

Li, Yangke

2015

Optimum height for surround recordings with height channels

Department of Music

<https://hdl.handle.net/10133/3764>

Downloaded from OPUS, University of Lethbridge Research Repository

**OPTIMUM HEIGHT FOR SURROUND RECORDINGS WITH HEIGHT
CHANNELS**

YANGKE LI
Graduate Degree, University of Lethbridge, 2015

A Thesis Submitted to the School of Graduate Studies of the University of Lethbridge in
Partial Fulfilment of the Requirements for the degree

MASTER OF MUSIC
(DIGITAL AUDIO ARTS)

Department of Music
University of Lethbridge
Lethbridge, Alberta, Canada

© 2015, Yangke Li

**OPTIMUM HEIGHT FOR SURROUND RECORDINGS WITH HEIGHT
CHANNELS**

YANGKE LI

Date of Defence: August 7, 2015

**Thilo Schaller
Supervisor**

Assistant Professor M.A.

**Dr. Rolf Boon
Thesis Examination Committee Member**

Associate Professor Ph.D.

**Dr. Paul Sanden
Thesis Examination Committee Member**

Assistant Professor Ph.D.

**Dr. Brian Black
Chair, Thesis Examination Committee**

Associate Professor Ph.D.

OPTIMUM HEIGHT FOR SURROUND RECORDINGS WITH HEIGHT CHANNELS

ABSTRACT

This research investigates the perceptual qualities of various microphone distances between the low channel microphones and height channel microphones in order to decide optimum height for height microphones in three-dimensional acoustic recordings.

First, the thesis introduces information about the added value of the height channels in audio recording, and then it presents the development of recording techniques and technology, and the current knowledge and limitations of the related recording methods. Next, the thesis discusses the methodology and the process of the research, followed by data analysis, discussion and conclusions about a listening test involving multiple distances between lower and height channels. Finally, the limitations of the research are addressed and suggestions for future research are offered.

The results of the listening test show significant results on depth perception and envelopment. However, the rest of the results show no major significant difference between the perceptions of different height channel setups.

Table of Contents

ABSTRACT	iii
List of Tables	v
List of Figures	vi
List of Examples	vii
INTRODUCTION	1
LITERATURE REVIEW	8
Framing of the research	14
RESEARCH	16
Methodology	17
<i>Recording</i>	17
<i>Listening Test</i>	18
<i>Data Analysis</i>	19
<i>Pilot test</i>	21
<i>Surround Recording with Height Channels Implementation</i>	22
<i>Listening Tests Implementation</i>	24
RESULTS	28
<i>Localization focus</i>	28
<i>Depth perception</i>	28
<i>Spatial impression</i>	30
<i>Envelopment</i>	30
<i>Overall timbre</i>	31
<i>Overall preference</i>	33
DISCUSSION	33
<i>Localization focus</i>	33
<i>Depth perception</i>	37
<i>Spatial impression</i>	38
<i>Envelopment</i>	39
<i>Overall timbre</i>	40
<i>Overall preference</i>	41
<i>Timbre influences by orchestration</i>	42
<i>Deviations caused by the playback order</i>	43
<i>Limitations of the test</i>	44
<i>Insufficient number of participants</i>	44
<i>Phrasing issues</i>	45
CONCLUSION	46
BIBLIOGRAPHY	50
APPENDIX I: QUESTIONNAIRE	52
APPENDIX II: LETTER OF CONSENT	56

List of Tables

Table 1, Relationship between sound attributes and sound-field types	3
Table 2, Comparison of different playback format profiles.....	15
Table 3, Descriptives of depth perception	29
Table 4, Multiple comparisons of different factors of analysis of depth perception	29
Table 5, Descriptives of spatial impression	30
Table 6, Descriptives of envelopment	31
Table 7, Descriptives of overall timbre	32

List of Figures

Figure 1, ITU-R BS.775-3, speaker arrangement for the 5.1 surround setup.....	5
Figure 2, Tape recorder head.....	11
Figure 3, Top view of the recording setup.....	23
Figure 4, Side view of the recording setup.....	24
Figure 5, Sample of the listening questionnaire.....	26
Figure 6, The interface subjects worked with.....	27
Figure 7, Means plots of depth perception.....	29
Figure 8, Means plots of envelopment.....	31
Figure 9, Means plots of overall timbre.....	33
Figure 10, Pie chart analysis of preference.....	33
Figure 11, Principle radiation region of a tuba.....	36
Figure 12, Setup of the recording session.....	36
Figure 13, Equal loudness contours.....	41

List of Examples

Examples 1, Oskar Böhme, Trompeten-Sextett, Op.30, i, mm 15-16.....	37
Examples 2, Oskar Böhme, Trompeten-Sextett, Op.30, i, mm 112-114.....	37

INTRODUCTION

The study of human perception of sound considers psychology and psychoacoustics. Psychoacoustics refers to the perception of physical sound stimuli,¹ which relies on the physical structure of the ear, sound pathways and their function, the human brain's interpretation of sound, and their interrelationships. It is the basis for the entire field of audio engineering. It is the foundation of perceptual coding such as MP3 and WMA. It also plays a vital role in architectural acoustics, informing us how a room's sound field is interpreted by the listener. Therefore psychoacoustics has always been a very important factor the recording industry takes into consideration in order to deliver different experiences to listeners. For example, in recording classical music, engineers aim to take the live experience and translate it into the recording; in recording popular music, the live experience often takes on a quite different form under the engineer and producer's aesthetic influence. In a natural acoustic listening environment, for instance, when listening to classical music in a concert hall, our ears not only perceive sound sources in front of us, but also reflections from the ceiling, floor, and the walls around us.

If a piano player plays a note in a concert hall, first we will hear the direct sound that travels from the musician to us. This information tells us where the sound is coming from. The time gap between the direct sound and the first reflection is called the initial time delay gap. It usually indicates the size of the room. From 15 milliseconds to approximately 50 milliseconds after the direct sound onset, we start hearing reflections known as early reflections from the ceilings, walls, and floors. Early reflections tell our

¹ Stanley R. Alten, *Audio in Media: The Recording Studio* (New York: Schirmer, 1995), 29.

brain the apparent source width² and the distance of the sound source. They also give us a hint of spaciousness. After the early reflections until the sound energy completely fades under the ambient noise,³ is the reverberant tail of the sound. It contains a massive amount of reflections and they decay in a logarithmic manner. From these reflections our brain interprets the spatial impression of the room.

When we are listening to a concert or any kind of live show in a real acoustic environment, we naturally perform an evaluation on various properties of these three factors, or in other words, the brain is capable of intuitively and spontaneously distinguishing between direct sound, early reflections and reverberation for spatial hearing.⁴ Our perception of distance away from the sound source(s), depth of the ensemble and feeling of sitting in a concert hall are provided by these primary concepts: initial time delay gap, early reflections, and reverberation. As Table 1 indicates, when sitting in a concert hall, a listener's perception of spaciousness and envelopment is closely related to indirect sound, such as early reflections, reverberation and background noise.

² Apparent sound width denotes a characteristic spatial spreading of the auditory events. The effect of spatial spreading of auditory events due to lateral reflections depends on their rise-time. If the events have a short rise-time compared to the arrival of major lateral reflections they will be perceived as sharply localised. Günther Theile and Helmut Wittek, "Principles in Surround Recordings with Height." *AES Convention 130*, 8403 (2011):6.

³ F. Alton Everest and Ken C. Pohlmann, *Master Handbook of Acoustics* (New York: McGraw-Hill, 2009), 152.

⁴ Theile and Wittek, "Principles in Surround Recordings with Height".

Table 1, Relationship between sound attributes and sound-field types⁵

SOUND ATTRIBUTES IN THE HALL	DIRECT SOUND	EARLY REFLECTIONS	REVERB	BACKGROUND NOISE
Direction/elevation	••	•		
Distance/depth		••		
Spatial impression		••	•	
Envelopment			•	••
Timbre	••	•	••	

In a play back situation, it is a completely different experience because the directional cues in the direct sound play an even less important role than they do in natural hearing. For mono playback, direct sound, pre-delay time, early reflections, and reverberation are all forced into a single spot creating a heavily colorized sound mixture.⁶ Therefore, it is very difficult for our ears to achieve an intuitive distinction between direct sound, early reflections and reverberation for spatial hearing. In conventional two-channel stereo recording and playback, the impression of spaciousness has to be recreated by two frontal speakers between a 60-degree angle, which are also used to provide localization cues of sound sources using direct sound. Inevitably sound engineers have to carefully balance the level of direct sound, indirect sound, and patterns of reflections without causing too much loss of intelligibility. Two-channel stereophony is entirely based on the knowledge of binaural localization, in other words, how directional information is perceived by the human hearing mechanism. When reproduced by loudspeakers in a two-channel system, the time or the level differences (or both) between the sound signals provide this directional information to the listener. If there is no time or level difference between the left and right channel, a sound source is reproduced at a 0

⁵ Ibid. In this table, the numbers of dots indicate the effectiveness of the attributes.

⁶ Ibid.

degree angle, directly in front of the listener. As time or level differences increase, localization of sound source will start shifting to one side. Stereophony achieves a huge improvement of realism compared to mono audio by spreading the sound elements onto a 60 degree angle plane, but it is still a significant downgrade in comparison to a real acoustic situation.

In the case of 5.1 surround sound, the listening experience is improved by the use of additional surround speakers. The ITU-R BS.775-3 standard specifies the arrangement of the five speakers (Figure 1). In order to create a realistic and coherent acoustic environment, it is recommended by the Audio Engineering Society that all five speakers (left, centre, right, left-surround, and right-surround) be identical in frequency response (20 Hz to 20 kHz), directivity index, nonlinear distortion attenuation, time delay, and dynamic range/maximum operating level.⁷ Studies have shown that with early reflections distributing to a three hundred and sixty degree horizontal plane, not only can the apparent source width be defined more easily, but also a much more natural and realistic spatial impression can be created.

⁷ Dolby, "5.1-Channel Music Production Guidelines" (2003): 36.

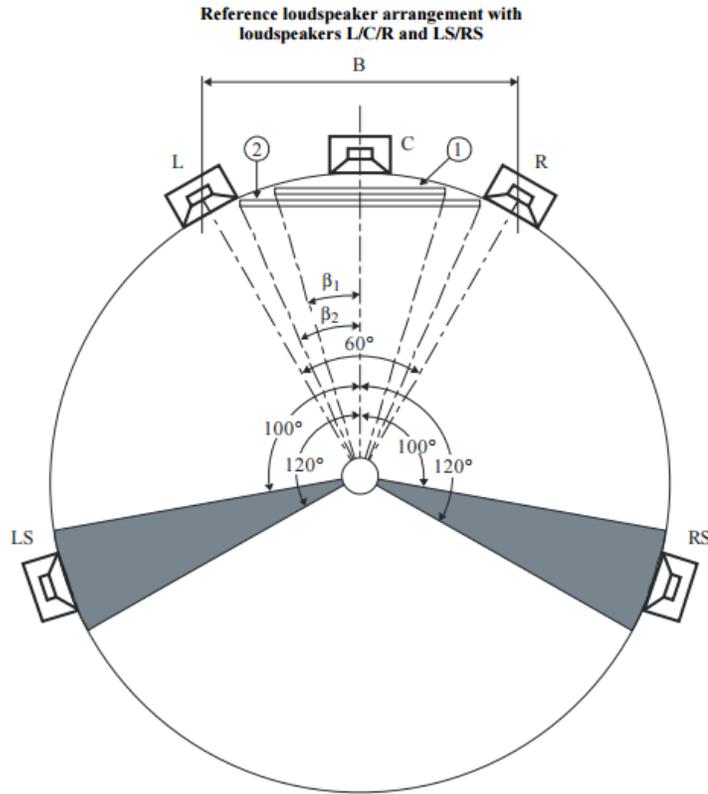


Figure 1 ITU-R BS.775-3, speaker arrangement for the 5.1 surround setup⁸

Since 5.1 surround is a different playback format than traditional stereo, both the direct sounds of a sound source and natural reflections have to be recorded in 5.1 format with at least five microphones capturing direct sound from the front and reflections from the sides and the back. At the same time, it has to be taken into consideration that about up to 50% of the indirect sound energy should be positioned into the left and right surround channels in post-production.⁹ Thus, the polar pattern of the microphones used in such a case and distance from sound sources should be carefully chosen in order to provide sufficient channel separation. According to many audio engineers' experience and

⁸ International Telecommunication Union, *Recommendation ITU-R BS.775-3, Multichannel stereophonic sound system with and without accompanying picture*. ITU, 2008, PDF.

⁹ Günther Theile, "Multichannel Natural Music Recording Based on Psychoacoustic Principles", *AES 19th International Conference* (2001): 26.

studies,¹⁰ recording in a 5.1 situation proves to be difficult, because there are more parameters and aspects affected by psychoacoustic principles and practical limitations. In addition, theory and a number of experiments have shown that lateral sound source localization is extremely unstable, and it is subjective to a listener's location and type of sound source even if the listener sits in the "sweet area."¹¹ Therefore, lateral areas in the 5.1 surround format are not suitable for stable sound source localization. Fortunately, the instability of lateral localization is not very crucial to reproducing a convincing spatial impression. Thus, 5.1 surround with added surround speakers significantly extends the possibilities of 2-channel stereophony with respect to realistic acoustic audio production compared to monaural and two-channel sound.

In a concert hall, the best listening position is usually far away from the stage. According to Collin Symes, sitting in the twelfth row of a concert hall often gives audiences a satisfactory listening experience.¹² However, in order to reproduce audio with a reasonable sense of realism using conventional two-channel stereophony or multi-channel playback configurations, engineers have to place microphones very close to the performers in order to increase the clarity. In addition, they have to carefully capture the seating arrangement of instruments using proper recording techniques. Sometimes engineers have to change microphone placement during recording in order to alter frequency spectra of certain instruments or polarities of the signal to achieve better intelligibility.

¹⁰ Theile and Wittek, "Principles in Surround Recordings with Height", 2.

¹¹ Theile and Wittek, "Principles in Surround Recordings with Height", 1.

¹² Colin Symes, *Setting the Record Straight: A Material History of Classical Recording* (Connecticut: Wesleyan, 2004), 62.

Psychoacoustics is not the only factor creating distinction in our comparative experiences of listening to recorded music and to live music. Psychology plays an important part as well. The human brain combines multiple senses including sight, hearing, taste, smell and touch to create a fully functioning all-dimensional cognition of the environment that surrounds us. Most of the cognitive acknowledgement of humans is done through sight. A concert taking place in a concert hall does not only contain sound. The internal structures, décor of the hall, musicians' and conductor's movements and their interpretation of the music also affect our perception of the surrounding environment. In addition, sitting together with five hundred people, hearing and seeing their applause will definitely be missing if the listening took place in a living room with two speakers in front of the listener. Different listening venues and different audiences might even give rise to different emotional feelings towards the music. Listening to music on a stereo system in a living room is a completely different emotional and psychological experience compared to enjoying a concert in a concert hall. Therefore it is very difficult to achieve absolute realism of classical music through loudspeaker playback.

In 2001, a surround recording format with added height channels was introduced by Werner Dabringhaus with his 2+2+2 recording technique.¹³ It uses two speakers at the back and two height speakers positioned on top of the left and the right channels of the traditional stereo playback system. The objective was to reproduce the sound from the concert hall as realistically as possible.¹⁴ In 2006, Wilfried Van Baelen introduced the Auro-3D format that involves four extra channels for height information. Due to the fact that this setup specifies four extra channels for height information, this thesis studies the

13 Theile and Wittek, "Principles in Surround Recordings with Height", 2.

14 Ibid.

perception of quality of different microphone distances between the low channel microphones and height channel microphones. It is believed that the conclusions of this research can be applied to the future study of application of height channel microphones in acoustic recordings, providing a further improvement on the limitations of the stereo and 5.1 surround formats, at least from the perspective of psychoacoustics.

LITERATURE REVIEW

In over a hundred years since acoustic recording became possible, the evolution of recording technology has included the introduction of monophony, stereophony, 5.1 surround, 7.1 surround and many other formats, and every one of them aimed to overcome the disadvantages of its predecessors, in order to create a more realistic listening environment. The 2+2+2 format, the first multi-channel recording with height information, was introduced in 2001. It utilizes the upper half of a room for reproduction of ceiling reflections in order to take the realism of sound recording one step further. This new format better accommodates our psychoacoustics under real situations through the implementation of height channels into audio production.

Despite the huge differences between recorded music and live music, the recording industry never ceases trying to faithfully reproduce the live concert experience, especially in the classical genre. As Symes puts it, “Humankind has long treasured the prospect of machines that had the capacity to preserve the transient beauties of music.”¹⁵ The audio timeline below from the Audio Engineering Society website summarizes the

¹⁵ Symes, *Setting the Record Straight: A Material History of Classical Recording*, 62.

course of development from documenting a single sound to aiming to deliver realistic listening experiences in about one and a half centuries.¹⁶

- 1877 Thomas Alva Edison succeeds in recovering Mary's Little Lamb from a strip of tinfoil wrapped around a spinning cylinder.
- 1929 Harry Nyquist publishes the "Nyquist Theorem," which is the mathematical foundation for the sampling theorem basic to all digital audio processing.
- 1931 Alan Blumlein, working for Electrical and Musical Industries in London, patents stereo. His seminal patent discusses the theory of stereo, both describing and picturing in the course of its 70-odd individual claims a coincident crossed-eights miking arrangement and a "45-45" cutting system for stereo disks.
- 1941 Arthur Haddy of English Decca invents the first motional feedback, lateral-cut disk recording head, later used to cut their "ffrr" high-fidelity recordings.
- 1953 Ampex engineers a 4-track, 35 mm magnetic film system for 20th-Century Fox's Christmas release of "The Robe" in CinemaScope with surround sound.
- 1954 The first commercial 2-track stereo tapes are released.
- 1958 The first commercial stereo disk recordings produced by Audio Fidelity.
- 1969 Dr. Thomas Stockham begins to experiment with digital tape recording.
- 1975 Digital tape recording begins to emerge in professional audio studios.
- 1980 3M, Mitsubishi, Sony and Studer each introduces a multitrack digital recorder.
- 1981 Philips demonstrates the Compact Disc (CD).
- 1990 Dolby proposes a 5-channel surround-sound scheme for home theatre systems.
- 1996 Experimental digital recordings are made at 24 bits and 96 kHz.
- 1999 Audio DVD Standard 1.0 agreed upon by manufacturers.

¹⁶ For a more complete history, see <http://www.aes.org/aeshc/docs/audio.history.timeline.html>.

In 1937, Bell Laboratories successfully incorporated two-channel stereophony into motion pictures, which was a milestone for sound reproduction for the film industry. However, it took another twenty years before two-channel stereo sound was accepted by the public and commercially available to the market. The invention of the compact disc by Sony and Philips in 1980 and its subsequent domination of the market pushed stereophony to its peak of popularity around the year 2000. In the early 1940s, the Disney studio started experimenting with animated movies with more than two channels, but it was Dolby labs that introduced the 5.1-surround format (three screen channels, two surround channels and one low-frequency effects channel) into the film industry in 1992. It has been a magnificent improvement of realism compared to the phonograph. From the earliest days of recording, and with each new development after that, it was regularly claimed that “the last word in sound had been reached and that records were indistinguishable from their sources. But in fact records fell far short of this.”¹⁷

The recording industry started with analogue audio, which was used in phonograph, gramophone, and tape recording. Analogue recording devices such as phonographs and gramophones use a continuous signal to represent and store analog sound waves mechanically. During playback, the needle or stylus traces the groove or the wiggles exactly to reproduce the motion of the diaphragm of the microphone at the time the recording was done.¹⁸ The principle of tape recording is very similar. The tape is a strip of plastic which has been coated with a material that is easily magnetized. Tape deck mechanisms have three heads, and all of them are constructed the same way: a C-shaped

¹⁷ Symes, *Setting the Record Straight: A Material History of Classical Recording*, 62.

¹⁸ Peter Elsea, “ANALOG RECORDING OF SOUND,” University of California, Santa Cruz, accessed July 1, 2015, http://artsites.ucsc.edu/EMS/music/tech_background/te-19/teces_19.html.

piece of metal with the very narrow gap of the "C" near the tape (Figure 2).¹⁹ A coil of wire around the metal can serve to either detect or produce magnetic fields at the gap. If a strong current is passed through the coil, a field is produced which creates a magnetic spot on the nearby tape. The amount of magnetism will be proportional to the amount of current. If the tape is moved and the current varied in a periodic way, a track of magnetic areas will be imprinted on the tape.²⁰

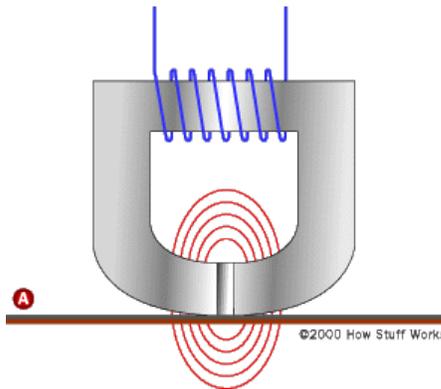


Figure 2, Tape recorder head²¹

Modern digital sound recording/reproduction devices convert continuous signals to streams of discrete numbers. The conversion is done by a device called an analog-to-digital converter (ADC). To play back the music, the stream of numbers is converted back to an analog wave by a digital-to-analog converter (DAC).²²

The analog wave produced by a DAC is amplified and fed to the speakers to produce the sound. There are two major variables involved in this process that decide the resolution of the signal conversion: the sampling rate controls how many samples are taken every second. The bit depth controls how many different gradations (quantization

¹⁹ "ANALOG RECORDING OF SOUND."

²⁰ Ibid.

²¹ "How Tape Recorders Work," accessed June 30, 2015, <http://s.hswstatic.com/gif/tape-head.gif>.

²² Marshall Brain, "How Analog and Digital Recording Works", accessed June 30, 2015, <http://electronics.howstuffworks.com/analog-digital3.htm>.

levels) are possible during sampling. In other words, sampling rate is how fast the analogue signal is being measured, and bit depth is how accurate the measurement is. Generally speaking, the higher sampling rate and bit depth are, the better the resolution with which the sound waves are heard.

It is necessary to state that analogue recording is not objectively superior to digital recording or vice versa. High quality sound reproduction is possible with both systems. However, digital systems do overcome some of the problems and inherent weakness of analogue systems, such as tape hiss and wear due to mechanisms and copying. In the acoustic era, there were mechanical handicaps inherent in the playback equipment that detracted from the realism of recorded sound.²³ In addition, performing errors were difficult to fix, therefore musicians had to be right the first time. The advent of magnetic tape in 1928 changed this, and together with overdubbing, patching, and splicing, engineers and producers were able to produce almost perfect performances.²⁴ Still, there were some restraints caused by the technology itself. For example, most of the analogue systems such as microgroove or magnetic tape are unable to reproduce the full dynamic range of a symphony orchestra. Their sound was thus a product of making acceptable compromises involving a compression of auditory reality.²⁵

On the other hand, digital systems are able to handle such a huge dynamic range over the entire audible frequency spectrum (20Hz to 20000Hz) if the bit depth is high enough, and the bit depth of a digital system is almost potentially and theoretically infinite. Both analogue and digital systems inevitably generate noise that is not related to

²³ Symes, *Setting the Record Straight: A Material History of Classical Recording*, 62.

²⁴ Ibid. Although tape was introduced in 1928, it wasn't actually widespread in the industry until the 1940s

²⁵ Ibid, 80.

the original music in the process of recording, storing and playback. The amount of noise that a piece of audio equipment adds to the original signal can be quantified.

Mathematically, this can be expressed by means of the signal to noise ratio. Analogue systems tend to have lower signal to noise ratios.²⁶ Magnetic tape and vinyl records suffer from tape hiss noise and background noise caused by structures and principles of how analogue systems function. Signal to noise ratio is significantly improved in digital systems, but the process of digital conversion and other types of processing will always add some noise. For example, signal-dependent noises of quantization errors are unavoidably introduced by the ADC because digital systems use discrete digital levels to measure the amplitude of incoming sound waves. Relatively speaking however, digital systems provide a higher level of signal to noise ratio, so the recorded content does not have to fight its way through background noise.

In addition, with vinyl records, there is some loss in fidelity every time it is played. This is due to the wear of the stylus in contact with the record surface. Magnetic tapes wear from friction between the tape and the heads, guides, and other parts of the tape transport as the tape slides over them.²⁷ Tapes can also suffer creasing, stretching, and frilling of the edges of the plastic tape base, particularly from low-quality or out-of-alignment tape decks. Digital systems, such as the compact disc, almost entirely eliminated noise related to physical degradation.²⁸ When a Compact Disc (CD) is played, there is no physical contact involved, and the data is read optically using a laser beam. In addition, error detection and correction of digital systems ensure that potentially corrupted data does not interfere with the reconstruction of a waveform. Therefore, there

²⁶ Jan Maes and Marc Vercammen, *Digital Audio Technology* (Massachusetts, 2001), 33.

²⁷ Fred R. Byers, *Care and Handling of CDs and DVDs* (Maryland: Gaithersburg, 2003), 12-15.

²⁸ *Ibid.*

is no media deterioration taking place,²⁹ and with proper care the CD will sound almost exactly the same every time it is played (disregarding aging of the player and the CD).³⁰

Framing of the research

As has been expounded, in a real acoustic environment, omni-directional hearing gives us a feeling of spaciousness and envelopment, which is an effect that traditional stereo and even surround playback systems cannot provide. Also because of limitations of the 5.1 surround format, new multichannel sound formats extend 5.1-surround playback systems with height channels that add the third dimension to recordings. In 2001, surround recording with added height channels was introduced by Werner Dabringhaus with his 2+2+2 recording technique.³¹ In 2005, Wilfried Van Baelen introduced the Auro 9.1 format that adds four extra channels for height information. This 9.1 loudspeaker setup (a system that adds four height speakers to 5.1-surround system (ITU-R BS.775-3)) is capable of delivering improved realism in terms of direct sound, reverberation, and ambience sound. The four additional speakers in the upper hemisphere are used to reproduce early reflections and diffuse sound with the purpose of offering a better perception of envelopment, distance, depth, and spatial impression. Furthermore, they provide the same stereo imaging as traditional stereo formats, giving this format the ability of reproducing sound sources that are located throughout the upper hemisphere (e.g., airplanes, birds, etc). Auro Technologies claim the Auro 3D format³² they developed is the most efficient format for three-dimensional audio reproduction. It meets

²⁹ "All about digital audio," Sound on Sound, accessed July 6, 2015, <https://www.soundonsound.com/sos/jul98/articles/digitalbasics3.html>. If data corruption happens, a digital system will try either concealing or correcting the error. If too many errors occur, system failure will happen.

³⁰ Byers, *Care and Handling of CDs and DVDs*, 12-15.

³¹ Theile and Wittek, "Principles in Surround Recordings with Height", 2.

³² Wilfried Van Baelen also introduced 11.1 and 13.1.

many of today's requirements for a universal and compatible future-oriented standard for digital cinema, games, broadcast, and the music industry.³³ Table 2 shows the comparison of different formats in terms of attributes of sound reproduction. As is shown in the table, Auro-3D 9.1 potentially offers better spatial impression and envelopment with added height channels.

Table 2 Comparison of different playback format profiles³⁴

ATTRIBUTES OF SOUND REPRODUCTION	2.0 STEREO	5.1 SURROUND	AURO-3D 9.1	WFS*	BINAURAL TECHNIQUES
Front direction	•	••	••	••	•
Surround direction		•	•	••	••
Elevation			(•)***		••
Height			•		••
Distance/depth	(•)**	•	••	••	••
Proximity to the head				•	••
Intra-active perspective / ₁				••	
Spatial impression	(•)**	•	••	•	••
Envelopment		•	••	•	••
Timbre	••	••	••	•	••

As always, this new type of playback format has been introduced into the film industry, and there has been quite extensive research on height speaker arrangement for this format such as integration with high-definition television and 3D video, influence of positions of height speakers on perceived sound quality, localization in the vertical plane, elevation perception, and so on.³⁵ However, due to the complexity of principles of human natural hearing, psychoacoustic phenomena and practical limitations, spatially realistic resolution of direct sound, early-reflections, and diffuse sound need to be taken into consideration; thus it has become very difficult to find a suitable microphone technique

³³ Theile and Wittek, "Principles in Surround Recordings with Height", 4.

³⁴ Theile and Wittek, "Principles in Surround Recordings with Height", 4.

³⁵ Barbour, "Exploration of the Height Dimension in Audio Reproduction," *Australasian computer music conference* (2004): 3.

that meets all the requirements.³⁶ Only a few microphone techniques for this format have been proposed, and they merely follow the existing horizontal 5.1 surround microphone techniques and add a height layer with largely spaced omni-directional or super-cardioid microphones from the lower layer. Hyunkook Lee and Christopher Gribben from the University of Huddersfield conducted a series of listening tests using impulse responses³⁷ and virtual instruments to investigate how different distances between base and height microphones affect perceived spatial impression and overall preference.³⁸ However, their research did not use real musical ensembles and it only involved twelve participants. In addition, the participants only observed two attributes—spatial impression and overall preference—to evaluate the quality of different height channel setups. Therefore it is necessary to conduct further research on this topic and look into the perception of general quality and spatial impression of different microphone spacings between the low microphones and height microphones.

RESEARCH

This study consisted of four different parts: A carefully designed and executed three-dimensional chamber music recording with different distances between the low and height channels, a listening test that involved mainly students and faculty of the Digital Audio Arts program at the University of Lethbridge, data analysis and discussion.

³⁶ Theile and Wittek, "Principles in Surround Recordings with Height", 4.

³⁷ In signal processing, the impulse response refers to the reaction of a system in response to external change.

³⁸ Hyunkook Lee and Christopher Gribben. "On the Optimum Microphone Array Configuration for Height Channels." *AES Convention*:134 (2013): 93.

Methodology

Recording

In natural multi-channel recording, in order to capture and deliver a satisfactory direct sound image, early reflections, reverberation, and ambient sound, a suitable recording technique with sufficiently low inter-channel cross talk needs to be designed.³⁹ However, experiments have already indicated that 100% inter-channel decorrelation results in a very stable phantom image but almost no spatial impression.⁴⁰ Therefore it is important to maintain a reasonable level of inter-channel correlation between channels in natural multi-channel recording, and it is particularly true for lateral and rear signals. Since this study only investigated the optimum height for height microphones in three-dimensional recording, it was important to maintain the consistency of the surround setup when the following listening tests were conducted, and therefore participants' perceptions would only be influenced by the height channel microphones and the constant surround layer. For the height layers, in order to eliminate possible discrepancies introduced by different frequency responses of the microphones, the same type of microphones with the same polar pattern were used. For professional musicians, there would still be differences if they were to play the same piece twice and those differences could have significant influence on the results of the listening tests. Therefore it was important to record the ensemble with microphones of all the layers instead of recording several times for different layers.

³⁹ Theile, "Multichannel Natural Music Recording Based on Psychoacoustic Principles", 10.

⁴⁰ *ibid*, 25.

Listening Test

Hedonic tests are used to measure liking, usually in terms of acceptability or preference. The most common methods of hedonic tests are: (i) paired comparison tests (usually paired preference); (ii) ranking tests; or (iii) scoring tests using category scales.⁴¹ Since the purpose of this study was to test participants' preference on different heights of microphone setups, a scoring test using scales was chosen to be the method. Twenty-three subjects were recruited as participants. Because of what this study intended to investigate, most of the testing subjects were Digital Audio Arts students who had training in the field of audio engineering. They were experienced subjects with a very clear idea about the attributes they were listening to.

The listening test was a single-blind test, which is a method of testing in which the participants are not informed about the predispositions or assumptions of the study. This test method prevents results from being influenced by any a priori information.⁴² In the field of audio, blind tests truly highlight what a listener is able to hear.⁴³ A single-blind experimental design is used where the researchers or experimenters either must know the full facts and so the researchers/experimenters cannot themselves be blind, or where the researchers/experimenters will not introduce further bias and so the experimenters need not be blind. However, there is a risk that subjects might be influenced by interaction with the researchers – known as the experimenter's bias.⁴⁴ Therefore, in order to eliminate any conscious or subconscious influence from the experimenter, this particular single-blind listening test was designed in a way that there

⁴¹ M.J Saxyby, *Food Taints and Off-Flavours* (New York: Springer, 2012),19.

⁴² "Blind Listening Tests," accessed July 2, 2015, http://www.audiocheck.net/blindtests_index.php.

⁴³ *Ibid.*

⁴⁴ David L. Sackett, "Bias in analytic research," *Journal of Clinical Epidemiology* 32 (1979): 51–63.

was no interaction between the researcher and the participants after the start of the tests by using a system, which utilized two computers. One was used by the subjects to navigate the audio signal through different samples at any time, and the other one was used to receive, complete and deliver the questionnaire.

The listening test used sample sets that were arranged as follows: 1. surround only (LS); 2. surround and first layer information (L1); 3. surround and second layer information (L2); 4. surround and third layer information (L3); 5. a random duplication of one of the combinations. The reason for this combo design was that height channels are add-ons that improve the realism of the recording. This research would have been too theoretical if only height information had been evaluated since in practice, multichannel recordings with height channels always contain surround information.

Randomization was applied as a method of experimental control in order to prevent selection bias and accidental bias.⁴⁵ The order of samples in each set was randomized, and the duplicated sample was randomly selected. Finally, the order of sample sets given to all the participants was randomized.

Data Analysis

The purpose of analyzing the acquired data was to examine participants' perception of quality and impression of different microphone spacing between the low channel microphones and height channel microphones. Furthermore, data analysis might reveal relationships, connections, patterns, or trends between the participants' perception and the height of upper channels, and the results would pave a way for future similar studies.

⁴⁵ Ibid.

In statistics, scientific statements can be split into testable hypotheses. The hypothesis or prediction that comes from a researcher's theory usually states that an effect will be present.⁴⁶ This hypothesis is called the alternative or experimental hypothesis and is denoted by H_1 . There is another type of hypothesis, and this is called the null hypothesis and is denoted by H_0 . The null hypothesis usually states that an effect is absent or not true. We need the null hypothesis because we cannot prove the experimental hypothesis by using statistics, but we can reject the null hypothesis. "If our data give us confidence to reject the null hypothesis then this provides support for our experimental hypothesis."⁴⁷ Rejecting the null hypothesis proves the experimental hypothesis, i.e. there are relationships between two or more factors. The incorrect rejection of a true null hypothesis is known as the type 1 error. The probability of making a type 1 error is the α level. When the level of significance—p value—is calculated, the researcher will compare the p and the α .⁴⁸ If $p > \alpha$, the result is not significant and therefore the experimental hypothesis is rejected. If $p < \alpha$, the result is significant and the null hypothesis is rejected meaning there are relationships between factors/phenomena. The α value is normally set at 0.05 (5%), but due to the fact that the listening test in this study only tested a very small group of participants, the α level was set at 0.1 (10%) to increase the probability of accepting an effect as genuine, or decrease the probability that we will reject an effect that genuinely existed.⁴⁹

An analysis of variance (ANOVA) was used to analyze the data. This method is used to analyze differences or variations between two or more groups on certain variables.

⁴⁶ Andy Field, *Discovering Statistics Using SPSS* (New York: SAGE Publications Ltd, 2013), 27.

⁴⁷ Ibid.

⁴⁸ P value is the probability level of alternative hypothesis. It is the estimate of rejecting null hypothesis.

⁴⁹ Field, *Discovering Statistics Using SPSS*, 27.

ANOVA offers statistical ways of testing if the means (average value) of several groups are significantly different.⁵⁰ Additionally, ANOVA provides detailed inspection of data by generating means, standard deviation,⁵¹ confidence interval,⁵² F-value,⁵³ sum of squares,⁵⁴ degrees of freedom,⁵⁵ mean square,⁵⁶ and minimum and maximum value of a set of data. One-way ANOVA also includes Post Hoc Tests. Since Post Hoc Tests perform null hypothesis tests by providing multiple comparisons between different factors, which are layers in the case of this research, correction of level of significance has to be made. One of the methods is to divide α by the number of comparisons in order to control for the type 1 error (the value of α). This method is known as the Bonferroni correction.⁵⁷ SPSS (Statistical Package for the Social Sciences) was used as the software to perform ANOVA. Since four samples (surround, layer 1, 2 and 3) were involved, one-way ANOVA was used in order to examine and compare means of several samples simultaneously.

Pilot test

A pilot listening test was conducted in advance of the listening test in order to examine the experimental design. After the pilot test, some problems were fixed, and a few adjustments were made to the questionnaire and the listening test design. For

⁵⁰ Significance is a statistical term that tells how possible difference or relationship exists.

⁵¹ Standard deviation measures the amount of dispersion/spreading of a set of data.

⁵² Confidence interval is a range of values that works as estimates of the unknown parameter. It indicates the probability of the range that represents the true parameter by giving a distribution of samples. It does not describe any single sample.

⁵³ F-ratio is the ratio of two mean square values. If a hypothesis is true, it is expected that F has a value close to 1.0 most of the time. A large F ratio indicates that the variation among group means is more than what is expected.

⁵⁴ The sum of squares represents a measure of variation or deviation from the mean.

⁵⁵ Degree of freedom is a measure of how many values can vary in a statistical calculation.

⁵⁶ In ANOVA, mean square refers to an estimate of the population variance based on the variability among a given set of measures.

⁵⁷ Field, *Discovering Statistics Using SPSS*, 373.

example, the original time slot for each individual subject was fifteen minutes. However, it was discovered that the participants' perception of music changed significantly as the orchestration of the music changes over time. It was decided that there would be no time limit for each subject so more accurate data could be obtained; all questions were condensed into a single page instead of five pages, which were used in the original design. Therefore, participants could re-examine and change their answers to any question without need for page changes.

Surround Recording with Height Channels Implementation

The first movement of Trompeten-Sextett, Op.30 (Brass Sextet) by Oskar Böhme, performed by *The Bridge Brass*, was recorded in the Recital Hall at the University of Lethbridge on March 5, 2015. The setup is shown in Figure 3. Four AKG C451s were used to capture signal for the left, right, surround left and surround right channel. One Oktava MK-012-01 cardioid was used to record the signal for the centre channel. The left and the right microphones were two meters apart, slightly angled pointing at the edges of the ensemble, and the surround left and surround right microphones were three meters from each other, slightly angled facing the back for lateral and back reflection capture. The distance between the front and the rear microphones was three metres. Different layers of height microphones with different distances from the lower layer were directly positioned above left, right, left surround and right surround.

All the height microphones were pointed at the ceiling with a purpose of suppressing direct sound and capturing ceiling reflections. The whole microphone system was three metres away from the ensemble. The spacings from the lower layer were 0.2m, 0.9m and 1.8m (Figure 4). The distances between the layers were increased to make the

differences more obvious. The same gain settings were used for all the microphones. The session was recorded with a sampling rate of 44.1 kHz and a bit depth of 24 bit using two RME Fireface 800 audio interfaces with four microphone inputs on each unit and two RME OctaMic with eight inputs on each unit. In addition, each instrument was recorded with a microphone at close distance. To prevent bias, they were only used for the post-production of the music, not for the listening test.

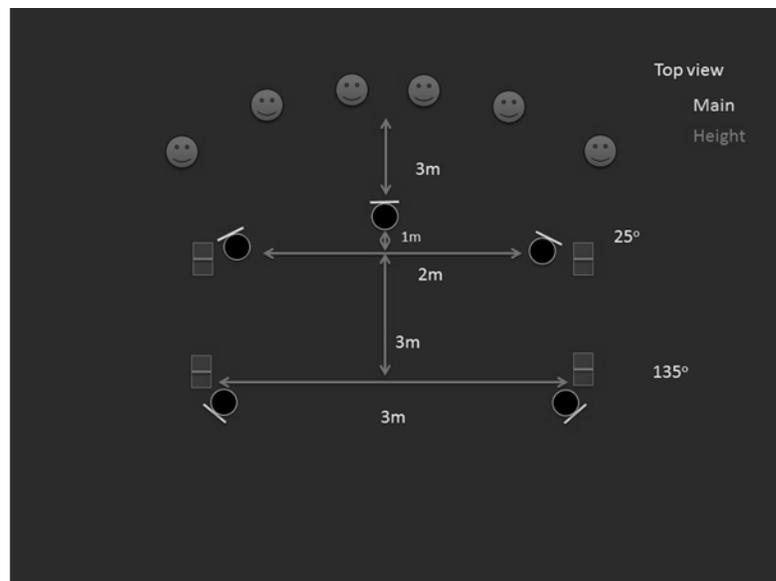


Figure 3, Top view of the recording setup

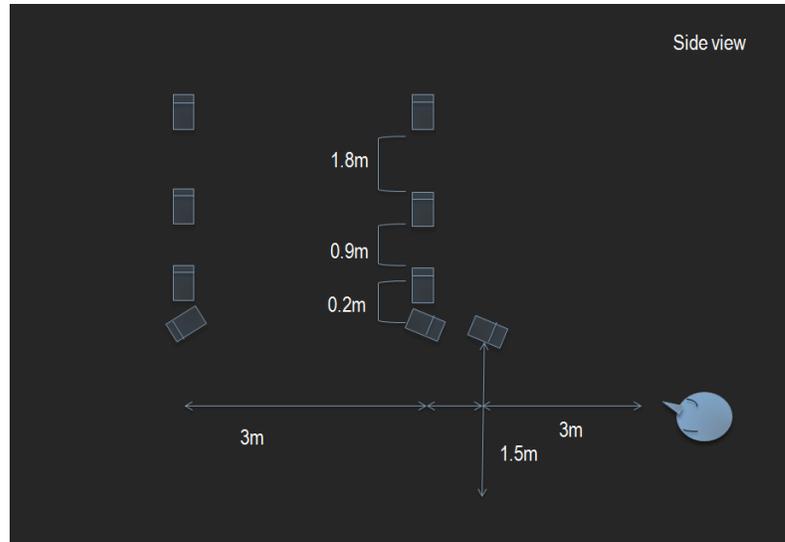


Figure 4, Side view of the recording setup

Excerpts from different takes of the recording were assembled according to musical quality in order to make sure that the listening material with the best performance was presented. No mixing was involved in order to avoid human influence.

Listening Tests Implementation

The questionnaire (see Appendix I) was an online survey using Google Forms. During the test, five samples of the recording were played to each participant. Each participant was asked to rate all the attributes of each sample based on:

- a) Localization focus: stability of instruments.
- b) Depth perception: perceived depth of auditory scene. For example, the distance between the musicians sitting in the front row and the ones sitting in the back row.
- c) Spatial impression: the impression/concept of the naturalness of an actual or simulated space to a listener. For example, the type and size of the room, tonal

balance of reverberation (if the reverberation contains too much high frequency content, it might sound excessively bright), etc.

- d) Envelopment: the feeling of being surrounded/enveloped by sound.
- e) Overall timbre: the general sound colour of the instruments/full ensemble.

All the attributes were rated on a scale of 1 to 10 with 1 being not effective and 10 being very effective. For depth perception, 1 represents not obvious and 10 represents very obvious. The last step of the questionnaire was a dropdown-selection question asking the participants which sample they preferred the most. Any information of each sample was not revealed to the subjects at any time during the test in order to avoid bias. Each sample was two minutes long, but the participants were free to switch to any sample at any time.

The same attributes for each sample were grouped into the same section of the questionnaire as shown in Figure 5. The purpose of this arrangement was to improve the effectiveness of the test and relevance of the questions by asking the participants to focus and rate each attribute of all the samples before moving to the next attribute. All the attributes were mandatory to rate. Considering the fact that the testing participants had to sit in the sweet spot, and they were allowed to switch to any sample and compare different sound attributes of different height microphone setups, the listening test sessions were done with one listening participant at a time.

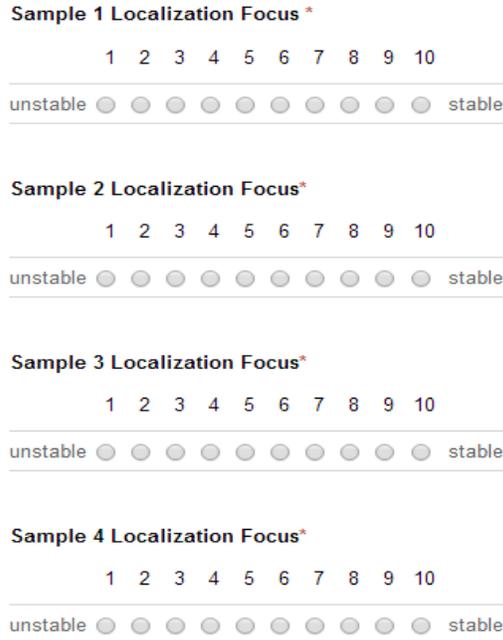


Figure 5, Sample of the listening questionnaire

The series of sets was randomized as shown below. In addition, a letter from A to H was assigned to each of the eight sets for reference purpose. Number 1 refers to layer 1 height information; number 2 refers to the second layer; number 3 refers to the third layer; number 4 refers to the surround layer:

- Set1: 1,4,3,4,2 A
- Set2: 2,4,3,1,2 B
- Set3: 4,2,1,3,1 C
- Set4: 3,3,1,4,2 D
- Set5: 2,2,1,4,3 E
- Set6: 3,1,3,2,4 F
- Set7: 3,2,4,1,2 G
- Set8: 1,3,1,4,2 H

Two RME Fireface 800 were used as audio interfaces for the listening test. All signal routing was done in RME TotalMix, which also served the purpose of letting the subject choose the audio signal. To avoid confusion and bias, all the windows of presets were resized and saved in a way that all the inputs and software returns were hidden and only the preset numbers were revealed (Figure 6).



Figure 6, The interface subjects worked with

The listening tests took place in the control room of Studio One at the University of Lethbridge. Nine Genelec 8040A speakers were used and arranged in a 9.1 playback format without the low frequency channel according to the recommendations by Auro Technologies⁵⁸. However, due to the fact that the ceiling of the studio is not high enough, the height speakers were lower than the recommendation. All the speakers were calibrated to 75 dB SPL (C-weighting) with pink noise.⁵⁹

The participants were seated in the sweet spot. They then received instructions for the experiment. They gave informed consent for their participation of the test. In order to avoid bias, they were asked not to reveal any information about the listening test to the next subject. The listening test had twenty-one participants.

⁵⁸ Theile and Wittek, "Principles in Surround Recordings with Height", 3.

⁵⁹ Frequency weightings are used to allow the sound level meter or noise dosimeter to measure and report noise levels that represent what we hear according to the equal loudness contours. 'C' Weighting is a standard weighting of the audible frequencies commonly used for the measurement of Peak Sound Pressure level

RESULTS

Localization focus

ANOVA revealed that change of distances between low and height channels had no significant influence on participants' localization perception [$F(3,101) = 1.781, p = 0.156$]. However, the mean of L2 is slightly higher than the rest of the layers with the smallest standard deviation. LS received the lowest rating.

Depth perception

ANOVA revealed that change of distances between low and height channels had significant influence on participants' depth perception [$F(3,101) = 3.923, p = 0.011$]. The descriptives of the results (Table 3) show that depth perception on L1 is significantly higher than other layers with a mean of 7.630, and the smallest standard deviation of 1.597. The α value was corrected to 0.017 according to the Bonferroni correction.⁶⁰ From the results shown in the multiple comparisons table (Table 4), it is clear that rating of L1 is very different from the ratings of other layers (Sig. of 0.007, 0.008 and 0.004). There is no significant difference between L2, L3 and LS in terms of depth perception. LS received the lowest rating but the difference between the lowest mean and the second lowest one is very small. The means plots demonstrate this fact (Figure 7).

⁶⁰ There are six comparisons: L1-L2, L1-L3, L1-Surround, L2-L3, L2-Surround and L3-Surround. Therefore the level of significance had to be divided by 6 ($\frac{0.1}{6} \approx 0.017$) in order to control type I error.

Table 3, Descriptives of depth perception

	Mean	Std. Deviation
L1	7.630	1.597
L2	6.210	1.969
L3	6.220	1.987
LS	6.040	2.121

Table 4, Multiple Comparisons of different factors of analysis of depth perception

layer	layer	Sig.
L1	L2	0.007
	L3	0.008
	LS	0.004
L2	L1	0.007
	L3	0.988
	LS	0.753
L3	L1	0.008
	L2	0.988
	LS	0.744
LS	L1	0.004
	L2	0.753
	L3	0.744

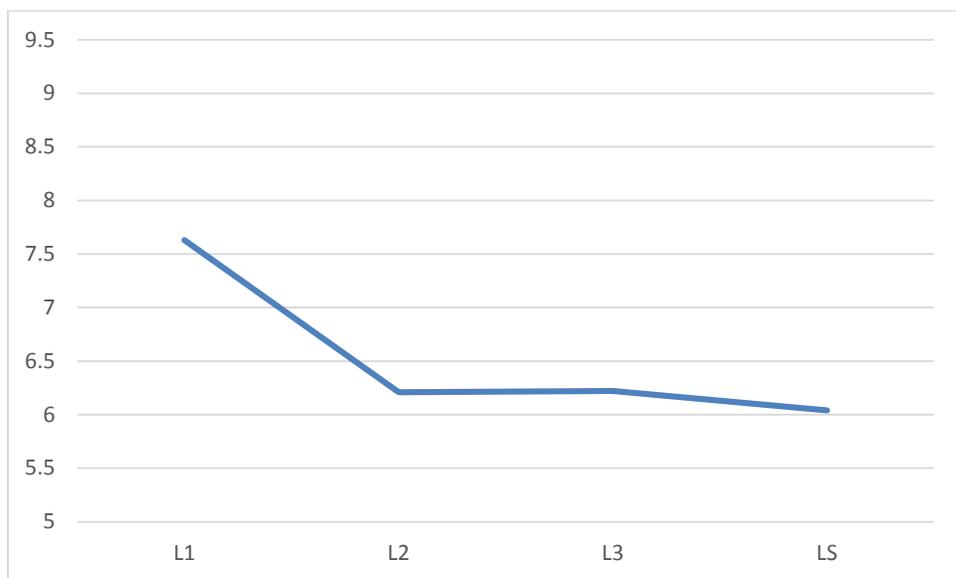


Figure 7, Means plots of depth perception

Spatial impression

ANOVA revealed that change of distances between low and height channels had no significant influence on participants' spatial impression [$F(3,101) = 0.520, p = 0.669$]. The table of the descriptives of the spatial impression (Table 5) lists the statistical parameters of each individual layer's rating. The means and the standard deviations are all very similar.

Table 5, Descriptives of spatial impression

	Mean	Std. Deviation
L1	7.040	2.066
L2	7.210	1.707
L3	6.810	1.688
LS	6.610	1.901

Envelopment

ANOVA revealed that change of distances between low and height channels had significant influence on participants' envelopment perception [$F(3,101) = 22.890, p = 0.000$]. The descriptive results (Table 6) show that the difference between the mean of L1 (the highest rating) and L2 (the second highest rating) is over 2. The difference between L1 and LS (the lowest rating) is almost 4. The value of mean and the standard deviation suggest the participants strongly favoured the first layer in terms of envelopment perception. Additionally, the values of mean indicate that when height channels are included, as the distance between the low and height channels increased, participants felt progressively less enveloped. This fact is reflected in Figure 8.

Table 6, Descriptives of envelopment

	Mean	Std. Deviation
L1	8.930	1.107
L2	6.640	1.967
L3	6.070	1.796
LS	5.000	2.023

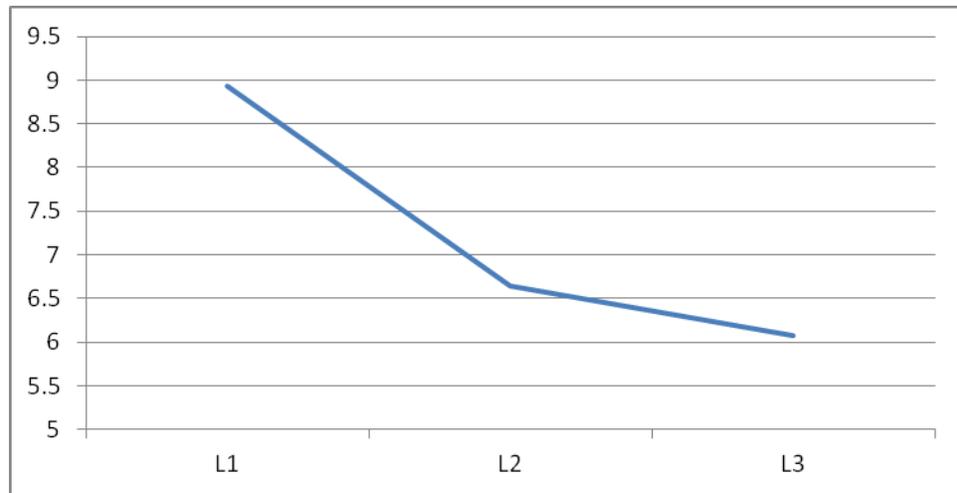


Figure 8, Means plots of envelopment

Multiple comparison shows: 1. L1 has all three positive mean differences (2.283, 2.852, 3.926) compared to the other three layers. 2. Rating of L2 is slightly higher than L3, but it is not significantly different. 3. LS received the lowest average rating, and it is significantly lower than the other three layers.

Overall timbre

ANOVA revealed that change of distances between low and height channels had no significant influence on participants' perception of overall timbre [$F(3,101) = 1.688, p = 0.174$].

Table 7, Descriptives of overall timbre

	Mean	Std. Deviation
L1	7.700	1.836
L2	7.430	1.136
L3	6.890	1.867
LS	6.780	1.953

The descriptives of overall timbre are shown in Table 7. The mean, the standard deviation, and the standard error, are very close between all the layers, even the maximum ratings are identical (10). Once more LS received the lowest rating, but the differences between LS and the other layers are not significant. The means plots (Figure 9) show a trend that is similar to the perception of envelopment (Figure 8) with smaller deviation.

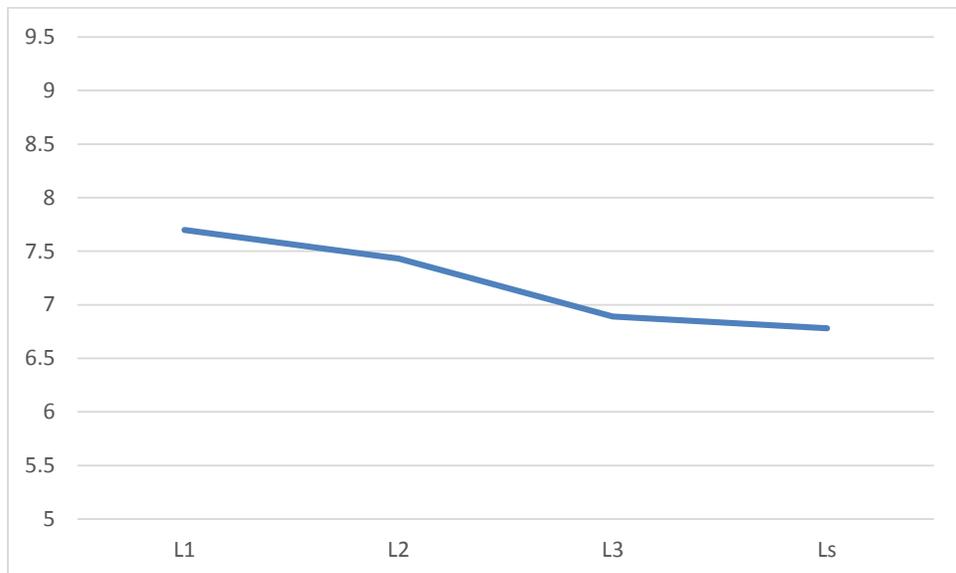


Figure 9, Means plots of overall timbre

Overall preference

The last question was intended to find out which layer the participants preferred the most. The results are shown in Figure 10 below. The numbers indicate the number of participants who voted for the particular layer:

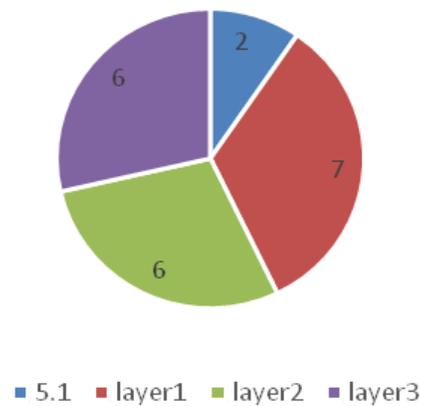


Figure 10, Pie chart analysis of preference

The pie chart does not show a clear difference among the samples with height channels. 33% of the participants favoured L1. Almost 30% of them selected L2 and 30% of them selected L3. The surround sample is the least favourite one with less than 10% of the voting.

DISCUSSION

Localization focus

The results of analysis of the localization focus did not suggest that there was a connection between the perception of localization focus and the increasing distance of the

height channels. The definition of localization focus given by the instruction was the stability of instruments in the stereo field. The localizations of the instruments were mostly provided by the level and time difference of the signal captured by the left-centre-right microphones of the surround configuration, and all the participants were sitting at the sweet spot.

Despite the fact that the results were not significant, L2 and LS received the highest and the lowest rating. This could be caused by the highly-directive characteristics of brass instruments. Different instruments project sound energy differently based on their structure and how they are held by musicians. For instance, near 500Hz, due to its large cone shape and its sound pressure radiating mainly toward the sides in the frequency region, the tuba's output measures approximately 10dB higher on the sides than it does in the front and back. At 800Hz the level of the direction perpendicular to the bell axis is about 20dB lower than the level of the direction of the bell axis. At 2000Hz this difference rises to 28dB.⁶¹ The statistical directivity factor⁶² also indicates the characteristics of the tuba. It has a value of 1.45 at 125Hz for the direction of strongest sound radiation. However, it rises to 2.0 at 400Hz, 4.5 at 1000Hz, and 6.6 at 2000Hz.⁶³ As the notes they play go higher, the sound intensity the brass instruments radiate becomes more focused along the direction where the bells are pointing at, but sound intensity at other directions attenuates rapidly (Figure 11). As a result, the direct signals captured by microphones that the horns were generally pointing at were higher than the signals captured by the rest of the microphones. Based on the setup of the recording

⁶¹ Jürgen Meyer, *Acoustics and Performance of music* (New York: Springer, 2009),136.

⁶² Directivity factor specifies by how much the sound level is higher in the direction considered than it would be for an omnidirectionally radiating sound source of equal power. The larger the factor is, the more directly the sound radiates.

⁶³ J. Meyer, *Acoustics and Performance of music*, 45.

session (Figure 12), the space between L1 and L2 was generally the area where the bells of the tuba and the euphonium were aiming at. Consequently, horns sounded more present or direct in some samples of L1 and L2 than the samples of the other layers at a few places during the recording, and therefore L1 and L2 being more direct might have led most participants to believe that L2 provided the most focused phantom sound sources. In addition, directional characteristics influenced two factors that were related to sound source localization. Localization focus of some samples of L1 and L2 decreased when the dynamic changed from *forte* to *piano* and it slightly increased when the dynamic changed from *piano* to *forte*, such as m.15 (Example 1) and m.112 (Example2); phantom sound source localizations of some instruments moved up and down when switching between different layers. The localization focus influenced by the changes in dynamics might have contributed to the non-significant results. In addition, due to the fact that L1 and L2 had a higher ratio of the direct sound to reverberant sound (D/R ratio), perception of localization focus influenced by changing between *forte* and *piano* was easier to be heard in L1 and L2 than in L3 and LS, and therefore it might have contributed to the results that L2 was the top-ranking layer and L1 received the lowest rating with the highest standard deviation among all the height layers. The change of the phantom sound source caused by directional characteristics might have been one of the reasons that LS received the lowest rating. The displacement of phantom sound sources was subtle when switching between L1, L2 and L3, but it was noticeable when LS was chosen. Therefore the participants might have believed that the localization of LS sounded unstable compared to the rest of the layers.

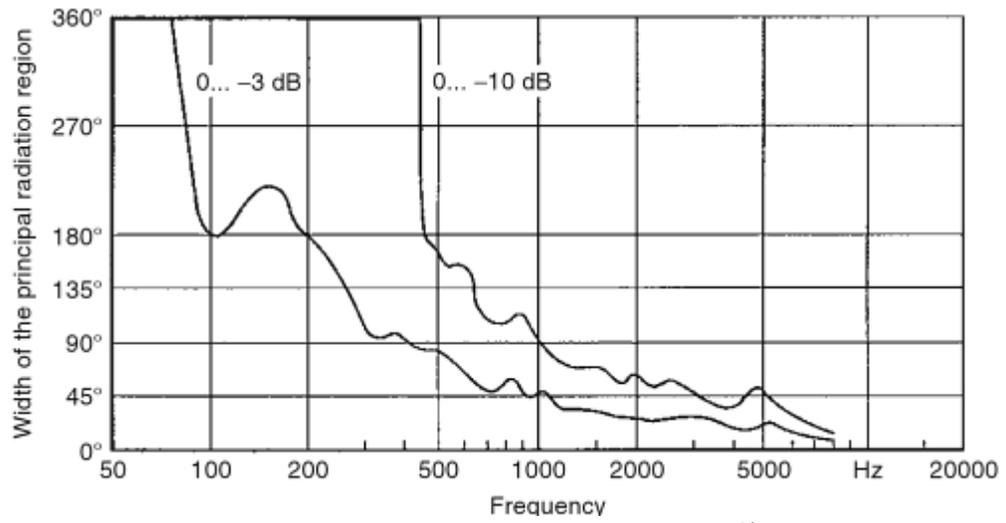


Figure 11, Principle radiation region of a tuba⁶⁴

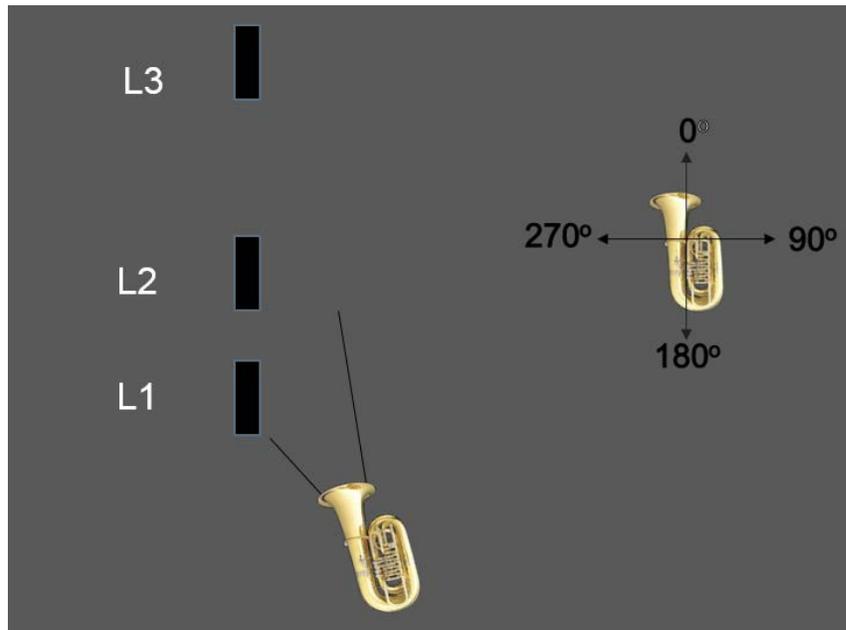


Figure 12 Setup of the recording session

⁶⁴ ibid, 137.

Examples 1, Oskar Böhme, Trompeten-Sextett, Op.30, i, mm 15-16.



Examples 2, Oskar Böhme, Trompeten-Sextett, Op.30, i, mm 112-114.



Depth perception

The results of depth perception suggest that L1 sounded larger in depth to the participants than the rest of the layers, and L2, L3 and LS sounded quite similar, or the differences between those three layers are so small that they can be neglected. The reason for this might also be related to directive characteristics of brass instruments. The D/R ratio was higher along the direction of principle radiation of brass instruments than other

directions. As a result, the microphones in L1 and L2 captured louder and more direct signal of the tuba and the euphonium than the microphones in the rest of the layers because of the fact that the French horn was aiming the bell towards the back of the stage and the trumpets were aiming downwards, making the tuba and the euphonium sound more direct than the French horn and the trumpets. A louder and more direct sound source is always perceived closer to a listener than the one that is quieter and less direct. As a result, in some of the samples of L1 and L2, the tuba and the euphonium appeared to be closer to the listener whereas the French horn and the trumpets sounded further away, which might have led the participants to believe that L1 had a larger auditory scene.

Spatial impression

The results showed that change of distance between the low and height channels had no significant effect on perception of spatial impression. Reverberation time is an essential criterion for the acoustical quality of a concert hall,⁶⁵ and therefore the instruction of the questionnaire asked the participants to use reverberation time and tonal balance of reverberation to help evaluate spatial impression. However, reverberation time does not change for different microphone positions. Furthermore, the reverberation time of the Recital Hall is fairly short, which made it difficult for the participants to evaluate the tonal balance of reverberation. The short reverberation time also made it difficult to hear the potential differences of the tonal balance of reverberation that were captured by microphones in different layers.

⁶⁵ J. Meyer, *Acoustics and Performance of music*, 203.

Envelopment

David Griesinger and his team conducted a series of experiments on envelopment using pink noise, speech, music, and sonic background. Their research indicated that the perception of envelopment was closely related to reverberation, and it was the lateral or diffuse component of the reverberation that mattered.⁶⁶ Their research identified two separate impressions that were related to envelopment: Continuous Spatial Impression (CSI) and Background Spatial Impression (BSI). The perception of envelopment is usually a combination of CSI and BSI. “CSI happens when lateral reflected energy interferes with a continuous sound source.”⁶⁷ It depends on the D/R ratio but it is not related to the loudness of the sound source. The perception of envelopment decreases as the D/R ratio increases. The perception of BSI does not depend on the D/R ratio. It depends on the energy of reverberant signal (especially late reflections that are more than 160ms after the direct sound), and thus the higher the level of a reverberant signal is, the greater perception of BSI will be.⁶⁸ However, the results of envelopment perception of this study contradicted Griesinger’s research. Because of directional characteristics, the D/R ratio of L1 and L2 was higher than L3, and the level of reverberant signal of L3 was higher than L1 and L2. Therefore L3 should have been the most enveloping layer and L1 should have been the least enveloping one of all the height layers. The definition of envelopment given by the instruction was the feeling of being surrounded/enveloped by sound. However, envelopment is difficult to define, and it is sometimes confused with spaciousness or room impression. Envelopment is the sense of immersivity and

⁶⁶ David Griesinger, “Spaciousness and envelopment in musical acoustics”, *AES 101th Convention* (1996): 2.

⁶⁷ David Griesinger, “Objective Measures of Spaciousness and Envelopment”, *AES 16th International Conference* (1999): 32.

⁶⁸ *Ibid*, 33.

involvement in a reverberant sound field, and it is regarded as a positive quality that is experienced in good concert halls.⁶⁹ Therefore, the definition given by the questionnaire presents difficulties when it comes to reproduced sound, especially in the event of artificial sound stage/sound sources, or in this case, listeners' perception being greatly influenced by directional characteristics. This is because sound sources could sometimes be arranged or the perception of the sound sources could be influenced such that direct sounds appear to be enveloping the listener, rather than reverberant sound. Since the D/R ratio of L1 and L2 was higher than L3, the participants might have felt less enveloped by the direct sounds as the height of height channels increased.

Overall timbre

This entire listening test was based on participants' subjective perception, and timbre might be the most subjective attribute. This could have been one of the reasons that the results suggested no significance. Additionally, tonal balance of reverberation influences timbre, yet the short reverberation time of the Recital Hall made it difficult to distinguish the possible timbre difference between the layers. Despite the non-significant results of overall timbre, the ratings on this attribute slightly favoured L1, and the ratings slightly decreased as the distance between low and height channels increased. The signal captured by microphones in L1 and L2 has higher D/R ratio than the one in L3 because of the directional characteristics of the brass instruments and the distances of the layers from the sound sources. Consequently, as the height increased, the level of direct signal became subordinate and reverberant signal became dominant. In addition, the loudness of L3 was slightly lower than L1 and L2. Loudness substantially affects auditory perception.

⁶⁹ Francis Rumsey, *Spatial Audio*, (Waltham: Focal Press, 2001), 38.

The equal loudness contours (Figure 13) reveal that perceived loudness varies greatly with frequency and sound-pressure level.⁷⁰ Therefore the directivity difference of the signal and the minor loudness difference between the layers could be the reason the participants gave L3 the lowest rating of all the height layers.

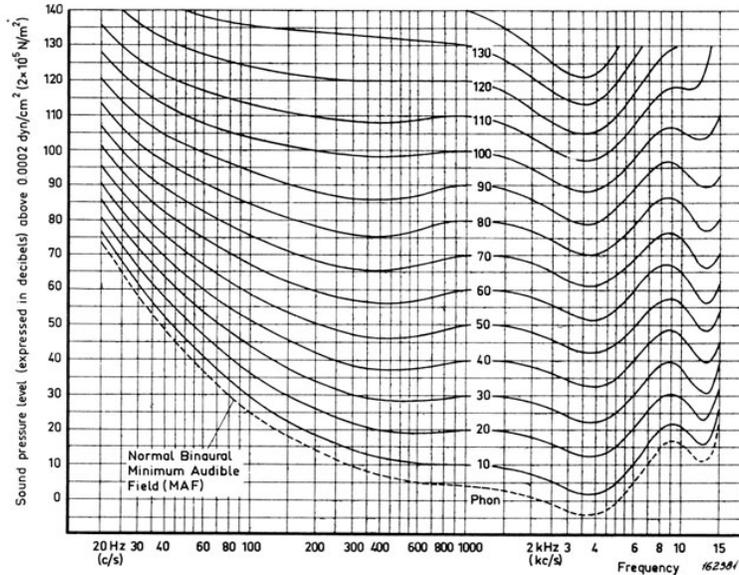


Figure 13, Equal loudness contours⁷¹

Overall preference

The loudness bias against LS made it the least favourite in terms of overall preference compared to the other layers. As the equal loudness contours indicate, if an audio stimulus is played at a relatively low level, the hearing mechanism will not be able

⁷⁰ Everest and Pohlmann, *Master Handbook of Acoustics*, 47. A sound-pressure level of 30 dB yields a loudness level of 30 phons at 1 kHz, but it requires a sound-pressure level of 58 dB more to sound equally loud at 20 Hz. The ear is less sensitive to bass notes than mid-band notes at low levels. This bass deficiency of the ear means that the quality of reproduced music depends on the volume-control setting. Listening to background music at low levels requires a different frequency response than listening at higher levels.

⁷¹ "Equal loudness contours," accessed May 5, 2015, <http://www.roger-russell.com/equalizers/equalizers.html>.

to interpret the low and high end of the frequency spectrum very well, leading to an unsatisfactory or unfavoured impression about the stimulus. Therefore having the participants compare just LS and the layers with heights was not a fair comparison. Furthermore, the results of overall preference show that demonstrating three-dimensional playback systems or microphone setups simply by switching the height channels on and off is not trustworthy. Any three-dimensional playback system or microphone setup would be considered superior than surround systems or microphone setups because of the loudness bias.

The participants did not show a strong preference between L1, L2 and L3. This might have been caused by various reasons: 1. The distances between different layers were increased but they might not be enough for the participants to distinguish. 2. The microphones in all three height layers were directional microphones pointing towards the ceiling, away from the sound sources. That is to say, the height microphones mainly captured reflections from the ceiling instead of direct signal. 3. Insufficient number of participants. In order to get more reliable data, this pool has to be enlarged substantially. This matter is addressed later under limitations of the research. 4. The acoustics of the Recital Hall, especially the short reverberation time might have made it difficult for the participants to hear the differences between the layers. 5. Directional characteristics of the brass ensemble might have played a significant role in the results.

Timbre influences by orchestration

The orchestration of Trompeten-Sextett, Op.30 might have affected the participants' judgement even though the results of the listening test do not show this influence. The perception of timbre slightly changes as the orchestration develops.

Trompeten-Sextett, Op.30 opens with a “somber introduction in chordal texture that serves as preface to the main theme.”⁷² The second subject is a variant of the main theme, given in harmonies by the trumpets and French horn, and the development section is based on an arpeggiated motive derived from the main theme. The chordal introduction of the piece sounds mellow and smooth in this recording. When the development section begins, not only does the sonority become slightly harsh, but the ensemble sounds closer to the audience. Change of timbre introduced by orchestration added another level of subjectivity into the listening test and therefore it might have been one of the reasons the results of overall timbre and overall preference were not significant. Additionally, the higher D/R ratio made the timbre change caused by orchestration more obvious in L1 and L2 than the rest of the layers, and the results of overall timbre slightly favoured L1 and L2. Perhaps some of the participants rated overall timbre during the development section and they preferred the timbre change of the brass instruments and the feeling of being closer to the ensemble.

Deviations caused by the playback order

Theoretically speaking, the same participant’s perception on the same audio samples should have been identical. However, the participants’ responses on the randomly duplicated sample stored in one sample set were greatly influenced by the presenting order of the samples despite the fact that the participants were given the freedom to switch to any sample at any time of the listening test. This might be due to the fact that, based on observation, all the participants were strictly following the sequential

⁷² Program notes by Dr. Richard E. Rodda, Chicago Chamber Musicians, accessed June 20, 2015, <http://www.encoreccm.org/pieces/287>

playback order during the first time of evaluation. If a replicated sample immediately followed the original one, all the participants were able to identify the duplicated one and provided the same rating as the original sample; if a duplicated sample was not immediately present after the original sample, most of the participants were unable to provide the same rating. Some of the ratings on two identical samples provided by the same participant had significant deviations. This might be one of the explanations for the high standard deviations of some of the ratings.

Limitations of the test

Insufficient number of participants

Hedonic experiments reveal what participants' preferences are. The results of these kinds of tests are usually related to many aspects. For example, some subjects might prefer chamber music while some might not. Their judgement is also influenced by the positions in which they sit, and their moods. Additionally, preference is very difficult to measure since it is not a concrete or definite concept. The results of this type of test could be influenced by various factors. For instance, preferences generated by hedonic tests need to be transformed using methods, such as converting scales into numbers for data analysis. The problem with this kind of measurement is that preference retains its subjectivity after the transformation. Taking the "depth perception" as an example, we might be confident that a participant who gave a rating of ten to L3 found its sound "larger" than those who gave a rating of five, but we cannot be certain that the first participant found L3 two times "larger" than the second participant. Similarly, we cannot be sure that two participants who gave a rating of five found the same sample equally

deep. However, the rating intervals of depth perception and envelopment suggested clear preference. Participants' ranges of ratings/preference were different. Some of them kept all of their ratings in a relatively small range (e.g., from 4 to 8) while some of them always rated the five samples from 1 to 10. For depth perception, 52% of the participants had smaller ranges of preference, and they all gave L1 the highest rating; for envelopment, 48% of them had smaller ranges of rating, and they ranked L1 as the top layer. Nevertheless, the participants' ratings depended on their subjective feelings. Therefore in order for hedonic tests to be effective, for subjectivity to be kept to a minimum, and for the survey results to be representative, it is critically important that the tests have a large number of randomly-selected participants in each group. However, with only twenty-one participants partaking in this research, the margin of error is fairly high with a value of roughly 23%.⁷³ That is to say, more participants would have helped improve the research.

Phrasing issues

The description of depth perception instructed the participants to use the distance between the musicians sitting in the front row and the ones sitting in the back row to evaluate the depth of the auditory scene. However, the brass ensemble only had one row and therefore this explanation was misleading and ambiguous. Defining the depth as the

⁷³ "Sample Size: How Many Survey Participants Do I Need?" http://www.sciencebuddies.org/science-fair-projects/project_ideas/Soc_participants.shtml. The margin of error is a statistic expressing the amount of random sampling error in a survey's results. It asserts a likelihood (not a certainty) that the result from a sample is close to the number one would get if the whole population had been queried. The likelihood of a result being "within the margin of error" is itself a probability, commonly 95%, though other values are sometimes used. The larger the margin of error, the less confidence one should have that the poll's reported results are close to the true figures; that is, the figures for the whole population. Margin of error applies whenever a population is incompletely sampled.

perceptual distance between the closest and farthest instruments might have yielded different results.

The definition of envelopment did not ask the participants to evaluate the perception based on envelopment in a reverberant sound field, and therefore most of the participants might have rated this attribute according to the feeling of being enveloped by sound in general.

CONCLUSION

Under some circumstances (i.e., most classical recording projects), the goal of recording is to deliver an experience that is consistent with classical music listeners' listening habits and expectations. Surround recording with height information definitely offers audio engineers an additional way of achieving this goal by utilizing the upper hemisphere for the reproduction of ceiling reflections and diffusion. In order to provide better spatial impression, envelopment, and timbre with those four speakers, capturing reflections from the upper hemisphere plays a crucial role, and how high the height microphones should be raised is an important factor in this process. This study conducted a surround recording with different spacings between the low and height channels and used a listening test to look into the perception of general quality and spatial impression of different microphone spacings. Based on the twenty-one participants' responses, two out of five attributes—depth perception and envelopment—were perceived as being significantly different, and L1 received the highest rating of all the layers on both attributes.

Practically speaking, any recording of any format is a unique situation. Different concert halls have different reverberation times and absorption; different instruments have different directional characteristics and timbres; listeners prefer various genres of music. That is to say, had the recording taken place in another location, had the ensemble been different (e.g., percussion ensemble), or had the genre of music been different, the results of the listening test might have varied substantially. This study suggests that recording engineers should take directive characteristics of instruments into consideration not only in stereo and surround recording, but in three-dimensional situations as well because now directional characteristics affect perception of spatial depth, spatial impression, envelopment, timbre, localization focus of phantom sound sources and other parameters in three-dimensions instead of two-dimensions in stereo or surround recording. Since additional microphones are involved in the case of surround recordings with height channels, the results of this study indicate that directional characteristics of instruments might create unsatisfactory perception of a small music ensemble by stretching or distorting the spatial depth between different instruments. Directional characteristics could also create undesired localization of phantom sound sources. For example, the localization of the instruments with their principle radiation direction facing the height channels might be perceived from higher positions versus those instruments with different directions of radiations. Directional characteristics might introduce problems to surround recordings with height channels, but it could also give recording engineers ideas or solutions to work around the situation when considering microphone setups for this kind of recording. For example, if undesired spatial depth occurs, engineers could balance it by adding spot microphone(s) for the instruments that do not project their

major sound energy at the height microphones or repositioning the main microphone system. If undesired phantom sound source localization occurs, engineers need to change the placement or the angles of the height microphones in order to offset the perception of height differences between the instruments. However, due to the existence of loudness bias of the listening material, an insufficient number of participants in the listening test, the phrasing issue of the questionnaire, the complexity of human natural hearing, and psychoacoustic phenomena, it is not entirely clear how relevant the testing result and spatially realistic resolution for height channels is.

For future study and research, according to what has been examined and discovered, a few points are worth taking into consideration. 1. It would be beneficial to conduct listening tests that involve different groups of listeners. For example, listeners who are experienced in audio engineering and listeners who are not. Different groups of listeners could be tested by either the same or different factors/attributes that are specifically designed for each different group. If the same attributes are used to test two groups of participants who have different levels of background knowledge about audio and ear training, researchers will be able to see if the responses given by different groups reflect the differences, or how important ear training and previous experience are in preference tests. If different factors/attributes are given to the two groups, researchers can get feedback from both professional and consumer level. 2. The listening tests took place in the control room of Studio One, which does not have enough height for the height speakers, and therefore it might have had some influence on the perception of the height channels. If a study on a similar topic is conducted, the results may be more accurate or closer to the purpose of study if the listening tests are carried out in spaces that have

proper heights, or studios dedicated to three-dimensional music production. 3. This study used a closely-spaced microphone setup. It might be interesting to try largely-spaced setups and investigate the relations between perception and different heights between low and height channels. 4. Since directional characteristics greatly influence perception, it is necessary to avoid placing height microphones at the principle radiation direction of the instruments in order to minimize their influence, which would allow participants to fully focus on the attributes. 5. Recording different genres of music or different types of ensembles and testing listeners' preference on different height channels will definitely widen the scope of microphone techniques for surround recordings with height channels and make the techniques more practically relevant.

BIBLIOGRAPHY

- Alten, Stanley R. *Audio in Media: The Recording Studio*. New York: Schirmer, 1995.
- Barbour, Jim. "Exploration of the Height Dimension in Audio Reproduction." *Australasian computer music conference, Wellington, New Zealand*. (2004).
- Beins, Bernard C. *e-Study Guide for Research Methods and Statistics*. Cram101, 2014.
- Brain, Marshall. How stuff works? "How Analog and Digital Recording Works." <http://electronics.howstuffworks.com/analog-digital3.htm>
- Bruck, Jerry. the late Al Grundy, and Irv Joel Audio. "History Timeline." <http://www.aes.org/aeshc/docs/audio.history.timeline.html>
- Dolby Laboratories Inc. "5.1-Channel Music Production Guidelines", 2003.
- Elsea, Peter. "ANALOG RECORDING OF SOUND", University of California, Santa Cruz, 1996, http://artsites.ucsc.edu/EMS/music/tech_background/te-19/teces_19.html
- Everest, F. Alton., and Ken C. Pohlmann. *Master Handbook of Acoustics*. New York: McGraw-Hill, 2009.
- Field, Andy. *Discovering Statistics using IBM SPSS Statistics*. New York: SAGE Publications Ltd, 2013.
- Fred R Byers. *Care and handling of CDs and DVDs: A guide for librarians and archivists*. Gaithersburg, Maryland: National Institute of Standards and Technology, 2003.
- Griesinger, David. "Spaciousness and envelopment in musical acoustics", *AES 101th Convention* (1996).
- Kim, Sungyoung, Doyuen Ko, Aparna Nagendra, and Wieslaw Woszczyk. "Subjective Evaluation of Multichannel Sound with Surround-Height Channels." *AES Convention: 135*, 9003 (October 2013).
- Lee, Hyunkook, and Christopher Gribben. "On the Optimum Microphone Array Configuration for Height Channels." *AES Convention: 134*, 93 (May 2013).
- Meyer, Jürgen. *Acoustics and the Performance of Music*. New York: Springer, 2009.
- Rabinowitz, Phil and Stephen B. Fawcett. "Evaluating Community Programs and Initiatives." <http://ctb.ku.edu/en/table-of-contents/evaluate/evaluate-community-interventions/collect-analyze-data/main>

- Research randomizer. <https://www.randomizer.org/>
- Rumsey, Francis. *Spatial Audio*. Waltham: Focal Press, 2001.
- Dr. Richard E. Rodda, Chicago Chamber Musicians. "Program notes."
<http://www.encoreccm.org/pieces/287>.
- Sackett, David L. "Bias in analytic research." *Journal of Clinical Epidemiology*. Volume 32, Issues 1-2 (1979): 51–63.
- Saxby, M.J. *Food Taints and Off-Flavours*. New York: Springer, 2012.
- Science Buddies. "Sample Size: How Many Survey Participants Do I Need?"
http://www.sciencebuddies.org/science-fair-projects/project_ideas/Soc_participants.shtml
- Silzle, Andreas, Stefan Geyersberger, Gerd Brohasga, Dieter Weninger, and Michael Leistner. "Vision and Technique Behind the New Studios and Listening Rooms of the Fraunhofer IIS Audio Laboratory." *AES Convention*:126, 7672 (May 2009).
- Suresh, KP. "An overview of randomization techniques: An unbiased assessment of outcome in clinical research." *Journal of Human Reproductive Sciences*.4(1) (2011): 8–11.
- Symes, Colin. *Setting the Record Straight: A Material History of Classical Recording*. Connecticut: Wesleyan, 2004.
- Theile, Günther. "Natural 5.1 Music Recording Based on Psychoacoustic Principles." *AES Conference:19th International Conference: Surround Sound - Techniques, Technology, and Perception*, 1904 (2001).
- Theile, Günther, and Helmut Wittek. "Principles in Surround Recordings with Height." *AES Convention*:130, 8403 (2011).
- Theile, Günther. "Multichannel Natural Music Recording Based on Psychoacoustic Principles." *AES Convention*:108, 5156 (2000).
- Tingay, James. "What are A, C & Z Frequency Weightings?", NoiseNews, 2011.
<http://www.cirrusresearch.co.uk/blog/2011/08/what-are-a-c-z-frequency-weightings>.
- Vercammen, Marc., and Jan Maes. *Digital Audio Technology: A Guide to CD, MiniDisc, SACD, DVD(A), MP3 and DAT*. Waltham, Massachusetts: Focal Press, 2013.
- Zielinski, Slawomir, Francis Rumsey, and Søren Bech. "On Some Biases Encountered in Modern Audio Quality Listening Tests-A Review." *AES Journal* Volume 56 Number 6 (2008).

APPENDIX I: QUESTIONNAIRE

In this listening test, five samples of a three-dimensional recording will be played. You will first be asked to grade all the attributes of each sample on a scale from 1 to 10, and then compare the three samples with each other. 1 being not effective at all and 10 being very effective. The duration of each sample is about 2 minutes, but you are free to switch to any sample at any time. The scale and all the attributes are explained below.

You may withdraw from the test at any time without reason by simply closing your browser without submitting your responses. If you do this, all information from you will be destroyed. Your email address, names or ID number will not be recorded or used at any stage of the research. All collected data will be stored on the researcher's Google Drive, secured in a password-protected account. Thank you for participating.

Example of scale

1 2 3 4 5 6 7 8 9 10

Frontal Image: Sound image perceived from the front.

- **Localization focus:** stability of instruments.
- **Depth perception:** perceived depth of auditory scene. For example, the distance between the musicians sitting in the front row and the ones sitting in the back row.

Spatial Impression: the impression/concept of the naturalness of an actual or simulated space to a listener. For example, the type and size of the room, tonal balance of reverberation (if the reverberation has too much high frequency content, it might sound excessively bright), etc.

Envelopment: the feeling of being surround/enveloped by sound.

Overall timbre: the general sound colour of the instruments.

Frontal Image

Sample 1 Localization Focus*

1 2 3 4 5 6 7 8 9 10

unstable stable

Sample 2 Localization Focus*

1 2 3 4 5 6 7 8 9 10

unstable stable

Sample 3 Localization Focus*

1 2 3 4 5 6 7 8 9 10

unstable stable

Sample 4 Localization Focus*

1 2 3 4 5 6 7 8 9 10

unstable stable

Sample 5 Localization Focus*

1 2 3 4 5 6 7 8 9 10

unstable stable

Sample 1 Depth Perception*

1 2 3 4 5 6 7 8 9 10

small large

Sample 2 Depth Perception*

1 2 3 4 5 6 7 8 9 10

small large

Sample 3 Depth Perception*

1 2 3 4 5 6 7 8 9 10

small large

Sample 4 Depth Perception*

1 2 3 4 5 6 7 8 9 10

small large

Sample 5 Depth Perception*

1 2 3 4 5 6 7 8 9 10

small large

Sample 1 Spatial Impression*

1 2 3 4 5 6 7 8 9 10

unnatural natural

Sample 2 Spatial Impression*

1 2 3 4 5 6 7 8 9 10

unnatural natural

Sample 3 Spatial Impression*

1 2 3 4 5 6 7 8 9 10

unnatural natural

Sample 4 Spatial Impression*

1 2 3 4 5 6 7 8 9 10

unnatural natural

Sample 5 Spatial Impression*

1 2 3 4 5 6 7 8 9 10

unnatural natural

Sample 1 Envelopment*

1 2 3 4 5 6 7 8 9 10

unenveloped enveloped

Sample 2 Envelopment*

1 2 3 4 5 6 7 8 9 10

unenveloped enveloped

Sample 3 Envelopment*

1 2 3 4 5 6 7 8 9 10

unenveloped enveloped

Sample 4 Envelopment*

1 2 3 4 5 6 7 8 9 10

unenveloped enveloped

Sample 5 Envelopment*

1 2 3 4 5 6 7 8 9 10

unenveloped enveloped

Sample 1 Overall Timbre*

1 2 3 4 5 6 7 8 9 10
unpleasant pleasant

Sample 2 Overall Timbre*

1 2 3 4 5 6 7 8 9 10
unpleasant pleasant

Sample 3 Overall Timbre*

1 2 3 4 5 6 7 8 9 10
unpleasant pleasant

Sample 4 Overall Timbre*

1 2 3 4 5 6 7 8 9 10
unpleasant pleasant

Sample 5 Overall Timbre*

1 2 3 4 5 6 7 8 9 10
unpleasant pleasant

Please choose a sample that you prefer the most.*

APPENDIX II: LETTER OF CONSENT

HUMAN SUBJECT RESEARCH

Letter of Consent

February 12, 2015

Dear Research Participant:

You are being invited to participate in a listening test to determine the optimum range of height microphones for three-dimensional recordings.

This test will require about 10 minutes of your time. During this time, you will be asked to listen to three different chamber music recordings. The test will be conducted in Studio One.

There are no anticipated risks or discomforts related to this research. The playback volume will be carefully calibrated before the tests to make sure that the sound volume is below the level at which hearing damage could occur.

Although there are no direct benefits to you from participating in this study, you may help me identify factors that will improve the audio production for three dimensional recordings.

Several steps will be taken to protect your anonymity and identity. At the beginning of each test, you will be asked to log into your Uleth email account in order to access the questionnaire. You will be reminded to log out once the test is done. Your name and email address will not be collected at any stage of the research. All collected data will be stored on the researcher's password-protected Google Drive. All information will be destroyed after 5 years time.

Your participation in this research is completely voluntary. However, you may withdraw from the test at any time for any reason. If you do this, all information from you will be destroyed.

The results from this study will be presented in writing in the researcher's thesis. If you wish to receive a summary of the results from this study, you may contact the researcher at the email address given below.

If you require any information about this study, or would like to contact the researcher, please email (*Yangke Li*) at (*yangke.li@uleth.ca*) at the University of Lethbridge. If you have any other questions regarding your rights as a participant in this research, you may also contact the Office of Research Ethics at the University of Lethbridge at 403-329-2747 or research.services@uleth.ca.

By proceeding to the questionnaire, your consent to participate is implied. Thank you.