Sample files
% Originated: 7/6/02
```

```
clear
clc
close all

screen = get(0,'ScreenSize');

% Setup preliminary variables

N=2048; % size of FFT window
overlap=floor(.5*N);
win=hamming(N);

% Perform freq. analysis for four Room Response recordings

m=0;
n=0;
fftsum=0;
finalfft=0;

while m<4 % there are four recordings that need averaged
  m=m+1;

  % Use pop-up window to allow user to select the file
  [filename, pathname] = uigetfile('*.wav','Choose "Room" Response WAV file to
ANALYZE',screen(4)/2,screen(3)/2);
     addpath(pathname);
  [y,fs,trash] = wavread(filename);

  yy=zeros(length(y)+2*overlap,1);% pad with zeros to avoid window effects at start and stop of data
  yy(overlap+1:overlap+length(y))=y;
  y=yy;

  % Take the FFT
     fftavg=0;
     fftdata=0;
     n=0;

     for i=1:overlap:(length(y)-N)
       n=n+1;
       data=y(i:i+N-1).*win;
       fftdata=abs(fft(data));
       fftavg=fftavg+(fftdata(1:N/2)); % sum the overlapped windows of FFT data
  end

  fftsum=fftsum+fftavg;     % Need to average the resutling fft of each recording
end

finalfft_room=fftsum/m; % Actual average

% Plot Room results
f=logspace(log10(fs/N),log10(length(y)*fs/N),N/2);
ydB=20*log10(finalfft_room);

figure
subplot(2,1,1)
```

```
semilogx(f,ydB);
title('  "Room" Average Impulse Response'  )
axis([20 20000 min(ydB)*1.1 max(ydB)*1.2])
v=axis;
xlabel('  Frequency in Hertz (Hz)'  )
ylabel('  Magnitude in Decibels (dB)'  )

% Now perform for Free Field recorded responses

m=0;
n=0;
fftsum=0;
finalfft=0;

while m<4 % Again,perform for 4 recorded files
  m=m+1;

  % Use pop-up window to allow user to select the file
  [filename, pathname] = uigetfile('  *.wav'  , '  Choose "Field" Response WAV file to
ANALYZE'  ,screen(4)/2,screen(3)/2);
     addpath(pathname);
  [y,fs,trash] = wavread(filename);

  yy=zeros(length(y)+2*overlap,1);% pad with zeros for windowing
  yy(overlap+1:overlap+length(y))=y;
  y=yy;

  % Take the FFT
  fftavg=0;
     fftdata=0;
     n=0;

     for i=1:overlap:(length(y)-N)
       n=n+1;
       data=y(i:i+N-1).*win;
       fftdata=abs(fft(data));
    fftavg=fftavg+(fftdata(1:N/2)); % sum the overlapped windows of FFT data
  end

  fftsum=fftsum+fftavg;     % keep running total for averaging
end

finalfft_field=fftsum/m; % acutal averaging

% Plot Free Field results

f=logspace(log10(fs/N),log10(length(y)*fs/N),N/2);
ydB2=20*log10(finalfft_field);

subplot(2,1,2)
semilogx(f,ydB2);
title('  "Field" Average Impulse Response'  )
axis([v])
xlabel('  Frequency in Hertz (Hz)'  )
ylabel('  Magnitude in Decibels (dB)'  )
```

```
% Now calculate actual room response, dividing out equipment (field) response

yfinal=finalfft_room./finalfft_field;
yfinal=20*log10(yfinal);

% Plot final room response

figure
subplot(2,1,1)
semilogx(f,yfinal);
title('Adjusted "Actual" Room Impulse Reponse')
axis([20 20000 min(yfinal) max(yfinal)])
v=axis;
xlabel('Frequency in Hertz (Hz)')
ylabel('Magnitude in Decibels (dB)')

% Smooth out final response with averaging

yfinalavg=yfinal;
for i=1:length(yfinal)
  if i<16
    yfinalavg(i)=mean(yfinal(i:i+14));
  elseif i>length(yfinal)-16
    yfinalavg(i)=mean(yfinal(i:i-14));
     else
    yfinalavg(i)=mean(yfinal(i-15:i+15));
  end
end

subplot(2,1,2)
semilogx(f,yfinalavg);
axis([v])
title('AVERAGED Actual Room Impulse Reponse')
xlabel('Frequency in Hertz (Hz)')
ylabel('Magnitude in Decibels (dB)')
```

<u>4) Spectrogram Matlab File</u>

```
% Rob Hartman
% Thesis - Spectrogram and Frequency Spectrum of Sample file
% Last Modified: 9/21/02
%
%   This program calculates and displays the spectrogram of a *.WAV file.  It prompts the user to select
%   the file to be analyzed.  It then, calculates the spectrogram directly from the left and right channel data.


clear
clc
close all

N=2048; % size of FFT window
win=hamming(N);
```

```
screen = get(0,'ScreenSize');

[filename, pathname] = uigetfile('*.wav','Choose WAV file to ANALYZE',screen(4)/2,screen(3)/2);
addpath(pathname);
[y,fs,trash] = wavread(filename); % stereo WAVS are read into 2 columns. col 1 is left, col 2 is right.

[blft,flft,tlft]= specgram(y(1:end,1),N,fs,win,floor(.75*N));
[blrt,flrt,tlrt]= specgram(y(1:end,2),N,fs,win,floor(.75*N));

figure

subplot(2,1,1)
imagesc(tlft,flft,20*log10(abs(blft)));
xlabel('Left Channel - Time in seconds (s)')
title('Spectrogram of Test Track Two - What is Hip?')
axis xy
colormap(jet)

subplot(2,1,2)
imagesc(tlrt,flrt,20*log10(abs(blrt)));
axis xy
colormap(jet)
xlabel('Right Channel - Time in seconds (s)')
ylabel('Frequency in Hertz (Hz)')
```

5) Loudspeaker Frequency Response Matlab File

```
% Rob Hartman
% Thesis - Calculate Directivity Levels from Microphone Recordings of M&K Loudspeakers
% Originated: 7/6/02

clear
clc
close all

ftemp=linspace(20,1000,20);
f(1:19)=ftemp(2:20);
f(20:49)=linspace(1000,10000,30);
f(50:99)=linspace(10000,20000,50);

screen = get(0,'ScreenSize');

% We know from the test signal that the tone plays for 1.75 sec and has a .25 sec silent interval between
% each freq. This means that once it starts, it will be on for an an index of 1.75*fs and off for .25*fs

b=1;
p=0;
while b<4
  if b==1
        [filename, pathname] = uigetfile('*.wav','Choose "LOW" Freq. Mic
Recording',screen(4)/2,screen(3)/2);
        addpath(pathname);
    [y,fs,trash] = wavread(filename); % WAVS are read into 2 columns. col 1 is left/mono, col 2 is right.
```

```
  elseif b==2
         [filename, pathname] = uigetfile('*.wav', 'Choose "MID" Freq. Mic
Recording',screen(4)/2,screen(3)/2);
         addpath(pathname);
    [y,fs,trash] = wavread(filename);  % WAVS are read into 2 columns. col 1 is left/mono, col 2 is right.
  else
         [filename, pathname] = uigetfile('*.wav', 'Choose "HIGH" Freq. Mic
Recording',screen(4)/2,screen(3)/2);
         addpath(pathname);
    [y,fs,trash] = wavread(filename);  % WAVS are read into 2 columns. col 1 is left/mono, col 2 is right.
  end

  y=abs(y);

     n=1;
     while y(n)<.05,
       n=n+1;
       kbegin=n;
     end

     kbegin=kbegin+floor(.1*fs); % ignore initial ,1 sec of each tone recorded
     klength=floor(2*fs); % tone sequence (tone/silence) is 2 seconds long
  kmeasure=floor(1.5*fs); % level will be the average max value of 1.5 second of data

  for m=kbegin:klength:length(y)-kmeasure
    if m==0
      % Used to double check data being used
         z=y(m-10000:m+kmeasure+10000);
         zz=zeros(length(z),1);
         zz(10001:10001+kmeasure)=y(m:m+kmeasure);
         plot(z)
         hold
      plot(zz,'g');
      hold
    end

    p=p+1;

    if p<100  %there are only 99 frequency points
         temp=sort(y(m:m+kmeasure));
         temp=flipud(temp);
      avgmiclevel(p)=mean(temp(1:floor(.1*(kmeasure))));  % Average the first 10% of max values
    end
  end
  b=b+1;
end


% Plot & Display results

figure
subplot(2,1,1)
stairs(avgmiclevel)
title('Microphone Recorded Levels')
xlabel('Tone Index')
ylabel('Recorded Signal Level')
```

```
axis([0 100 0 .8])

% Plot normalized decibel scale

normmiclevel=1.42*avgmiclevel/max(avgmiclevel);  %normalize to max of +3dB (1.42)
normmicdB=20*log10(normmiclevel);

subplot(2,1,2)
semilogx(f,normmicdB)
xlabel('  Logarithmic Frequency'  )
ylabel('  Normalized Level in dB'  )
axis([80 20000 -12 3])
```

Appendix D - Impulse Response

To ensure a thorough definition of the acoustic space, the impulse response of the test room was determined. Specifically, an impulse response of the Gusman "dead" room was recorded and compared to a "free field" impulse response. This was performed using the following equipment:

- **Microphone** - B&K type 4003 with Black Diffusion Cap
- **Recording -** Portable DAT machine, TASCAM model DA-P1
- **Impulse Generator -** Standard rat trap modified with a metal striker

In the "dead" room, the microphone was setup to be at the center -head position of the test subjects at a height of 1.22m from the floor. The impulse generator was then mounted atop the speaker stands, which would be used during the experiments. The stand was located directly in front of the microphone at the same height and at a distance of 2m (see Figure 41). A string was tied to the release of the rattrap, which ran outside of the room and allowed remote triggering.

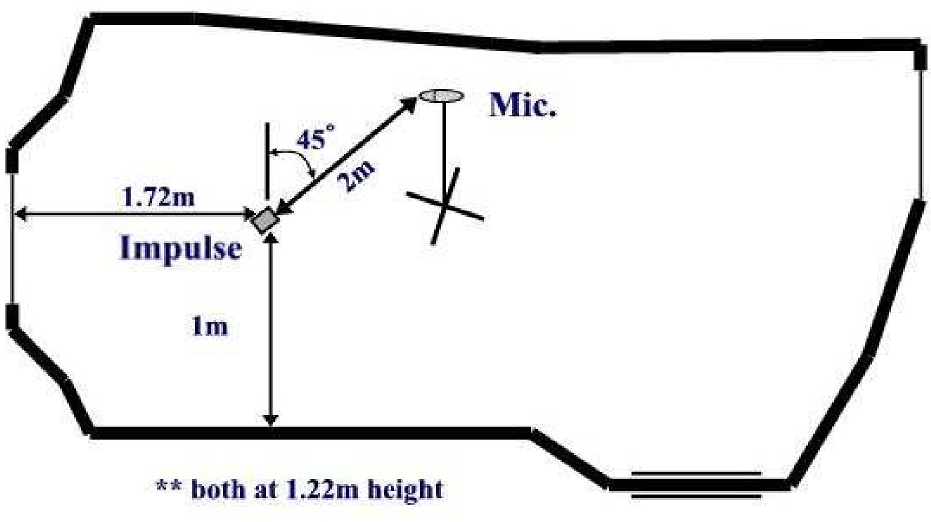


Figure 41: "Dead" room Impulse Response Setup

The "free field" impulse responses were generated by using the same setup previously described, this time located in the approximate middle of the University of Miami's Intramural (IM) fields (south of the percussion studio). The t est was performed late in the evening to minimize interference of other sounds. The approximate location can be seen in Figure 42.

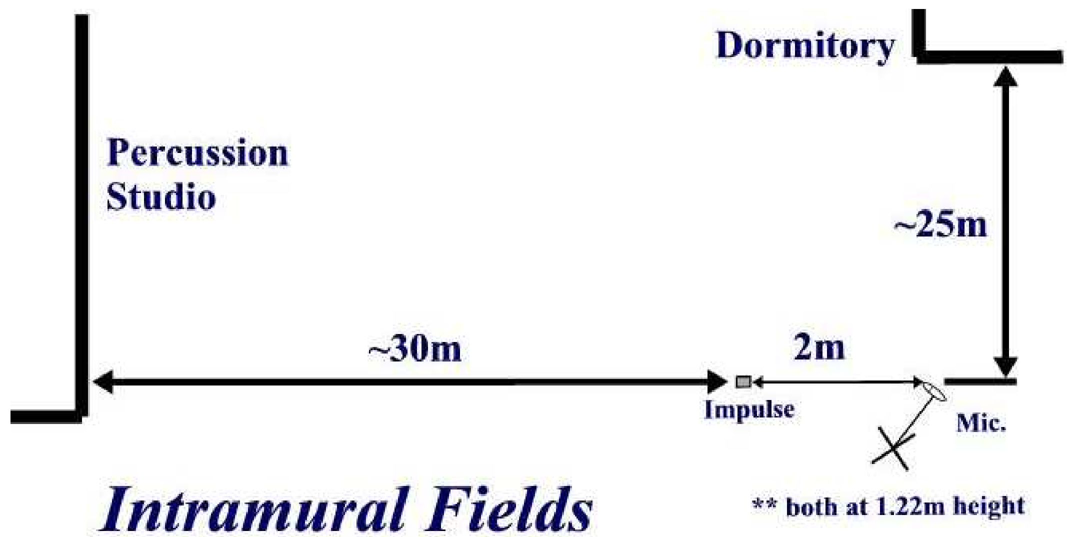


Figure 42: "Free field" Impulse Response Setup

In both cases, several test trials were run in order to maximize the recorded signal level without clipping. Five impulse responses were then recorded in each acoustic space. These recordings were transferred from the Digital Audio Tape (DAT) to a CD audio track and then converted to a mono WAV digital audio file at 44,100 Hz and 16 bits. The WAV files were truncated to 100 ms lengths using Sound Forge in preparation for the frequency analysis using Matlab (see Appendix C for code).

The mathematics and theory behind this exercise should be quickly reviewed. Particularly, recording the resulting sound of an impulsive sound source can capture the impulse response of an acoustic space. However, this recording would contain both the response of the room and the additional unwanted response of the recording equipment. It is desirable to remove the equipment's response from the overall recorded response. Therefore an additional recording in an anechoic environment can be used to obtain the equipment's response. With both frequency responses, the equipment's response can be removed from the room's recording, resulting in the desired frequency response of the room.

In these recordings, the impulse response generator is assumed to generate an ideal impulse. Additionally, the "field" recordings are assumed to represent the needed anechoic condition. Consider the discussed analysis in a mathematical context with the output (Y), input (X) and impulse response (H):

$$Y_{ROOM}(w) = X(w) \cdot H_{EQUIP}(w) \cdot H_{ROOM}(w)$$
$$Y_{FIELD}(w) = X(w) \cdot H_{EQUIP}(w) \cdot H_{FIELD}(w)$$

Assuming an ideal anechoic environment allows $H_{FIELD}(w) = 1$, and an ideal impulse response source gives $X(w) = 1$, which results in:

$$Y_{FIELD}(w) = H_{EQUIP}(w)$$

and

$$\Rightarrow H_{ROOM}(w) = \frac{Y_{ROOM}(w)}{Y_{FIELD}(w)}$$

The recorded impulse responses were input to Matlab, which was used to calculate the frequency responses (H) using the Discrete Fourier Transform (DFT). This analysis showed that one recording from each location seemed to be outside the expected DFT, and therefore was eliminated. The remaining four responses were averaged and normalized to create the plots shown in Figure 44 (see Appendix C for Matlab code). Also shown is a temporal plot of the recorded impulse response (Figure 43) and the resulting impulse response (Figure 45).

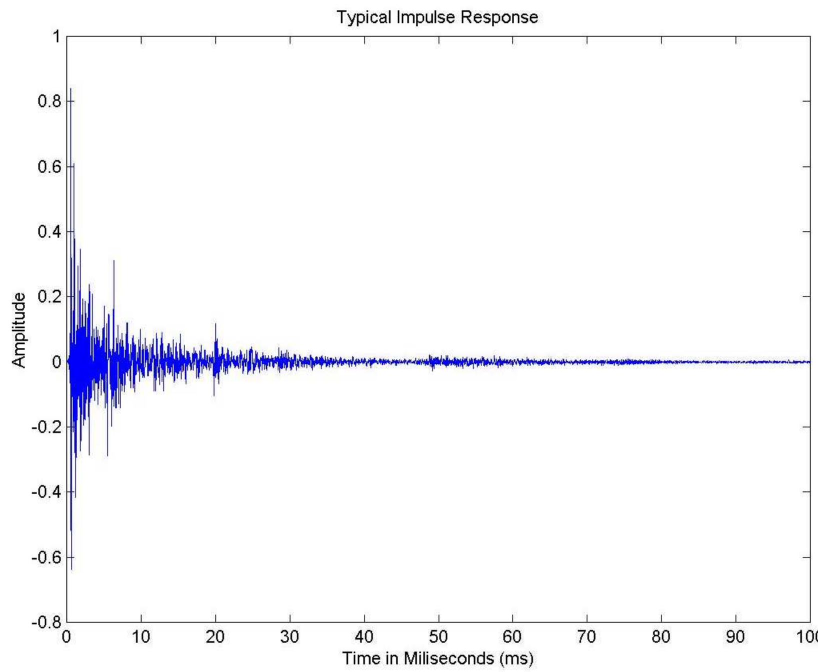


Figure 43: Temporal Plots of Impulse Responses

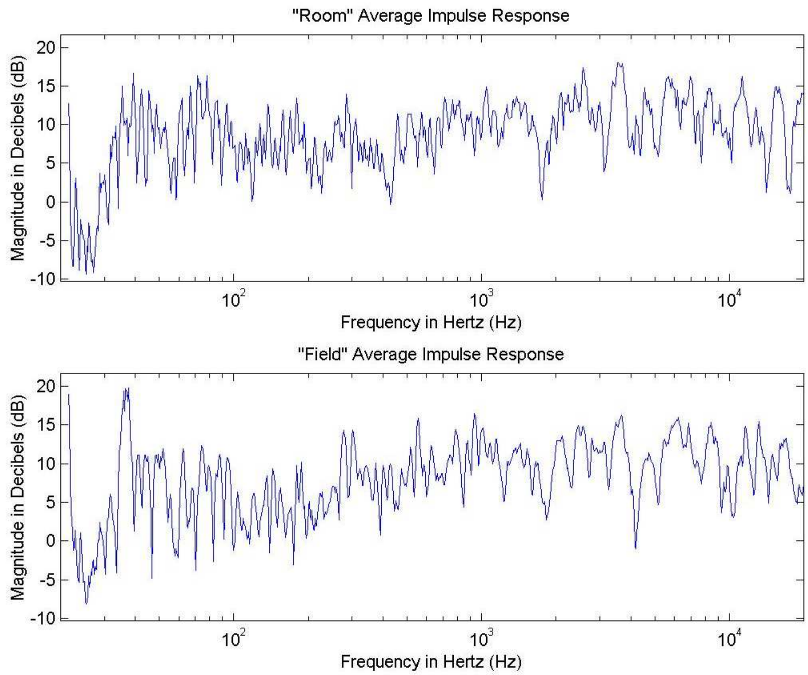


Figure 44: Spectral Plots of "Room" (top) and "Field" Impulse Response

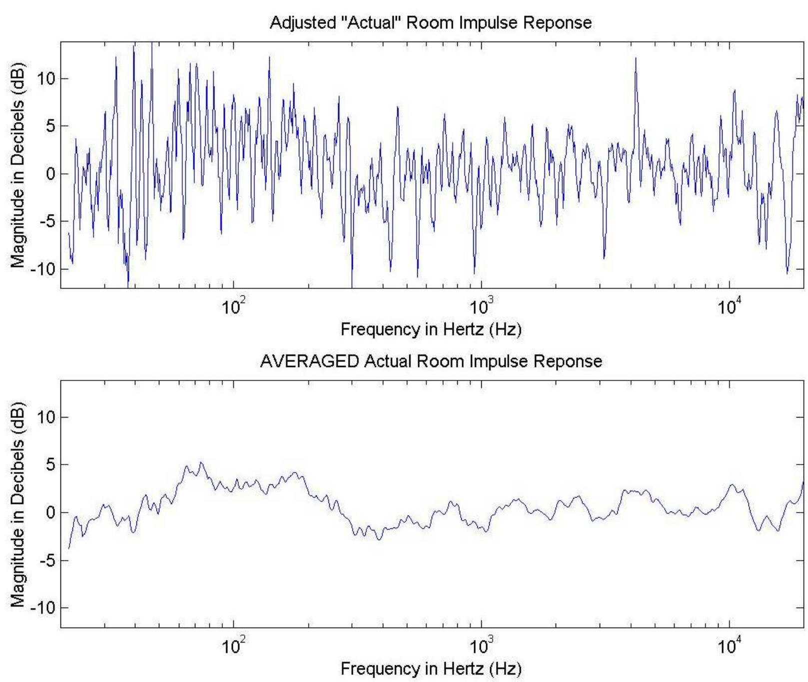


Figure 45: Resulting Impulse Responses of "Room"

Appendix E - Loudspeaker Frequency Response

In addition, it was necessary to verify the frequency response of the loudspeakers. The following equipment was used for this portion of the experiment:

- **Microphones** - B&K 4003 with black diffusion cap
- **Recording -** Portable DAT machine, TASCAM model DA-P1
- **Loudspeakers -** M&K model MPS-1610 loudspeakers previously discussed.
- **Speaker Stands -** Studio Tech SN-A adjustable metal speaker stands

A CD audio track was played; having one hundred sinusoidal signal bursts ranging from 20-20,000 Hz was played. Each frequency was held for 1.75 seconds, and there was .25 seconds of silence between successive tone increments. The recording was played through the loudspeaker under test and recorded, as shown in Figure 46. The microphones were setup at the same height as the center of the woofer and at a distance of 1 meter.

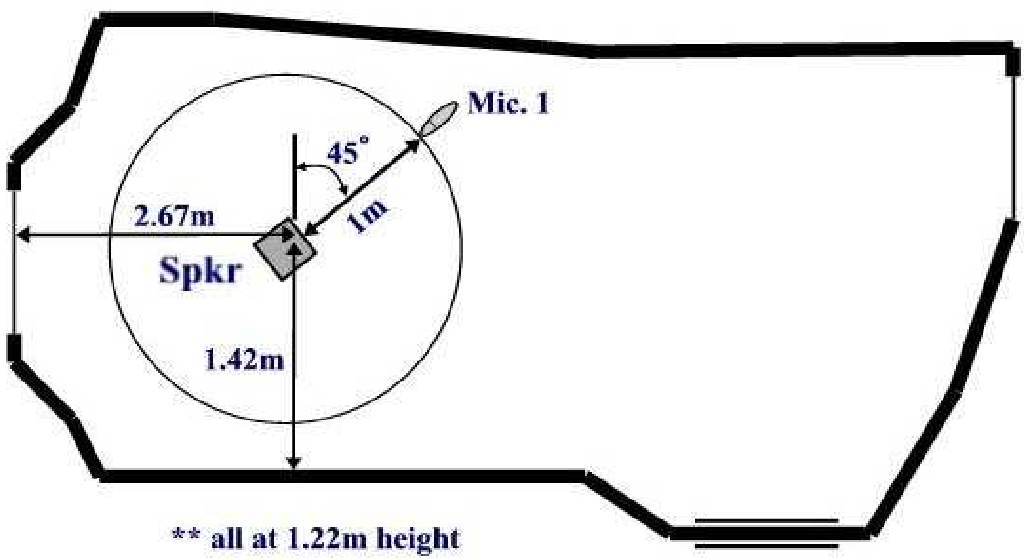


Figure 46: Frequency response measurement setup

As mentioned, each of the recorded microphone tracks was converted from DAT to CD and then to a WAV 16 bit, 44.1 kHz mono digital audio file. The WAV files were then analyzed using Matlab (see Appendix C for code). The frequency response can be shown by plotting the data of the microphone directly in front of the speaker versus frequency (see Figure 47).

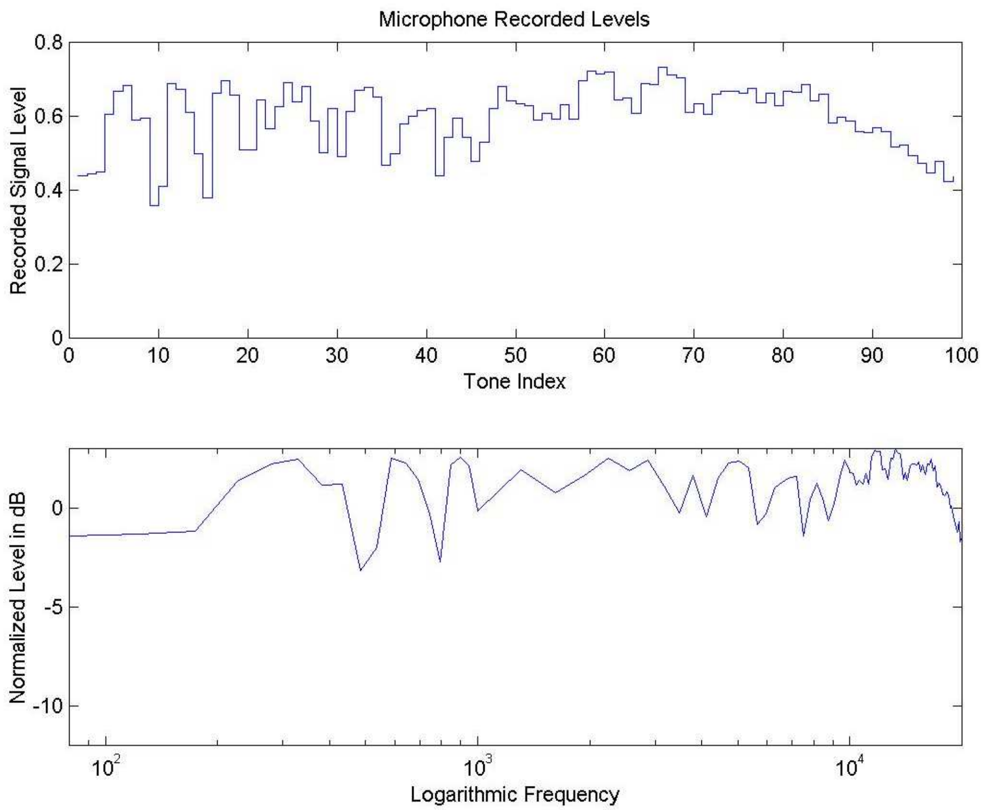


Figure 47: Frequency Response of MPS-1610 Loudspeakers

VITA

Robert Grant Hartman was born in New Brighton, Pennsylvania, on May 9, 1974. His parents are Robert Grant Hartman and Marilyn Lorraine Hartman. He received his elementary education at Ewing Park Elementary School and his secondary education at Lincoln High School in Ellwood City, Pennsylvania. In August 1992 he entered the College of Engineering at the Pennsylvania State University from which he graduated with the BS degree in Electrical Engineering in June 1997. Throughout school, Robert was involved as a cooperative education student, working for Lutron Electronics in Coopersburg, PA. After his undergraduate graduation, he accepted employment at Ford Motor Company in Dearborn, Michigan where he worked for two and one half years before deciding to return to pursue his master’s degree via an educational leave.

In January 2000 he was admitted to the Graduate School of the University of Miami where he was granted the degree of Master of Science in Music Engineering Technology in May 2003. After receiving his graduate degree, Robert was employed at Niles Audio in Miami, Florida.

Permanent Address: 407 Beaver St., Wampum, PA 16157