5-19-2014

Investigating the interaction between positions and signals of height-channel loudspeakers in reproducing immersive 3d sound

Antonios Karampourniotis

Follow this and additional works at: http://scholarworks.rit.edu/theses

Recommended Citation
Investigating the interaction between positions and signals of height-channel loudspeakers in reproducing immersive 3d sound

Antonios Karampourniotis
May 19, 2014
Rochester, New York, USA

A research Thesis Submitted in Partial Fulfillment of the Requirements for the degree of Master’s of Science in Telecommunications Engineering Technology (MSTET)
Thesis Supervisor
Dr. Sungyoung Kim
Department of Electrical, Computer and Telecommunications Engineering Technology
College of Applied Science and Technology
Rochester Institute of Technology
Rochester, New York, USA

Approved by:

____________________________________
Dr. Sungyoung Kim
Thesis Advisor, Department of Electrical, Computer and Telecommunications Engineering Technology

____________________________________
Professor Mark J. Indelicato
Committee Member, Department of Electrical, Computer and Telecommunications Engineering Technology

____________________________________
Dr. Clark Hochgraf
Committee Member, Department of Electrical, Computer and Telecommunications Engineering Technology
Dedicated to my mother…
# TABLE OF CONTENTS

Table of Figures  
Table of Appendices  
Abstract  
1. Introduction  
   1. History of Multichannel Audio  
      1. Mono  
      2. Stereo  
      3. Multichannel Reproduction Systems  
         1. 5.1 & 7.1  
         2. NHK 22.2  
         3. Auro3D  
         4. Wave Field Synthesis (WFS)  
         5. Virtual Surround  
2. Room Acoustics in Multichannel Audio  
   1. Anechoic Chamber  
   2. Reverberation  
   3. Impulse Response  
3. Psychoacoustics in Multichannel Audio  
   2. Sound Localization
3. Loudness
   1. Phons
   2. Sones
   3. Weighting Functions
   4. Sound Pressure Level

4. Timbre
5. Pitch

2. Method
   1. Configuration
   2. Software
      1. EASERA
      2. MAX/MSP
      3. MATLAB
   3. Level Adjustment
   4. Participants
   5. The Interface
   6. Stimuli
   7. Playback Equipment

3. Experimental Stages and Results
   1. Stage I
      1. Results
   2. Stage II
      1. Results
3. Stage III
   1. Results
4. Discussion
5. Future Research Topics
6. Conclusion
7. References
## TABLE OF FIGURES

<table>
<thead>
<tr>
<th>Figure Number</th>
<th>Figure Caption</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 1</td>
<td>The ITU-recommended loudspeaker configuration for multichannel audio reproduction (ITU-R BS 775-1), which is often referred to 5.1-channel audio or 5.1-surround</td>
</tr>
<tr>
<td>Figure 2</td>
<td>Illustration of the NHK 22.2 configuration</td>
</tr>
<tr>
<td>Figure 3</td>
<td>Illustration of all the current versions of Auro3D configurations</td>
</tr>
<tr>
<td>Figure 4</td>
<td>Inside the Fraunhofer Institute for Digital Media Technology (IDMT), WFS Lab</td>
</tr>
<tr>
<td>Figure 5</td>
<td>Illustration of HRTF variables</td>
</tr>
<tr>
<td>Figure 6</td>
<td>Inside an anechoic chamber while measuring the environmental noise generated by a sports car</td>
</tr>
<tr>
<td>Figure 7</td>
<td>Illustration of how the reflections can distort the auditory image</td>
</tr>
<tr>
<td>Figure 8</td>
<td>Relationship of phons and dB</td>
</tr>
<tr>
<td>Figure 9</td>
<td>Behavioral chart of A,B,C and D weighting functions</td>
</tr>
<tr>
<td>Figure 10</td>
<td>Illustration of the position of loudspeakers used in particular experiment</td>
</tr>
<tr>
<td>Figure 11</td>
<td>Microphone array used to capture the IRs for this thesis</td>
</tr>
<tr>
<td>Figure 12</td>
<td>Time vs Amplitude analysis of the Anechoic Signal</td>
</tr>
<tr>
<td>Figure 13</td>
<td>Time vs Amplitude analysis of an Impulse Response</td>
</tr>
<tr>
<td>Figure 14</td>
<td>Time vs Amplitude analysis of the Anechoic Signal</td>
</tr>
<tr>
<td>Figure 15</td>
<td>Illustration of the schematic used for Stage I and II of this thesis. S1 and S2 are the snakes used to connect the audio interfaces to the actual loudspeakers</td>
</tr>
<tr>
<td>Figure 16</td>
<td>The RME UFX audio interface used for the purposes of the experiment</td>
</tr>
<tr>
<td>Figure 17</td>
<td>Illustration of the interface for Stage I</td>
</tr>
</tbody>
</table>
### TABLE OF TABLES

<table>
<thead>
<tr>
<th>Tables</th>
<th>Caption</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table 1</td>
<td>Table of all configurations</td>
</tr>
<tr>
<td>Table 2</td>
<td>Stage I saved data format</td>
</tr>
<tr>
<td>Table 3</td>
<td>Overall summed ranking</td>
</tr>
<tr>
<td>Table 4</td>
<td>The results from Wilcoxon’s test for all configurations</td>
</tr>
<tr>
<td>Table 5</td>
<td>Friedman’s test results</td>
</tr>
<tr>
<td>Table 6</td>
<td>Example of function’s output format</td>
</tr>
<tr>
<td>Table 7</td>
<td>Spread sheet of all adjectives used by the participants</td>
</tr>
</tbody>
</table>
ABSTRACT

Since transmission capacities have significantly increased over the past few years, researchers are now able to transmit a larger amount of data, namely multichannel audio content, in the consumer applications. What has not been investigated in a systematic way yet is how to deliver the multichannel content. Specifically, researchers’ attention is focused on the quest of a standardized immersive reproduction format that incorporates height loudspeakers coupled with the new high-resolution and three-dimensional (3D) media content for a comprehensive 3D experience. To better understand and utilize the immersive audio reproduction, this research focused on the (1) interaction between the positioning of height loudspeakers and the signals fed to the loudspeakers, (2) investigation of the perceptual characteristics associated with the height ambiances, and (3) the influence of inverse filtering on perceived sound quality for the realistic 3D sound reproduction. The experiment utilized the existence of two layers of loudspeakers: horizontal layer following the ITU-R BS.775 five-channel loudspeaker configuration and height layer locating a total of twelve loudspeakers at the azimuth of ±30°, ±50°, ±70°, ±90°, ±110° and ±130° and elevation of 30°. Eight configurations were
formed, each of which selected four height-loudspeakers from twelve. In the subjective evaluation, listeners compared, ranked and described the eight randomly presented configurations of 4-channel height ambiances. The stimuli for the experiment were four nine-channel (5 channels for the horizontal and 4 for the height loudspeakers) multichannel music. Moreover, an approach of Finite Impulse Response (FIR) inverse filtering was attempted, in order to remove the particular room’s acoustic influence. Another set of trained professionals was informally asked to use descriptors to characterize the newly presented multichannel music with height ambiances rendered with inverse filtering. The experimental results indicate the significance of the positioning of the loudspeakers with respect to the signals being fed to those loudspeakers in delivering a 3D sound field. Furthermore, it has been revealed that the perceptual characteristics that listeners linked for multichannel music with height ambiances include envelopment, elevatedness and fullness. Last but not least, after applying the inverse filtering the subjective preference was not affected significantly. This allows for the author to believe that, in fact, the room’s influence with respect to the subjective evaluation is not as important as the appropriate loudspeaker-positioning for the multichannel-reproduced music with height ambiances.
1. INTRODUCTION

Although many researchers have carried out experiments and extensive studies in contemplation of determining where and how the optimum positioning of height layer loudspeakers occurs [1, 2], perceptual characteristics of height channels has not yet received the proper attention. This thesis presents an experimental study that investigated the perceptual characteristics of height loudspeaker configurations throughout various alterations.

Surely, sound recording and reproduction techniques have gone a long way since the first appearance of the phonograph [3], but now more than ever progress comes with leaps and bounces. In the footsteps of Gernemann [4], but with a decade of technological advances apart, when height layer of loudspeakers is being mentioned, it refers to the existence of a second layer of loudspeakers above the listener’s head located 100 inches (2.54m) from the surface of the floor, in this specific research. Starting from mono to stereo and surround, comes in early 21st century the idea of height channel incorporation that allows the 22.2 multichannel [5] and Audio3D [6] reproduction, just to name a few. The incorporation of such a layer allows for the creation of an immersive sound field, crucial to the desired holistic 3D experience.
With this short background, a fundamental research question has been formed:

Is there an interaction between height channel loudspeaker configuration and height channel signal influencing the listeners’ perceived quality of 3D-Audio content?

The following sections of this thesis contain the methods, experimental results, discussions, future research topics and conclusions. But first, some background information about the history of the technological advances throughout the years in the audio reproduction research and development, as well as some information on several key concepts of this research.

1.1 HISTORY OF MULTICHANNEL AUDIO

1.1.1. MONO

A monophonic sound reproduction, or monaural, refers to the existence of one microphone to capture the sound and/or one loudspeaker to reproduce the signal. In the scenario of multiple microphones, the signals are being mixed down to a single track in order to be fed to the loudspeaker. Although stereo has replaced
mono reproduction in many of today’s entertainment applications, monaural sound is still being used in telephony, hearing aids, as well as some FM radio stations that still use mono to broadcast, mainly talk shows. The reason behind it is that monophonic signals have better signal strength while being broadcasted and experience less information and quality loss.

1.1.2. STEREO

A stereophonic sound is described as a method of sound reproduction that gives the listeners directional illusion and creates the image of an audible perspective. With respect to the impression of the sound as being perceived from various directions just as in natural hearing, the stereophonic sound can be achieved by employing at least two audio channels through a configuration of at least two loudspeakers. A common misconception is that stereo reproduction only refers to the playback of sound only from two sources at once. This is the most common configuration, but stereophony could also refer to quadrophonic or even surround sound systems.

Although Clément Ader introduced the idea of two channel audio sound in Paris Opera in 1881, it wasn’t until 1931 when an electronics engineer named Alan Blumlein came to make waves and changed the way we perceive audio for ever.
Since 1931 a number of changes and additions have been made to the stereo format to keep up with the technological progress and consumer preferences. Stereo reproduction systems have managed to stay atop and be the default format for the music and movie industry for almost a century. Stereo sound remains today’s standard, but may not as bandwidth and technology can support the delivery of new multichannel formats and the creation of an immersive three dimensional sound field.

Stereophony also introduced a new term, the sweet spot. The term is being used in the audiophile community in addition to the recording and audio engineering community to describe the point between an undefined number of loudspeakers, in which an individual can completely and clearly perceive all the effects that are added to a sound source and experience the full extent of that source.

1.1.3. MULTICHANNEL

Multichannel reproduction systems are an attempt to augment the sound reproduction quality of an audio source with supplementary audio channels from additional loudspeakers that surround the listener, producing sound from a 360° radius in the horizontal two dimensional plane.
Multichannel reproduction adds to the perception of sound spatialization by handling sound localization; a listener's capability to pick out the location or origin of a perceived sound in direction and distance. Normally this is accomplished by using multiple individual audio channels fed to an assortment of loudspeakers.

1.1.3.1. SURROUND SYSTEMS (5.1 & 7.1)

The concept of a 5.1 surround system dates back to 1976 when in Dolby Labs the engineers modified the then traditional use of the six analogue magnetic sound-track tracks to the format that we know today. On the front of the sweet spot the listener can find a center loudspeaker located in 0°, a front-left and front-right loudspeaker located ±30° from the center loudspeaker, two loudspeakers in the rear located ±110° from the sweet spot and last but not least a Low Frequency Effect (LFE) to generate the low frequencies that the loudspeakers can not reproduce. Scientists however have standardized a version of the 5.1 surround system that they called a Full Range 5.1 surround system, where in this particular system all the loudspeakers can reproduce a wide bandwidth of the audible range of frequencies rendering the LFE less needed. All the loudspeakers are positioned equidistant from the sweet spot. A 7.1 surround system consists of a center loudspeaker, two loudspeakers positioned ±30°, two side loudspeakers at ±70° and
finally two rear loudspeakers located $\pm 130^\circ$. The 7.1 surround system can deliver a wider surround image due to larger number of loudspeakers and its additional rear loudspeakers.

Figure 1. The ITU-recommended loudspeaker configuration for multichannel audio reproduction (ITU-R BS 775-1), which is often referred to 5.1-channel audio or 5.1-surround [7]

1.3.2. NHK 22.2-CHANNEL SYSTEM

Surround sound does not refer to the existence of loudspeakers in just the horizontal plane. Surround can also include height layer of loudspeakers, such as the 22.2 configuration. The need that drove the creation of the NHK (translates to Japan Broadcasting Corporation) 22.2 configuration was the development of the ITU approved Ultra High Definition TeleVision (UHDTV) which runs in two digital
video formats, 4K and 8K. 4K (2160p) and 8K (4320p) refer to the definition of the screen, delivering unparalleled clarity and depth of color. As a comparison, think of the High Definition (HD) screens, as they have a resolution of 720p or 1080p. The lowest UHDTV format has two times the definition of the default HDTV. NHK Science & Technical Research Laboratories released both the UHDTV and the 22.2 configuration. The particular configuration can be perceived as the surround sound component of UHDTV, wherein it utilizes a total of 24 loudspeakers arranged in three layers; an upper layer of nine channels, a middle layer of ten channels, and a lower layer of three channels and two channels for LFE. Hamasaki et al. [8] mentioned that “the 22.2 multichannel audio system can reproduce a greater sensation of presence over a wider listening area than the conventional multichannel audio system, and the upper layer of loudspeakers is essential to reproducing better presence”.

Figure 2. Illustration of the NHK 22.2 configuration
1.1.3.3. AURO3D

In the never-ending search of an audio format that can deliver a much wider range of spatial sound effects and allow more realism of spatial reproduction in terms of direct sound, early and late reflections, reverberation and ambience sound, Wilfried and Guy Van Baelen introduced Auro3D to the public. Currently there are three versions of Auro3D, the Auro3D 9.1, 10.1 and 11.1. The Auro3D 9.1 version features the use of the typical 5.1 surround system and in addition the use of a given upper (height) layer of loudspeakers right above the left (−30°), right (+30°), rear left (−110°) and rear right (+110°) height loudspeakers. The 10.1 version of the audio format utilizes the addition of another loudspeaker on top of the 9.1 version but in the same upper layer, located right above the center loudspeaker (0°). That loudspeaker is called Height Center. Essentially, Auro3D 10.1 is a duplication of the ITU-R 5.1 surround system in the height layer. Lastly the Auro3D 11.1 employs the use of yet another loudspeaker located on the ceiling facing down vertically to the ground labeled as the Top Speaker, introducing a second height layer. This particular format has a significant impact on several perceptual characteristics including spatial depth, spatial impression, envelopment, ambient atmosphere, as well as directional stability within the sweet spot.
1.1.3.4. WAVE FIELD SYNTHESIS (WFS)

Wave field synthesis (WFS) is a method of recreating an exact acoustic replication of a sound field employing only the theory of waves and wavefronts. A wavefront is described in physics as the locus of points that have the same phase. The initial idea was born more than 20 years ago in Delft University of Technology [10]. The basis of WFS is the Huygens' Principle which manifests that the points on a wave front serve as individual point sources of spherical secondary waves.
To portray Huygens' principle and make the concept clearer, let us consider a simple example. A rock (or primary source) thrown in the middle of a pond generates a wave front that propagates along the surface. Huygens' principle indicates that an identical wave front can be generated by simultaneously dropping an infinite number of rocks (secondary sources) along any position defined by the passage of the primary wave front. In order to perfectly synthesize the initial wave, the knowledge of the passage is required.

Figure 4. Inside the Fraunhofer Institute for Digital Media Technology (IDMT), WFS Lab

This synthesized wavefront will be perfectly accurate outside of the zone determined by the secondary source distribution. The wavefronts (secondary sources)
have the ability to replicate the original wavefront in absence of a primary source [13].

In the audio engineering world, in order to apply Huygens' Principle, a large number of loudspeakers placed close to each other, forming a loudspeaker array, is needed. That array, if properly programmed, can deliver an acoustic hologram which can accommodate perfect temporal, spectral and spatial properties throughout the listening room. Although extensive research is being conducted related to WFS currently, there are still some remaining issues. These are:

I. **Sensitivity to room acoustics**: In order for WFS to simulate the acoustic attributes of a recording space, the acoustics of the rendition area must be suppressed. A way to address the issue is to arrange the walls in an absorbing and non-reflective way, hence non-parallel walls. The second possibility of prevention is playback within the near field. There are two ways for this to work productively. First the loudspeakers must be coupled very closely at the listening zone, or second, the diaphragm surface must be very large.

II. **High cost**: Due to the large population of transducers, in contemplation of reducing spatial aliasing to a minimum, the cost of such systems are extremely high.
III. Aliasing: Aliasing refers to the existence of artifacts between continuous and discrete signals. Aliasing arises when a signal is discretely sampled at a rate that is insufficient to capture the changes in the signal. And safe way to avoid aliasing is always making sure that adequate samples have been captured to demonstrate various changes of the signal, in the time and the frequency domain.

IV. Truncation effect: Since the final spherical wavefront is synthesized by elementary waves, an unexpected alteration of pressure could possibly occur if no more speakers deliver elementary waves, where the loudspeaker array ends.

1.1.3.5. VIRTUAL SURROUND

Virtual Surround is an audio technology that aims to create the illusion that there are more sound sources than actually present. That can be accomplished by driving the human auditory system to believe that a perceived sound is coming from another direction from where it actually is. Such systems can create multiple virtual loudspeakers and provide the listener with the feeling that they are listening to a multichannel reproduction system. That illusion can be accomplished even with the presence of a single loudspeaker. In order to achieve this, several psychoacoustics principles (such as Head-Related Transfer Function (HRTF)) have to
be put into practice. In addition, the reflections of the walls, ceiling and floor can affect the reproduction of sound.

HRTF is a response that makes distinctive how a human ear becomes aware of a sound from a certain location. A couple of HRTFs can create a binaural sound radiating from a particular point in space, since humans have two ears. HRTFs are transfer functions, representing the way a sound from a discrete point will arrive to the outer rim of the auditory canal and into the eardrum. As illustrated in Figure 4, a sound source reaches to the two ear position resulting in two distinct transfer functions—\( h_L(t) \) and \( h_R(t) \). A listener can perceive the 3-dimensional sound source position through these two transfer functions.

![Figure 5. Illustration of HRTF variables](image)

The way humans perceive sound is unique for each individual, since some of the factors that determine the uniqueness is the size and shape of the head, pinna and
torso. External factors such as the environment in which sounds are being generated and heard or elements such as diffraction and reflection can cause modifications to the perceived sounds.

1.2. ROOM ACOUSTICS IN MULTICHANNEL AUDIO

The term room acoustics describes how sound behaves in an enclosed space. Any multichannel audio system is influenced by the enclosure wherein people listen the reproduced music/sound. Room acoustics can also alter the perceptual characteristics of the original wave if not addressed properly. In this section, the various room acoustics will be reviewed to systematically understand the influence on the waves and how it can be addressed.

1.2.1. ANECHOIC CHAMBER

An anechoic chamber is a room designed to utterly absorb reflections of either sound or electromagnetic waves. It is specially designed to also be soundproof and thoroughly isolated from exterior noise sources thus it can be used as a recording environment with no ambient noise and reflected sound. The walls, ceiling and floor of the anechoic chamber are lined with a sound absorbent material
(glass-fibre wedges). The reason is the attempt of simulation of a quiet open space of infinite dimensions in order to conduct acoustical and psychoacoustical experiments. Anechoic chambers come in many different sizes and types. The sizes can range from a small household item to aircraft hangars. Some available types, currently, include anechoic chambers that block radio frequencies for radar testing, acoustic anechoic chambers and semi-anechoic chambers. The size and the type of the anechoic chamber primarily depend on the size and the type of the object being tested.

![Image of anechoic chamber with a sports car](image)

**Figure 6.** Inside an anechoic chamber while measuring the environmental noise generated by a sports car [14]

**1.2.2. REVERBERATION**
Reverberation is the prolongation of sound, or as known as, the resonance. It is the persistence of sound even though the generating source has stopped producing sound. Unlike echo, reverberation can be identified by the continuous stream of sound. Reverberation, and echo, are caused by reflections.

Reflections, much like multipath propagation (or multipath fading) in the telecommunications field, if not properly addressed can cause modification to the original signal. The reason behind the modification is the distortion of the primary signal when convolved with the out-of-phase reflected secondary signals. Early and late reflections determine the auditory depth of the human perception. Sound waves, as any other mechanical wave, reflect off of objects the same way billiard balls bounce off the bumpers of a pool table—the angle of incidence equals the angle of reflection. A sound wave hitting a flat wall at 45° will reflect off it at 45°. The reflected wave can interfere with the original wave, producing constructive and destructive interference. The interference can increase the convolved signal’s amplitude or, with phase cancellation, decrease its amplitude.

In many cases architects collaborate with audio engineers to physically minimize the acoustic influence of a hall or venue. The minimization the effect of the room’s acoustics, by defusing reflections, will allow easier manipulation of reflec-
tions (perhaps virtually generated), allowing modification of the room reverbera-

tion acoustics.

Figure 7. Illustration of how the reflections can distort the auditory image [15]

1.2.3. IMPULSE RESPONSE

The signal processing disciplinary defines the impulse response (IR) of a dynamic system as the output of that system when presented with a finite (brief) input signal - an impulse. In general, an IR refers to the reaction of any dynamic system in response to some external change. The IR describes the response of the system as a function of time, or perhaps as a function of any other predefined independent variable that configures the dynamic behavior of the system. In all these cases, the dynamic system and its IRs could be a number of physical ob-
IRs have varied applications. Their use is seen in loudspeaker design, electronic processing, control systems, acoustic/audio applications and economics. In acoustic and audio applications, IRs represent the acoustical attributes (acoustic fingerprint) of a specific position in a space. The IRs can be measured using maximum-length sequence (MLS) or time-stretched pulse (TSP) method [16]. Convoluting a dry sound source with the measured IR can simulate the auditory impression of the position where the IR was captured.

1.3. PSYCHOACOUSTICS IN MULTICHANNEL AUDIO

Psychoacoustics is a branch of psychophysics and the scientific study of sound perception. This includes the way humans perceive sound, psychological responses and the physiological impact of music and/or sound on the human nervous system. In the vast field of psychophysics and more specifically of psychoacoustics, the terms music, sound, frequency, and vibration are fundamentally indistinguishable, since they are different concepts describing one and the same essence [17]. Primarily, the perception of sound is in fact a behavioral response to the physical
attributes of the sound that includes intensity, frequency, and characteristics in the
time domain that allows for the auditory system to determine direction, distance,
loudness, timbre, pitch, and more. However, psychoacoustics is not only being
used for audio purposes.

Today psychoacoustics is being applied in many fields, such as software devel-
opcodevelopment where developers map out proven and experimental mathematical pat-
terns of the human auditory system. Another prominent field is in defense systems
design where scientists and engineers are trying to create an acoustic weapon than
can impair, harm and potentially kill others [18].

1.3.1. SOUND LOCALIZATION

Binaural hearing is directly correlated to the fact that ears are separated by some
distance (approximately 17cm [19]). This allows for the localization of sound by
deciphering the differences in arrival time, phase and intensity. According to
Helmut Haas, the human brain can actually detect time differences as low as 30
milliseconds [20]. Since this is a three-dimensional world, localization can be re-
lated with terms of three-dimensional positioning: the azimuth or horizontal an-
gle, the elevation or vertical angle, and the distance (for sounds that remain static
with respect to time) or velocity (for sounds that change in space and time) [21, 22].

1.3.2. LOUDNESS

Loudness can be described as "that attribute of auditory sensation in terms of which sounds can be ordered on a scale extending from quiet to loud" [23]. Loudness is highly correlated with subjective measurement rather than objective measurements of sound intensity, such as Sound Pressure Level (SPL) or sound power. Loudness can be affected by many parameters such as frequency, duration and critical bands. The frequency bandwidth of the auditory filter created by the cochlea is referred to as critical band. The concept of critical band was introduced by Fletcher in 1940 and since then has been extensively tested. Studies [24] have shown that critical bands are narrower in low frequencies rather than in high frequencies. Out of all the critical bands, three-fourths are created under 5kHz. The frequency bandwidth of each critical band is particularly dependent on the fundamental frequency of the sound arriving in the ear canal. If the fundamental frequency changes, new critical bands will be dynamically created, centered on the new fundamental frequency.
1.3.2.1 PHONS

A phon is a unit of measuring loudness for pure tones. Pure tones are tones with a sinusoidal waveform (sine or cosine), which means that regardless of other attributes of the signal (phase, amplitude), the wave is composed of one single frequency. Two sounds of equal intensity do not correspond to equal loudness. The auditory sensitivity for humans varies according to frequency. If a pure tone of 1kHz is chosen as a reference, then the number of phons of a sound is the deciBel Sound Pressure Level (dB SPL) of a pure tone of 1 kHz which is received to sound just as loud.

In other words, if a sound is perceived to be 60 phons of loudness this means that it can be translated as a pure tone of 1000Hz perceived in 60dB [24]. In extension this can imply that 0 phons is the threshold of perception and that some sounds can have negative phon levels [Figure 8].

1.3.2.2 SONES

A sone is used to characterize the subjectively perceived loudness of a sound. It has been calculated that a sound of 1 sone corresponds to a sound of 40 phons, the loudness level of a 1 kHz pure tone at 40 dB SPL [25]. However, a pivotal differ-
ence between phons and sones is that the phon scale measured level is expressed in dB and not loudness. The result of that difference is that the sone and phon scales are not proportional. The correlation of these two units translates as a 10 phon increase (+10 dB of a pure tone of 1kHz) results in the doubling of loudness expressed in sones.

![Figure 8. Relationship of Phons and dB](image)

### 1.3.2.3 WEIGHTING FUNCTIONS

There are four types of weighting functions when measuring loudness. The A-weighting function (weighting filter) is often used to highlight frequencies around the area of 2-6 KHz where the human ear is most responsive, while at the same time attenuating lower and higher bands of frequencies where the human ear is less sensitive. A-weighting is used to describe quieter sounds due to its initial de-
sign with the principle basis on the 40-phon Fletcher–Munson equal-loudness contour (one of many set of equal-loudness contours).

An equal-loudness contour is a measure of dB SPL in the spectral domain, in which a listener perceives a constant loudness when presented with pure steady tones. The B-weighting and C-weighting curves were deliberately designed for louder sounds while the D-weighting function is used in evaluating loud aircraft noises (that would explain the boost from 1KHz to 10KHz in Figure 9).

The explained weighting functions can be applied in telecommunications, audio reproductions and broadcasting equipment, as well as in measuring loudness, environmental noise, radiation and sunlight.

![Figure 9. Behavioral chart of A,B,C and D weighting functions [26]]
1.3.2.4. SOUND PRESSURE LEVEL

Sound pressure level (SPL) is the result of the logarithm of the effective sound pressure divided by the reference value, where $\rho_{\text{ref}} = 20\mu\text{Pa}$. ($p_{\text{rms}}$ is measured in dB).

$$L_p = 10 \log_{10} \left( \frac{p_{\text{rms}}^2}{p_{\text{ref}}^2} \right) = 20 \log_{10} \left( \frac{p_{\text{rms}}}{p_{\text{ref}}} \right) \text{ dB},$$

1.3.3. TIMBRE

Timbre, or tone color, is the character of a sound allowing us to distinguish between various instruments that reproduce the same frequency, pitch and loudness. Timbre is the concept that allows us to differentiate a piano from a guitar or a trombone from a french horn. McAdams and Bregman in 1979 once said that timbre can be identified as “the psychoacoustician’s multidimensional waste-basket category for everything that cannot be labeled pitch or loudness” [27].

1.3.4. PITCH
In the field of psychoacoustics, pitch is contemplated as the psychological perception of frequency. Although pitch and frequency are associated, they are not equivalent. The most important difference to bear in mind between those two terms is that frequency is an objective concept while pitch is subjective. Pitch is one of the principal auditory characteristics of musical tones, along with duration, loudness and timbre [28].

2. METHOD

In order to test the previous research hypothesis, an experiment has been designed and conducted. The experiment consisted of three Stages, where each Stage will be described and analyzed in the following sections. Stage I investigated the optimum positioning of height loudspeakers with respect to the appropriate 3D music reproduction, while Stage II indicates all the attributes that synthesize an immersive audio field from the subjects’ point of view. Last but not least, in Stage III of the experiment, an inverse Finite Impulse Response (FIR) filter had been implemented in order to identify the affect of the room acoustics to the listeners’ preferences.

2.1. THE CONFIGURATION
The experiment utilized the existence of two layers of loudspeakers. The horizontal layer, following the standard five-channel loudspeaker configuration (as illustrated in Fig. 1) and the height layer with elevation of 30°. A total of twelve loudspeakers were located at ±30°, ±50°, ±70°, ±90°, ±110° and ±130° in with respect to center loudspeaker of the horizontal layer [Fig. 10, Table 1]. The horizontal layer loudspeakers were located at listener’s ear height (~1.2m) as per the ITU-R standard, while the height layer loudspeakers were located 100 inches from the ground (~2.5m).

Figure 10. Illustration of total loudspeaker placement
The initial (reference) configuration was deemed to be deficient, thus more height
loudspeaker configurations needed to be taken into account. A large number of
loudspeakers were used so that listeners could compare multiple reproduction
conditions instantaneously. Among 12 loudspeakers, a subject listened to height
ambiences from four loudspeakers each time, the combination of which depended
upon the configuration the subject had selected.

<table>
<thead>
<tr>
<th>Configuration</th>
<th>Height Loudspeaker Position</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>±30°</td>
</tr>
<tr>
<td>2</td>
<td>±30° ±110°</td>
</tr>
<tr>
<td>3</td>
<td>±30° ±130°</td>
</tr>
<tr>
<td>4</td>
<td>±50° ±90°</td>
</tr>
<tr>
<td>5</td>
<td>±50° ±110°</td>
</tr>
<tr>
<td>6</td>
<td>±50° ±130°</td>
</tr>
<tr>
<td>7</td>
<td>±70° ±110°</td>
</tr>
<tr>
<td>8</td>
<td>±70° ±130°</td>
</tr>
</tbody>
</table>

Table 1. Table of all configurations

2.2. SOFTWARE USED IN EXPERIMENT

2.2.1. EASERA

EASERA is software for electronic and acoustic analysis. It provides both data
acquisition with a variety of stimulus signals including TDS, sweeps MLS or
noise excitation signals and a post processing engine to calculate acoustical parameters according to ISO Standard 3382 and higher. The real time analyzer provides multiple ways to perform a fast onsite analysis or to obtain a precise view of the surrounding acoustic environment. In this thesis, EASERA was primarily used to capture the transfer function between each loudspeaker and the listening position, and to calibrate the reproduced sound pressure level [29].

2.2.2. MAX

MAX, developed by Cycling ’74, is a visual programming language for music and multimedia. Its main market is among researchers and artists/performers. A noteworthy attribute of MAX is the freedom for third-party developers, not affiliated with the developing company, to create external customizable objects. Max has been characterized as the “lingua franca” (bridge language) for developing interactive music performance software [30]. In this work, MAX was used to reproduce the sound sources, to randomize the configuration, and to collect the listeners’ responses via a customized graphic user interface.

2.2.3. MATLAB
MATLAB is a high-performance programming language for scientific and technical computing. It combines computation, visualization, and programming in an easy-to-use environment where problems and solutions are expressed in a familiar mathematical notation [31]. Typical uses include:

- Math and computation
- Algorithm development
- Modeling, simulation, and prototyping
- Data analysis, exploration, and visualization
- Scientific and engineering graphics
- Application development, including Graphical User Interface building

An additional package, SIMULINK, adds graphical multi-domain simulation and Model-Based Design for dynamic and embedded systems. Although MATLAB is intended primarily for numerical computing, other toolboxes are specifically engineered for use in various fields [32]. Some fields include:

- Communications Systems
- Computational Biology
- Computational Finance
- Control Systems
- Digital Signal Processing
MATLAB can bring together users from different backgrounds and disciplines of engineering, science, and economics. It is extensively used in academic and research institutions as well as industrial enterprises, since it has become a standard [33]. In this thesis, MATLAB was used to analyze the collected data, run statistical analysis (Friedman’s Test, Wilcoxon rank sum) and design inverse filters for Stage III of the experiment.

2.3. LEVEL ADJUSTMENT

In all three Stages, the reproduced level of loudspeakers have been level matched and calibrated using pink noise fed individually to each loudspeaker. The calibrating device (SPL meter) was placed to the listener’s fixed position, making sure that the average playback level was set on 79 dB in a C weighting function (or 79 dBC).
2.4. PARTICIPANTS

Twelve subjects participated in the experiments. Their ages ranged between 19-25 years. All of the participants were students, from various disciplines, at Rochester Institute of Technology and had taken a technical ear training course. Although previous studies by Toole [34] have shown that fully trained professional listeners are more exigent and deliver better results, the authors called those particular participants in order to represent average consumers. All subjects were compensated for their time despite the fact that their participation was completely optional and voluntary.

2.5. THE INTERFACE

The participating subjects were presented with a custom-made Graphical User Interface (GUI) created in MAX. Using the interface the subjects could rank the eight randomly presented configurations of height loudspeakers from best to worst, based on overall perceived appropriateness. In other words, the listeners were asked to evaluate the level of appropriateness when given height channels were integrated with the horizontal surround sound field in producing a 3D immersive sound field and delivering high quality multichannel music with height ambiences. The interface allowed for the participants to compare all the different
configurations of four loudspeakers in a random sequence. After the subjects had gone through the eight combinations they were asked to perceptually rank them from best to worst. Although the system as a whole has been calibrated to emit on a max of 79 dBC, the listeners had the option of adjusting the level to a lower volume, if desired. Each listener repeated the ranking process four times, once for each stimulus. More information will be given in the sections below.

2.6. STIMULI

For the purposes of this experiment the IRs from two venues (having RT60500Hz of 1.4 sec and 2.51 sec respectively) were captured at 31bit/96KHz resolution using an 8-channel surround microphone array. The height information was captured [35] via two figure-of-eight microphones with azimuths ±90° and elevation of +45° while the microphone’s positive lobes were directed upward and were positioned at an overall distance of 1 meter from each other. The microphones (denoted as HL and HR in Fig. 11) dedicated in capturing the height channel information, height-left and height-right, were centered above a 2m x 2m array of a cluster of four omnidirectional microphones, front-left, front-right, rear-left, and rear-right. Additionally, two figure-of-eight microphones were pointing to the front and to the rear wall. A variety of IRs were captured in several positions and in an
assortment of heights (2, 3 and 4m) in those two venues. A professional recording engineer, after evaluating all the rendered IRs, selected a set of representative IRs from each hall that could deliver to the listeners the most spectrally balanced, unvaried and coherent room impression. Two anechoic recordings were selected, a male choir and a solo clarinet. Those specific musical pieces were selected from an assortment of anechoic recordings, due to their easy listening attributes. The anechoic recordings were convolved with the selected IRs, resulting in four nine-channel stimuli.

Figure 11. Microphone array used to capture the IRs for this thesis [35]
Figure 12. Time vs Amplitude analysis of the Anechoic Signal

Figure 13. Time vs Amplitude analysis of an Impulse Response
Figure 14. Time vs Amplitude analysis of the Convolved Signal
2.7. PLAYBACK EQUIPMENT

The signals were reproduced from the hard drive of the laptop through two audio interfaces—an RME UFX (Figure 14) and RME 400. A total of 17 matched loudspeakers (Genelec 8020B) were used to generate the incoming signals (5 for the horizontal and 12 for the height layer). For the signal flow and routing, readers are advised to refer to Fig. 13 below.

Figure 15. Illustration of the schematic used for Stage I and II of this thesis. S1 and S2 are the snakes used to connect the audio interfaces to the actual loudspeakers.
3. EXPERIMENT DESIGN AND RESULTS

3.1 STAGE I

Stage I investigated the influence of the height loudspeaker positions and their signals on perceived appropriateness of 3D-reproduced music. Both layers of loudspeakers were placed on the circumference of a circle with a diameter of two hundred inches (~5m). Four height channels were reproduced through eight different configurations where each configuration utilized four height loudspeakers [36]. In the continuous search for data in support of technological advancement in this area, an import question is posed:
What would be the perceptual characteristics that influence the listeners’ rank data?

3.1.1 STAGE I RESULTS

The twelve listeners that participated in the experiment were asked to compare the randomly presented eight configurations via the use of the GUI developed in MAX/MSP, and rank these configurations based on the appropriateness of the integration of the height layer [Figure 15].

Figure 17. Illustration of interface for Stage I
Unfortunately, a few students seemed unable to comprehend the task of the experiment, hence they were asked to rank the configurations based on their individually perceived sound quality. The particular method of ranking had been chosen in order to force the participants to actively differentiate the eight configurations, since no equal ranking between different configurations was a valid option in this case. Table 2 demonstrates the format in which the data has been recorded:

<table>
<thead>
<tr>
<th></th>
<th>0</th>
<th>1</th>
<th>6</th>
<th>2</th>
<th>2</th>
<th>3</th>
<th>5</th>
<th>4</th>
<th>1</th>
<th>5</th>
<th>7</th>
<th>6</th>
<th>4</th>
<th>7</th>
<th>0</th>
<th>8</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>1</td>
<td>4</td>
<td>2</td>
<td>0</td>
<td>3</td>
<td>6</td>
<td>4</td>
<td>7</td>
<td>5</td>
<td>5</td>
<td>6</td>
<td>1</td>
<td>7</td>
<td>2</td>
<td>8</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>7</td>
<td>2</td>
<td>0</td>
<td>3</td>
<td>2</td>
<td>4</td>
<td>4</td>
<td>5</td>
<td>3</td>
<td>6</td>
<td>5</td>
<td>7</td>
<td>6</td>
<td>8</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>6</td>
<td>2</td>
<td>0</td>
<td>3</td>
<td>1</td>
<td>4</td>
<td>5</td>
<td>5</td>
<td>3</td>
<td>6</td>
<td>7</td>
<td>7</td>
<td>4</td>
<td>8</td>
<td>2</td>
<td></td>
</tr>
</tbody>
</table>

Table 2. Stage I saved data format

The first number shown in every row indicates which signal, out of all four, the row refers to. In the second set of numbers, the even columns refer to the order of the height loudspeaker configuration, while the odd columns indicate the ranking that this particular listener has assigned to the specific configuration.

The author gathered a total of 48 rankings (4 stimuli and 12 listeners). Table 3 shows the sum of the collected rankings of the eight configurations, which indicates that configuration 4 had the highest rank sum.
Subsequently a Wilcoxon’s rank sum test was performed [Table 4]. The Wilcoxon rank sum test is a nonparametric test for two populations when the data are independent. If X and Y are independent samples with different sample sizes, the test statistic which rank sum returns is the rank sum of the first sample [37]. In order for the configurations to be labeled as statistically the same, the rank sum result should be greater to 0.5 (p>0.5).

<table>
<thead>
<tr>
<th></th>
<th>C1</th>
<th>C2</th>
<th>C3</th>
<th>C4</th>
<th>C5</th>
<th>C6</th>
<th>C7</th>
<th>C8</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1</td>
<td>1.00</td>
<td>0.0088</td>
<td>0.0002</td>
<td>0.5924</td>
<td>0.0234</td>
<td>0.0001</td>
<td>0.0262</td>
<td>0.0014</td>
</tr>
<tr>
<td>C2</td>
<td>0.0088</td>
<td>1.00</td>
<td>0.0410</td>
<td>0.0089</td>
<td>0.9499</td>
<td>0.0439</td>
<td>0.8706</td>
<td>0.1508</td>
</tr>
<tr>
<td>C3</td>
<td>0.0002</td>
<td>0.0410</td>
<td>1.00</td>
<td>0.0001</td>
<td>0.0698</td>
<td>0.8820</td>
<td>0.0506</td>
<td>0.3512</td>
</tr>
<tr>
<td>C4</td>
<td>0.5924</td>
<td>0.0089</td>
<td>0.0001</td>
<td>1.00</td>
<td>0.0120</td>
<td>0.0001</td>
<td>0.0182</td>
<td>0.0013</td>
</tr>
<tr>
<td>C5</td>
<td>0.0234</td>
<td>0.9499</td>
<td>0.0698</td>
<td>0.0120</td>
<td>1.00</td>
<td>0.0806</td>
<td>0.8621</td>
<td>0.3054</td>
</tr>
<tr>
<td>C6</td>
<td>0.0001</td>
<td>0.0439</td>
<td>0.8820</td>
<td>0.0001</td>
<td>0.0806</td>
<td>1.00</td>
<td>0.0482</td>
<td>0.4790</td>
</tr>
<tr>
<td>C7</td>
<td>0.0262</td>
<td>0.8706</td>
<td>0.0506</td>
<td>0.0182</td>
<td>0.8621</td>
<td>0.0482</td>
<td>1.00</td>
<td>0.1893</td>
</tr>
<tr>
<td>C8</td>
<td>0.0014</td>
<td>0.1508</td>
<td>0.3512</td>
<td>0.0013</td>
<td>0.3054</td>
<td>0.4790</td>
<td>0.1893</td>
<td>1.00</td>
</tr>
</tbody>
</table>

Table 4. The results from Wilcoxon’s test for all configurations
While conducting the rank sum test for the remaining configurations it is revealed that a grouping, in 3 categories, is possible. The first category in order of summation, features configurations number 1 and 4. The second category includes configurations number 2, 5 and 7, while the third category includes configurations number 3, 6 and 8 [Table 5].

<table>
<thead>
<tr>
<th>Configurations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group 1</td>
</tr>
<tr>
<td>C1</td>
</tr>
<tr>
<td>C4</td>
</tr>
<tr>
<td>—</td>
</tr>
<tr>
<td>Group 2</td>
</tr>
<tr>
<td>C2</td>
</tr>
<tr>
<td>C5</td>
</tr>
<tr>
<td>C7</td>
</tr>
<tr>
<td>Group 3</td>
</tr>
<tr>
<td>C3</td>
</tr>
<tr>
<td>C6</td>
</tr>
<tr>
<td>C8</td>
</tr>
</tbody>
</table>

Table 5. Grouping of statistically same configurations

It can be noticed that the subjective preference is highly influenced by the positioning of the rear height loudspeakers. Group 1 shares the same rear height loudspeaker with azimuth of ±90°, group 2 has rear height loudspeaker with azimuth of ±110° and group 3 utilized rear height loudspeaker with azimuth of ±130°. Furthermore, I conducted the Friedman’s test [Table 6], a nonparametric test for ranked data, which is similar to the classical balanced two-way analysis of variable (ANOVA). This test investigates column effects after adjusting for possible row effects. The results from Friedman’s test show that the column effect (loudspeaker configuration) accounting for the influence of the row effect (sound
source) was statistically significant. This implies the importance of the positioning of the height loudspeakers over the signals fed to the height loudspeakers. However the noteworthy aspect of the study is that the results indicate that despite the perceptual differences related to the room IRs, the perceived (overall) quality is significantly influenced by the positioning of the four height loudspeakers.

<table>
<thead>
<tr>
<th></th>
<th>C1</th>
<th>C2</th>
<th>C3</th>
<th>C4</th>
<th>C5</th>
<th>C6</th>
<th>C7</th>
<th>C8</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1</td>
<td>1.00</td>
<td>0.0095</td>
<td>0.0001</td>
<td>0.3677</td>
<td>0.0330</td>
<td>0.0001</td>
<td>0.0295</td>
<td>0.0016</td>
</tr>
<tr>
<td>C2</td>
<td>0.0095</td>
<td>1.00</td>
<td>0.0600</td>
<td>0.0091</td>
<td>0.8986</td>
<td>0.0293</td>
<td>0.9660</td>
<td>0.1586</td>
</tr>
<tr>
<td>C3</td>
<td>0.0001</td>
<td>0.0600</td>
<td>1.00</td>
<td>0.0002</td>
<td>0.0456</td>
<td>0.9312</td>
<td>0.1108</td>
<td>0.3672</td>
</tr>
<tr>
<td>C4</td>
<td>0.3677</td>
<td>0.0091</td>
<td>0.0002</td>
<td>1.00</td>
<td>0.0059</td>
<td>0.0002</td>
<td>0.0606</td>
<td>0.0024</td>
</tr>
<tr>
<td>C5</td>
<td>0.0330</td>
<td>0.8986</td>
<td>0.0456</td>
<td>0.0059</td>
<td>1.00</td>
<td>0.0893</td>
<td>0.8311</td>
<td>0.2632</td>
</tr>
<tr>
<td>C6</td>
<td>0.0001</td>
<td>0.0293</td>
<td>0.9312</td>
<td>0.0002</td>
<td>0.0893</td>
<td>1.00</td>
<td>0.1257</td>
<td>0.5500</td>
</tr>
<tr>
<td>C7</td>
<td>0.0295</td>
<td>0.9660</td>
<td>0.1108</td>
<td>0.0606</td>
<td>0.8311</td>
<td>0.1257</td>
<td>1.00</td>
<td>0.1538</td>
</tr>
<tr>
<td>C8</td>
<td>0.0016</td>
<td>0.1586</td>
<td>0.3672</td>
<td>0.0024</td>
<td>0.2632</td>
<td>0.5500</td>
<td>0.1538</td>
<td>1.00</td>
</tr>
</tbody>
</table>

Table 6. The results from Friedman’s test for all configurations indicate the significance of the proper height loudspeaker positioning in delivering an immersive sound field

### 3.2. STAGE II

In order to answer the question stated in 3.1, a second part of the experiment was conducted. The same set of subjects, as in Stage I, was asked to critically listen and describe perceptual characteristics associated with reproduction configurations. They first were asked to compared three conditions—conditions of the highest rank, and the lowest rank, and the references which were labeled A, B and
C. The set of choices is unique for each participant due to the fact that the formation of the set is based on the individual listener’s answers while perceptually ranking the randomly presented eight height loudspeaker configurations.

The participants, after listening to all three configurations that were again randomly presented in furtherance of ensuring legitimacy, had to match the two configurations that sounded the most similar to them. In addition, they were asked to use an adjective to describe their impression of the sound field. Furthermore the remaining configuration was deemed to be the most dissimilar and again the subjects were asked to elicit a descriptor to characterize the difference. This process took place four times, once for each signal. Those descriptors were later used to reveal the perceptual characteristics associated with surround music with height channels.

The GUI was programmed so that all the data were saved in two text files. One text file (having the file extension of .txt) contained the format in which the configurations and with their corresponding rankings were saved. Another text file contained the subject’s response for which configurations were perceived as most similar or dissimilar and coupled with the descriptor for each signal individually.
The triadic comparison consists of presenting items or objects in sets of three. The triadic method can be used to collect either similarity or ordered data. For similarity data, participants are asked to pick the item that is most dissimilar from the other two. For ordered data, the participants are asked to rank the items from “most” to “least” based on some other attributes (depending of the use), as an alternative of just choosing “the most dissimilar.” For Stage II, the subjects were asked to compare the similarity data. The method of triadic comparison has also
been broadly used in anthropology, psychology, and in other fields of the social sciences.

### 3.2.1 STAGE II RESULTS

A MATLAB function was created where the input would be the output .txt file from the GUI which contained the rankings of the subjects per configuration. The output of that function is demonstrated below [Table 7]:

| W8R2B1 |
|---|---|---|
| Txt File | Ranking | Configuration |
| W8 | Worst | Eighth |
| R2 | Reference | Second |
| B1 | Best | First |

**Table 7. Example of function’s output format**

whereas W(orst) indicates the configuration ranked lowest by the subject, R(efence) the reference configuration and B(est) the configuration ranked highest. The number next to each letter indicates the height loudspeaker configuration, so in this example the configuration ranked lowest was the eighth while the one ranked highest was the first and the reference was the second configuration of height sur-
round loudspeakers (configuration 2 is the reference configuration). In the case where the subject has indicated the best and worst configurations as the most similar, the descriptors were not taken into account.

Simultaneously, an EXCEL sheet was created containing all the descriptors used for each configuration from all the participants. Moreover the configurations were divided in two categories. One category accumulated all the descriptors used to accommodate a positive aspect of the configuration and the other for the negative descriptors. It needs to be noted that the descriptors were subjective and there was no list from which the participants could chose adjectives. After methodically arranging the descriptors, the following table has been created [Table 8].

<table>
<thead>
<tr>
<th>Configuration 1</th>
<th>Configuration 2</th>
<th>Configuration 3</th>
<th>Configuration 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Best</td>
<td>Worst</td>
<td>Best</td>
<td>Worst</td>
</tr>
<tr>
<td>Full</td>
<td>Wide</td>
<td>Response</td>
<td>Dark</td>
</tr>
<tr>
<td>Surrounded</td>
<td>Full</td>
<td>Unbalanced</td>
<td>Booming</td>
</tr>
<tr>
<td>Wide</td>
<td>Surrounded</td>
<td>Flat</td>
<td>Expanded</td>
</tr>
<tr>
<td>Narrow</td>
<td></td>
<td>Tight</td>
<td>Large Room Reverb</td>
</tr>
<tr>
<td>Wide</td>
<td></td>
<td>No Width</td>
<td>Enveloping</td>
</tr>
<tr>
<td>Configuration 1</td>
<td>Configuration 2</td>
<td>Configuration 3</td>
<td>Configuration 4</td>
</tr>
<tr>
<td>-----------------</td>
<td>-----------------</td>
<td>-----------------</td>
<td>-----------------</td>
</tr>
<tr>
<td>Full</td>
<td>Full</td>
<td>Surround</td>
<td>Full</td>
</tr>
<tr>
<td>Narrow</td>
<td>Spacious</td>
<td>Large</td>
<td>Elevated</td>
</tr>
<tr>
<td></td>
<td>Quiet</td>
<td></td>
<td>Sense of Elevation</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Crisp</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Configuration 5</th>
<th>Configuration 6</th>
<th>Configuration 7</th>
<th>Configuration 8</th>
</tr>
</thead>
<tbody>
<tr>
<td>Best</td>
<td>Worst</td>
<td>Best</td>
<td>Worst</td>
</tr>
<tr>
<td>Wide</td>
<td>Quiet</td>
<td>Direct</td>
<td>Distancing</td>
</tr>
<tr>
<td>Full</td>
<td>Narrow</td>
<td>Surround Reverb</td>
<td>Frontal</td>
</tr>
<tr>
<td>Surround</td>
<td>Broad</td>
<td>Harsh</td>
<td>Narrow</td>
</tr>
<tr>
<td></td>
<td>Dry</td>
<td>Flat</td>
<td>Balanced</td>
</tr>
<tr>
<td></td>
<td>Compressed</td>
<td>Low to the Ground</td>
<td>Dry</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Natural</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Sense of Height</td>
</tr>
</tbody>
</table>

Table 8. Spread sheet of all adjectives used by the participants
Nevertheless, there are a number of adjectives that are synonyms and can be used to portray the same aspect, hence a grouping of descriptors was necessary in order to alleviate the complexity and diversity of the issue at hand. The accumulated descriptors are in direct correlation with the descriptors used in Stage II for each configuration. It is obvious from Table 8 that configurations 1 and 4 received the largest number of positive descriptors but also the largest total number of adjectives used to characterize them. Furthermore, taking into account the positive descriptors used to portray configurations 1 and 4, it is seen that the perceptual characteristics consumers search for in a sound field that incorporates height ambiances, are envelopment, fullness and elevated-ness.

3.3 STAGE III

In Stage III, two configurations have been featured. One configuration utilized the loudspeakers positioned in ±50° and ±90°, which was ranked overall best. This was configuration 4. The second configuration utilized loudspeakers positioned in ±50° and ±130° was ranked overall lowest. This was configuration 6 and these results are seen in Stage I of the experiment. In this particular case, based on the text exports from EASERA that contain a matrix of the amplitude of the generated pure tone, expressed in dB Full Scale (dB FS) with respect to the frequencies of
the pure tone, a set of arbitrary response filters has been generated. The reason being is that the specific room added unwanted elements and colored the perceived sound, since it was not acoustically treated. As a consequence of the pursuit of eliminating the artifacts and being able to deliver a realistic reproduction of the two original venues, a set of inverse filters was created. A new set of participants was invited. The three new listeners were trained professionals.

3.3.1. STAGE III RESULTS

An informal experiment was conducted and three trained subjects evaluated the newly rendered and filtered audio field. The new subjects first went through Stages I and II to comprehend the process and purpose of this experiment. The experimental results, while informal, indicate that the subjective responses were not affected significantly by the spectral flattening despite the alteration on the perceived sound field. In simple words, the newly invited set of trained participants, after careful examination, ranked configuration 4 as the most appropriate configuration to deliver height ambiences to deliver a comprehensive and holistic 3D experience. However, all three subjects’ impression was unanimous that the filter-rendered sound field could be described as muddy, but the attribute of fullness remained intact. This allows the author to believe that digitally removing the
acoustic influence of the play back room, does not significantly change the auditory perception, despite the potential colorization of the generated sound field. This leads the author to believe that room acoustics have no significant effect while trying to create an immersive audio reproduction field, in this particular study.

4. DISCUSSIONS

In this section of the dissertation various concerns and questions that emerged during the experimentation will be discussed. For Stage I of the experiment, the initial assumption as to why the participants chose configuration 4 to be the most appropriate configuration, was due to the fact that the horizontal layer left an auditory gap (areas where there is a sound gap, or lack of sound) of 80º between the front-side (±30º) and rear (±110º) loudspeakers. Configuration 4 (±50º, ±90º) could fill in the auditory gap since it featured loudspeakers positioned right in the “empty space”, delivering height information while enveloping the listeners. Though, from similar experiments conducted in different countries of the world (Canada, Japan), the author cannot help but notice the fact that the results were significantly different. Hence, the previous assumption that the chosen configuration filled in the sound gap created by the horizontal
layer could be dismissed. There are three assumptions as to why the results were different. One assumption could include the fact that the replicated experiments were conducted in acoustically treated rooms, while another assumption is that the participating subjects were trained professionals. Moreover, the fact that cannot be excluded is the cultural variety of the participants, since it is known that subjective preference can be highly affected by cultural standards and living norms. Expanding in the fact that the replicated experiments were conducted in appropriate acoustically treated rooms a question of great importance arose:

_Do the room acoustics affect the listener’s decision?_

The difference of the experimental results from Stage I allows for the author to believe that, in fact, the room’s inherent acoustic response does play a role after all. Nevertheless, if a comparison takes place between the significance of a room and the positioning of the loudspeakers, the author believes that the appropriate positioning of the loudspeakers is more critical. That assumption is based purely on today's state of the art technology, as is it uncertain what would happen if a new loudspeaker design is released.
In addition, in Stage II of the replicated experiments the descriptors used to characterize the recreated audio field shared similarities with the ones used in this experiment. This gives the author ground to believe that the room has an active role in the immersive audio field creation. However, it is empirically known that the room’s acoustic influence is inversely proportional to the total number of speakers located in a single room.

In order to create the inverse filter in Stage III, a small number of obstacles needed to be overcome. Those obstacles are explained below. First, the exported information from EASERA was used (concerning the IRs from each loudspeaker) and with the help of the built-in MATLAB function (filterbuilder) the FIR filter was created. The IRs were captured in a 1/3rd of an octave increments (resolution) since this relates to the auditory sensitivity of the human system. From the early stages of the filter creation it became clear that there was a major drawback using that method. That assumption was based on the process time generating an FIR filter of 256 orders, which was unjustifiably time consuming relatively to the processing power of the hardware. The order of the filter was determined by the limitations of MAX software. Further on it was decided not to use the “filterbuilder” function, but rather manually program the filter using the “fdesing.arbmag” function. The processing time dropped from minutes to seconds. However,
the filter response was unexpected as it did not follow the inverse wave file pattern (magnitude vs. frequency). In order to alleviate that problem it was decided to modify the increments from 1/3rd of an octave to 1/24th of an octave, which had as a consequence the accumulation of larger sets of data. A normalization was deemed necessary as the filter’s absolute magnitude was greater than 1, which had as an outcome the clipping (loss) of valid data. A set of filters was designed for each of the 11 loudspeakers individually (5 on the horizontal layer and 6 on the height layer).

Moreover, in Stage III after removing the acoustic influence of the room, the informal experimental results indicate the small affect the room acoustics have in the subjective evaluation. The hypothesis is based purely upon the results, since the participants’ preference was not affected after applying the filter. There was however a distinguishable difference of color to the musical pieces, due to the application of the filtering. There is not yet any hard evidence to support the previous statement.

A few known problems will be listed below. First, the listeners that participated in Stage I and II of the experiment were not fully trained professionals, lacking a high level of critical listening skills. Another known issue that could compromise the experimental results would be the playback room that lacked acoustic treat-
ment, in an effort to simulate an average listening space and crowd. While analyzing the data, it has come to the author’s attention that a small number of subjects did not complete the entire process of ranking and the process of the triadic comparison. This problem led the author to perceptually manipulate the incomplete data and allow him to make assumptions as to the meaning the subjects were trying to convey.

5. FUTURE RESEARCH TOPICS

This research topic could be broadened and expanded by deepening the incorporation of the boundless field of psychoacoustics to cover practical technological needs. For the research at hand, that includes the enlargement of the sweet spot to a “sweet area”, meaning the widening of the optimal listener’s position, simultaneously accommodating a larger number of listeners. The sweet area could deliver the desired immersiveness to a larger group of audience, furthering the consumers' preferences.

It is empirically known that the room’s acoustic influence is inversely proportional to the number of loudspeakers psychically present in one room. However, there is not yet a precise formula that can calculate these inversely proportional ele-
ments. A topic of future research could be the investigation of the optimum number of loudspeakers according to room size. Primarily the question would be:

*How much does the room acoustics affect the perceived sound field*

Is there a formula that describes the effect of the room acoustics versus the number of loudspeakers in that room?

How can we use that formula to evaluate the appropriateness of various configurations in deliverance of height ambiences and recreation of an immersive audio field? If 10.2 and 22.2 perceptually have the same, or similar ranking, can we minimize the number of loudspeakers and still produce the same 3D experience?

6. CONCLUSIONS

This study investigated the interaction between the height loudspeaker positioning and their respective signals in delivering high quality multichannel music incorporating height ambiences. In Stage I of the article the results lead the author to believe that the significance of the positioning of the height channel loudspeakers is in fact greater than the actual signals being fed to said loudspeakers, with respect to the perceived quality of the sound field. Stage II answers the question “What
are the perceptual characteristics of a preferred sound field”. The participants listened back to the overall best and worst configurations along with the reference configuration. Later they were asked to elicit a descriptor to portray their subjective impression of the sound field. It has been revealed that the perceptual characteristics the listeners linked in multichannel music with height ambiences include envelopment, elevated-ness and fullness. Stage III is an attempt to subtract the room effect from the equation and see whether the individual perception has been altered. The experimental results indicate that in every case the subjective preference was indeed affected by the alterations made, but not in a degree that would thoroughly change their individually perceived quality.
7. REFERENCES


[25] Stanley Smith Stevens: A scale for the measurement of the psychological magnitude: loudness. See: Psychological Review. 43, Nr. 5, APA Journals, 1936, pp. 405-416


       68th Audio Engineering Society Convention, Los Angeles. Preprint 1766


[36] Antonios Karampourniotis, Sungyoung Kim et al., “Significance of loudspeaker positioning for perceived immersive sound field reproduction”
       Acoustical Society of America, May 2014