SAE Institute London

Major Project BSc

Capturing 3D Audio:

A pilot study on the spatial and timbral auditory perception of 3D recordings using main-array and front-rear separation in diffuse field conditions.

Student Full Name: Janet Grab
Student Number: 95477
Course Code: CMN6301
Assessment Code: CMN6300.1
Qualification: BSc (Hons) Audio Production
Project Supervisor: Tommaso Perego
Date of Submission: 12 April 2019
Word Count: 8332

Declaration

I hereby declare that I wrote this written essay on my own and without the use of any other than the cited sources and tools and all explanations that I copied directly or in their sense are marked as such, as well as that the essay has not yet been handed in neither in this nor in equal form at any other official commission.

…………………………...12 April 2019, London.

Abstract

This research is a preliminary pilot experiment into the subjectively perceived differences between the recordings resulting from three different 3D microphone arrays: A Bowles-Array with a vertical coincident height-channel microphone layer, a Fukada-Tree/Hamasaki-Cube configuration and a Hybrid-Array containing the signals from the Bowles-Array main layer and the Hamasaki-Cube height layer. It was hypothesised that the arrays in concern will produce recordings that shall lead each to an increased perception of specific attributes for all sources tested (cello, violin, handpan, djembe, guitar). In order to detect possible patterns in spatial and timbral auditory perception subjective listening tests included direct scale magnitude estimations for the attributes Naturalness, Presence, Preference, Width, Localisation Accuracy, Distance/Depth, Envelopment, Spatial Balance, Room Perception, Vertical Image Shift, Vertical Image Spread, and Vertical Frequency Separation, and category scaling for the assessment of timbral attributes. Results suggest that none of the arrays in concern conveyed an increased perception of any of the attributes for all sources, which disproves the hypothesis and indicates a source-dependent performance. Simultaneously patterns in the subject responses have been detected which could be explained through psychoacoustic findings focussing on the correlation of perception between the attributes in question. Furthermore, by trying to explain the obtained differences in auditory perception between the different arrays, some assumptions could be made upon what components of which array could have contributed to a specific perception. These findings could serve as a reference for future experiments in the fields of 3D recording techniques or psychoacoustics.

Acknowledgements

I would like to express my appreciation to Tony Faulkner for his practical advice and to Hyunkook Lee for his further explanations on his research.

Thanks to Antonio de Robertis and Tommaso Perego for their academic support throughout the project and Romain Zemiri from SAE Brussels for giving me access to the Auro-3D studio.

Thanks to Bernadette Contini for her assistance in the recording session and Saint Lewis-Green for his visual documentation.

And finally, many thanks to Gerard Day, Carola Krebs, and Mark Seow for their playing in the recording.

Table of Contents

List of Tables

List of Figures

Abbreviations

APE:	Acoustic Pressure Equaliser
ASW:	Apparent Source Width
DF:	Diffuse field
D/R ratio:	Direct-to-Reverberant ratio
GUI:	Graphical User Interface
HF:	High Frequency
IAC:	Interaural Cross-Correlation
ICC:	Interchannel Cross-Correlation
ICCC:	Interchannel Cross-Correlation Coefficient
ICCT:	Interchannel Cross Talk
ICLD:	Interchannel Level Difference
ICTD:	Interchannel Time Difference
IR:	Impulse Response
L:	Left
LCR:	Left-Centre-Right
LEV:	Listener Envelopment
LF:	Low Frequency
LS:	Left Surround
R:	Right
RS:	Right Surround
RT60:	Reverberation Time
SRA:	Stereo Recording Angle
S/R ratio:	Signal-to-Noise Ratio
VIS:	Vertical Image Spread

1. Introduction

1.1 Rationale, Significance and Aims of the Research

The importance of 3D recording techniques comes into play within the framework of a quickly growing 3D audio market aiming to improve the natural sound experience (Transparency Market Research, 2018). The ability to do so was confirmed by Hamasaki & Van Baelen as *"the three-dimensional multichannel sound system [implying 3D recording techniques] can produce better sensations of spatial sound quality, reality, and presence than the two-dimensional and the one-dimensional multichannel sound systems."* (2015, p. 6) This was confirmed by Howie *et al.* stating that *"the addition of height channels allows the recording engineer to enhance the presentation by improving the depth, presence, envelopment, naturalness, and intensity of the recordings."* (2015, p. 2)

However, as *"3D audio has gained traction only recently"* (Transparency Market Research, 2018) the field of 3D recording yet offers many things to explore. This was affirmed by research papers openly raising the lack of literature dealing with 3D recording techniques, such as Werner *et al.* declaring that "*there is hardly any literature on 3D audio [recording]. Therefore, close collaboration with experts who have practical experience in this field and are ready to share their thoughts is essential."* (2014, p. 1) The same publications appealed to develop new 3D recording techniques through experimentation (Bowles, 2015; Hamasaki & Van Baelen, 2015; Theile & Wittek, 2012) as *"capturing audio in three dimensions is becoming a required skill for many recording engineers."* (King *et al.*, 2016, p. 1) To allow for further insight into the field of 3D recordings, many studies have been published which investigate into the psychoacoustic influence of one parameter such as vertical microphone spacing (Lee & Gribben, 2014) or polar patterns (Howie *et al.*, 2015) on the perception of specific attributes. Comparisons of 3D microphone techniques for music recordings have also been made, as in Riitano & Victoria (2018), Howie *et al.* (2016) or Ryaboy (2015), amongst others. However, to the author's knowledge, only few comparisons have been conducted between the recordings of a 3D main microphone array system and a system entailing front-rear separation, as outlined in Howie *et al.* (2018), Riaz *et al.* (2017), or the proceedings of the ICSA 2011 (2011).

At the same time, the need to do so was declared by Hamasaki & Van Baelen with the suggestion to use a Hamasaki-Cube in the F/R configuration (2015, pp. 4 and 7). The necessity of this comparison seems further justified through previous results of Kassier *et al.* (2005), who found a high listener preference in the Fukada-Tree / Hamasaki-Square configuration when comparing different combinations of front-rear arrays for surround sound. Based on this, some of the latest findings in psychoacoustics considering height-channels, and the evidence that some established recording engineers almost exclusively use the approach of a Main-Array (Lindberg, 2015), it was decided to compare a Bowles-Array entailing a vertical coincident height-channel microphone layer (Main-Array) with a Fukada-Tree/Hamasaki-Cube configuration (F/R-Array). A direct comparison between these techniques has not been made to the present according to the author's knowledge. Following the previously mentioned appeal for experimentation, a hybrid version containing the signals of the Bowles-Array for the main layer and the signals of the Hamasaki-Cube for the height layer was included in the comparison.

This comparison is a preliminary pilot experiment investigating the differences between the recordings of these microphone arrays regarding their spatial and timbral auditory perception. The attributes derived to distinguish possible differences were Naturalness, Presence, Preference, Width, Localisation Accuracy, Distance/Depth, Envelopment, Spatial Balance, Room Perception, Vertical Image Shift, Vertical Image Spread, Vertical Frequency Separation and Timbre. The response format of the subjective listening tests was a direct scale magnitude estimation, whereas category scaling was applied to establish perceptual differences in Timbre. In addition, referring to Berg & Rumsey (2003, p. 2), objective measures, such as dummy head recordings, IRs, RT60 and spectral analysers have been included in the analysis of the obtained listening test results to gain further insight into the acoustics of the recording space or as an objective reference for human hearing.

The current investigation aims to gain further insight into a possible existence of patterns amongst the listener responses for the different spatial and timbral attributes and the recordings of the microphone arrays in concern. Since each of these arrays operates based on different psychoacoustic principles (see Appendix Chapters 9.1.1, 9.1.2, 9.2 and 9.5) it could be assumed that if a pattern can be established its origin will lie in the different array types. Or, in other words, that each array will produce recordings which will lead to an increased perception of specific attributes, for all sources (cello, violin, handpan, djembe, guitar). These findings can be considered significant as they might indicate what arrays or what parts of the arrays might have contributed to a certain auditory perception. This insight, on the other hand, could serve as a basis of a *"perceptual 'handbook'",* as proposed by Williams (2011, p. 5). Its aim is to ease the control of certain parameters to create a specific spatial or timbral perception in a 3D recording scenario;

> *"Since true identity [referring to the recording space or phantom images] is rarely possible or desirable, some means of creating and controlling adequate illusions of the most important subjective cues for consumer enjoyment could be held as the primary aim of recording and reproducing techniques."* (Rumsey as in Roginska & Geluso, 2018, p. 216)

Therefore, the outcomes of this investigation could further inform recording engineers experimenting with 3D recording techniques about a suitable approach according to the circumstances of a given recording situation, or a desired aesthetic. Furthermore, as filling a gap in literature with one of the first comparisons of this kind to the author's knowledge, the report of the different experimental techniques and the discussion of their outcomes could serve as a possible basis for further psychoacoustic research and new experimental approaches by other engineers. Indirectly, the study also reaches out to 3D audio consumers, which closes the circle to the importance of the 3D audio market.

1.2 Hypothesis and Research Questions

<u>Hypothesis</u>

The arrays in concern will produce recordings that shall lead each to an increased perception of specific attributes for all sources tested.

<u>Research Questions</u>

1. In the context of a subjective listening test, can patterns in auditory perception be detected between the recordings of a Bowles-Array with a vertical coincident height-channel microphone layer, a Fukada-Tree/Hamasaki-Cube configuration and a Hybrid-Array containing the signals of the Bowles-Array main layer and the Hamasaki-Cube height layer?

2. If so, what is the nature of these patterns and can they be explained by previous psychoacoustic research or by objective measures such as IRs, dummy head recordings or RT60 using spectral analysers?

3. How do these patterns possibly relate to the different arrays as a whole or some of their components?

1.3 Structure of the Thesis

This section provides an overview of the individual chapters and their content throughout the paper.

Introduction:

Summarises the necessity, impact and aims of the current investigation, states the hypothesis and research questions and provides an overview of the thesis.

Contextualisation:

Provides further background on the hypothesis and research questions, and some operating principles in 3D audio recording and reproduction.

Methodology:

Outlines and justifies the methods used in the listening test design and stimuli creation. A detailed justification of the recording setup and process, as well as the gathering of objective measures, can be found in the Appendix. Furthermore, the limitations of the pilot experiment are outlined.

Results:

Visualises, describes and summarises the results of the subjective listening tests in relation to the research questions and hypothesis.

Discussion:

Confronts the results of the subjective listening tests with previous psychoacoustic research and objective measures to provide possible explanations for the outcomes in relation to the research questions and hypothesis.

Summary and Conclusions:

Draws general conclusions about the outcomes of this investigation and suggests further research.

Appendices:

Provide further background information about the principles of stereophonic microphone arrays and psychoacoustics, a justification of the chosen arrays, the chosen microphones and their placement, a detailed overview of the recording process and further information about objective measures.

2. Contextualisation

This Chapter gives a background on the 3D microphone arrays in question and explains the psychoacoustic principles they are based on to ease the understanding of the discussion in Chapter 5.

2.1 Functions of the Main and the Height Microphone Layer

Within the community of practice, it is commonly agreed that *"the main (lower) microphones are used for source positioning, whilst the height microphones are used to increase spatial impression without affecting localisation."* (Wallis & Lee, 2014, p. 1) This gives rise to *"the need to separate the main and height layers entirely [in the recording technique], and to keep the height channel information specific to sound arriving from above.*" (Bowles, 2015, p. 2) These statements go in line with William's claim that the height layer *"must not generate information that will be in conflict with the localisation cues already generated in the loudspeakers making up the horizontal plane information."* (Williams, 2012, p. 5) Furthermore, it has been genuinely mentioned that a rather large portion of the signal must be present in the height channels to be effective (Lee, 2018a; Martin *et al.*, 2015). The reproduction system used in studies dealing with both, psychoacoustics and recording techniques, such as Lee & Gribben (2014), Howie *et al.*, (2015) or Ryaboy (2015), to name a few, was the Auro-3D 9.1 setup (Figures 2.1.1 and 2.1.2), where the microphone signals from both layers are routed discretely to their respective loudspeakers.

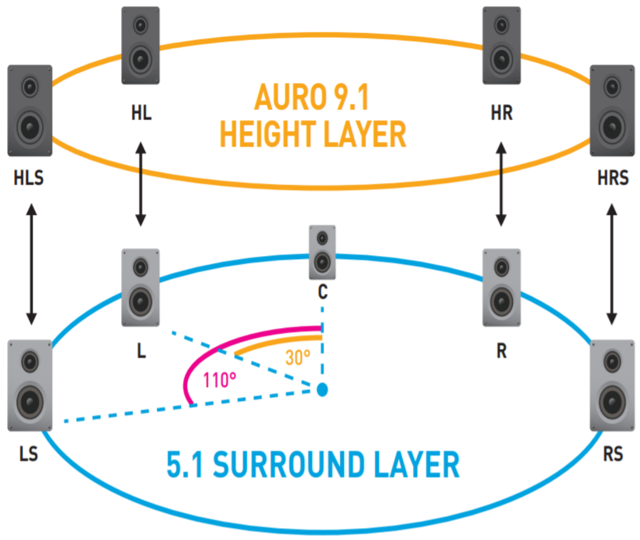


Figure 2.1.1: Auro-3D 9.1 Setup: Basic Layout (Auro-3D® Home Theater Setup Installation Guidelines, 2015 p. 12)

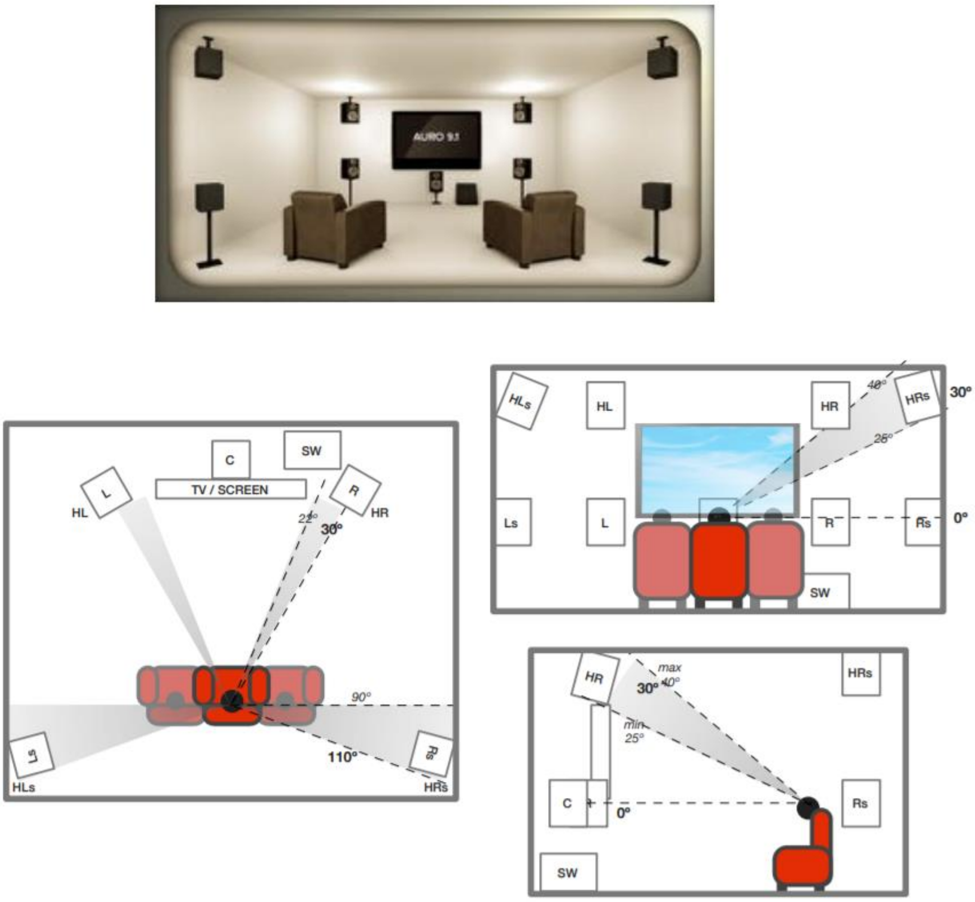


Figure 2.1.2: Auro-3D 9.1 Setup: Side and Top View (*ibid.*)

2.2 3D Main and Front-Rear Arrays

2.2.1 Overview

Originally suggested as a way of classifying microphone techniques for 5.1 surround sound, Rumsey declares the classification of two main approaches, based upon the purpose of the rear channels in the technique; the *"five-channel 'main microphone' arrays"* (2013, p. 190), subsequently referred to as Main-Arrays, and the *"separate treatment of front imaging and ambience"*, subsequently referred to as F/R-Arrays (*ibid.,* p. 196). Figure 2.2.1 provides a direct comparison between a 3D Main-Array (*"Type A"*) and a 3D F/R-Array technique (*"Type B"*):

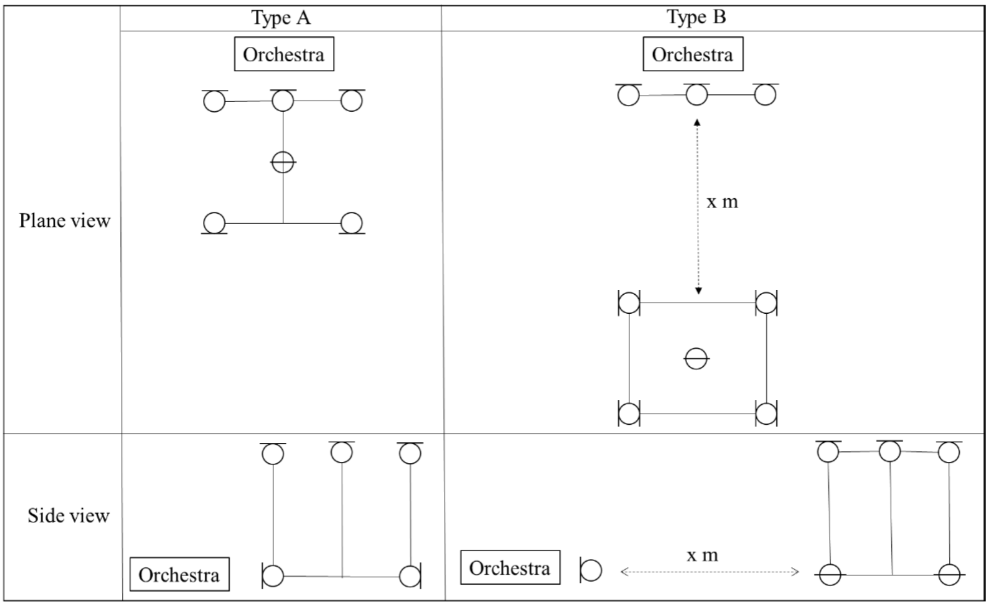


Figure 2.2.1: Main- and F/R-Array (Hamasaki & Van Baelen, 2015, p. 5)

Rumsey defines a Main-Array as *"a single array of microphones in reasonably close proximity to each other… usually based on some theory that attempts to generate phantom images with different degrees of accuracy around the full 360° in the horizontal plane."* (2013, p. 188) Main-Arrays normally consist of a front triplet with two microphones at the back and aim to provide satisfying directional images and spatial impression simultaneously. An example of a Main-Array can be seen in Figure 2.2.2:

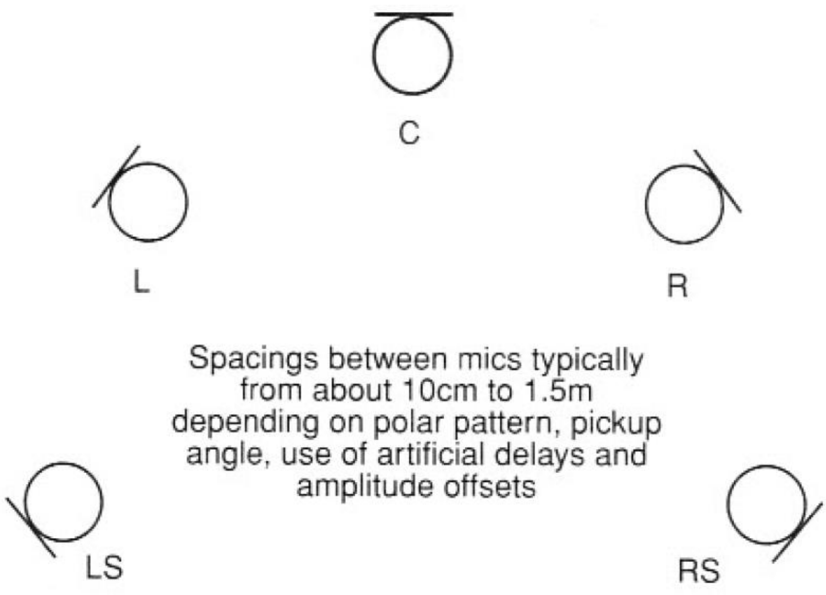


Figure 2.2.2: Generic Layout of Five-Channel Main Microphone Arrays (Rumsey, 2013, p. 191)

The F/R-Array approach, on the other hand, treats the front and rear channels separately and has usually *“a front array providing reasonably accurate phantom images in the front, coupled with a separate means of capturing the ambient sound of the recording space (often for feeding to all channels in carrying degrees).”* (*ibid.,* p. 188) Different front arrays can be combined with different back arrays to achieve the desired image and spatial qualities. A schematic visualisation of the principle of this technique can be seen in Figure 2.2.3 and 2.2.4.

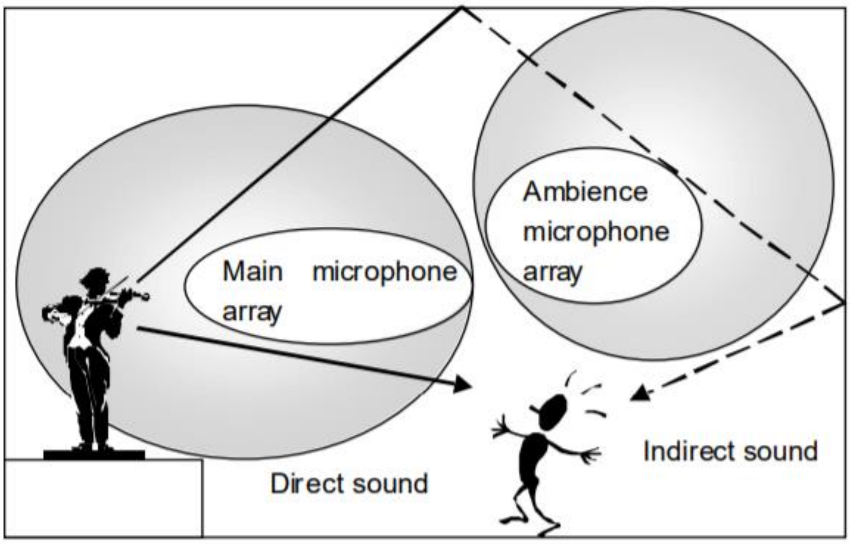


Figure 2.2.3: F/R-Approach (Hamasaki *et al.*, 2001, p. 5)

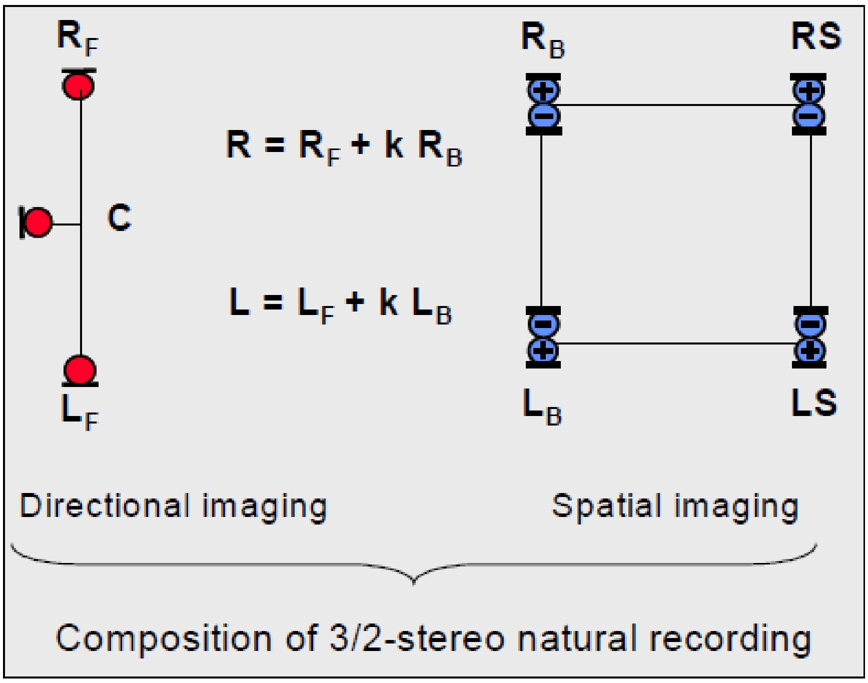


Figure 2.2.4: F/R-Approach with a Three-Channel Front Array for Directional Imaging and an Ambience Array (Hamasaki-Square) for Spatial Imaging (Theile, 2001, p. 22)

It should be noted that the term main array in the context of the F/R-Approach refers to the array optimised to capture direct sound. The main array within the F/R-Approach can be seen in Figure 2.2.5:

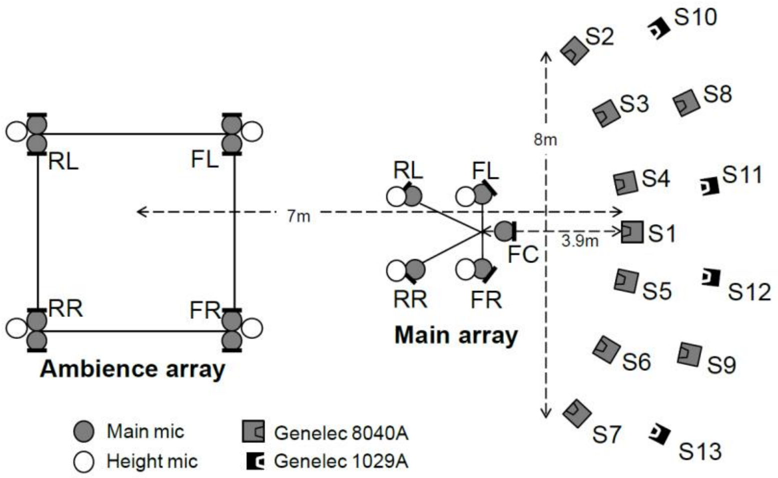


Figure 2.2.5: F/R-Array Approach by Lee & Millns (2017, p. 2)

Furthermore, Rumsey confirms the scarcity of these techniques to provide a separate feed for the LFE (Rumsey, 2013, p. 188). This means that they are in fact 9-channel and not 9.1-channel microphone techniques.

2.2.2 Practical Implications

In the context of the current research, it is worth having a look at the basic principles 3D Main- and F/R-Arrays rely on to understand what the practical implications are.

There are several authors describing advantages of Main-Arrays, for example, King *et al.* who claim that "*staying near the source results in less background noise and a more dynamic, and therefore interesting ambient program."* (2016, p. 2) This approach is used by some well-known engineers such as Lindberg (2018b; 2015), and also Theile & Wittek are aware of its benefits:

> *"The application of a main microphone [sic] appears to be advantageous if suitable recording conditions are given and the correct microphone location can be found to ensure the adequate directional image as well as the adequate balance of direct and indirect sound. This is even more true if naturalness and a 'being there' impression is intended by means of 3D recordings."* (2012, p. 18)

However, the same authors also clearly outline the challenges of the Main-Array approach, as *"the microphone setup has to pick up the direct sound, early reflections, reverberation, and enveloping sources (e.g. applause), and to deliver the complete nine-channel mix which must satisfy with respect to many parameters, such as sound colour, directional imaging, spatial imaging, and envelopment as described above. The parameters are governed by psychoacoustic principles and practical constraints leaving not much room to get everything well in any application. Suitable recording conditions must be given. Thus, the scope of applications for a specific main microphone [array] is limited."* (*ibid.*, p. 17)

These considerations raised the question of whether the approach of an F/R-Array would be generally preferred. Especially the increased possibilities of control speak in its favour:

"And there is not enough time to strictly adjust the position of microphones in such productions. Therefore, it is necessary to have proper microphone techniques to adjust the ratio of direct sound and indirect sound easily to maintain the highest quality of three-dimensional multichannel sound reproduction. In order to solve this issue, a microphone technique of type B [F/R array with a Hamasaki-Cube as ambience array] is developed. This will enable us to control spatial impressions easily and realize the stable sound source localization in the frontal sound field. During a recording, this technique makes it easy to control the ratio of direct sound and indirect sound without any disturbing effects on sound sources localization, because the ambience microphone array catches mainly the indirect sound." (Hamasaki & Van Baelen, 2015, p. 5)

This becomes even more important considering the genuinely agreed aim to achieve the best possible D/R ratio in the field of classical music recording (Howie *et al.*, 2015, p. 2; Lindberg, 2015) as *"reproducing the spatial impression of a reverberant sound field such as a concert hall is one of the principal aims of three-dimensional multichannel audio."* (Hamasaki & Van Baelen, 2015, p. 1)

However, the flexibility of the F/R-Array stands in contrast with the advantage of the Main-Array to convey a strong sense of presence and naturalness. Within the framework of this research, it thus makes sense to aim for a widely-defined comparison providing some further insight into the behaviour of these two opposed techniques, and to experiment with a hybrid version.

2.3 Psychoacoustic Background

This chapter contextualises the chosen arrays with the status quo of psychoacoustic research. For further explanations of psychoacoustic concepts mentioned in this chapter, the reader is referred to the Appendix (Chapter 9.5). To summarise it can be said that all array types are worth comparing because they are based on different psychoacoustic principles.

Firstly, there is the vertical coincident Main-Array, being preferred and perceived as more *"spacious"* than vertically spaced Main-Arrays (Lee & Gribben, 2014, p. 833). The reason for that is assumed on the one hand in a VIS of the vertical coincident array, caused by ICCT not reaching the masking threshold (Wallis & Lee, 2017, p. 17; Lee & Gribben, 2014, p. 879), and on the other hand in only having limited comb-filtering as no ICTD between the main and height microphone layer is at work (Lee & Gribben, 2014, p. 881).

On the contrary, the Main-Array seems to have limitations in adjusting the D/R ratio as the presence of direct sound in the height channels would lead to a vertical phantom image shift once the localisation threshold is exceeded (Wallis & Lee, 2017, p. 2). In that regard, the F/R-Array allows for more freedom in the mixing stage. This is due to the almost exclusive capture of ambient sound by the Hamasaki-Cube (Hamasaki & Van Baelen, 2015, p. 5) which will unlikely lead to vertical ICCT as ICCT conditions a certain amount of direct sound (Wallis & Lee, 2017, p. 1). Consequently, due to the absence of vertical ICCT, the risk of comb-filtering and vertical phantom image shift is reduced (Wallis & Lee, 2017, p. 1; Lee *et al.*, 2014, p. 7).

However, as the F/R-Array most likely will lack ICCT, the chance of a VIS is reduced, and VIS was found to have a positive influence on *"preference."* (Lee & Gribben, 2014, p. 879) In return, the Hamasaki-Cube with its large spacings is known to have a very low ICC, being the reason for its increased *"horizontal image spread."* (Gribben & Lee, 2018, p. 537) In the context of the Hamasaki-Cube, this implicates an increased perception of environmental width. However, as interchannel decorrelation was found to be inefficient in the median plane (*ibid.*), the performance of the F/R-Array with its low vertical ICC should be observed against the performance of the Main-Array with the possible effect of VIS.

When comparing the outcomes of Wallis and Lee (2017, p. 2) with Lee *et al.* (2014, p. 7), it is assumed that both approaches add tonal colouration to the main layer. The nature of these colourations, however, was unknown by that time and further studies dealing with these phenomena have been suggested (Wallis & Lee, 2017, p. 17).

Apart from that, Robotham *et al.* stated that *"early reflections arriving vertically are suggested to have a greater impact on a perception of timbre of reproduced sound [than early reflections with a lateral incidence angle]."* (2016, p. 2) If one combines this with the statement that *"spectral content from a vertical reflection could result in coloration when summed with the direct signal"* (*ibid.*) and their finding that *"timbral auditory sensations were responsible for the overall preference rating"* (*ibid.*, p. 7), the importance of the factor timbre besides the spatial qualities within the current research becomes evident.

3. Methodology

3.1 Listening Test Design

When designing the listening test, the procedure described in “*Quantification of Impression”* (Bech & Zacharov, 2006, pp. 39-96) was taken as a reference. A summary of this process can be seen in Figure 3.1.1 (see next page):

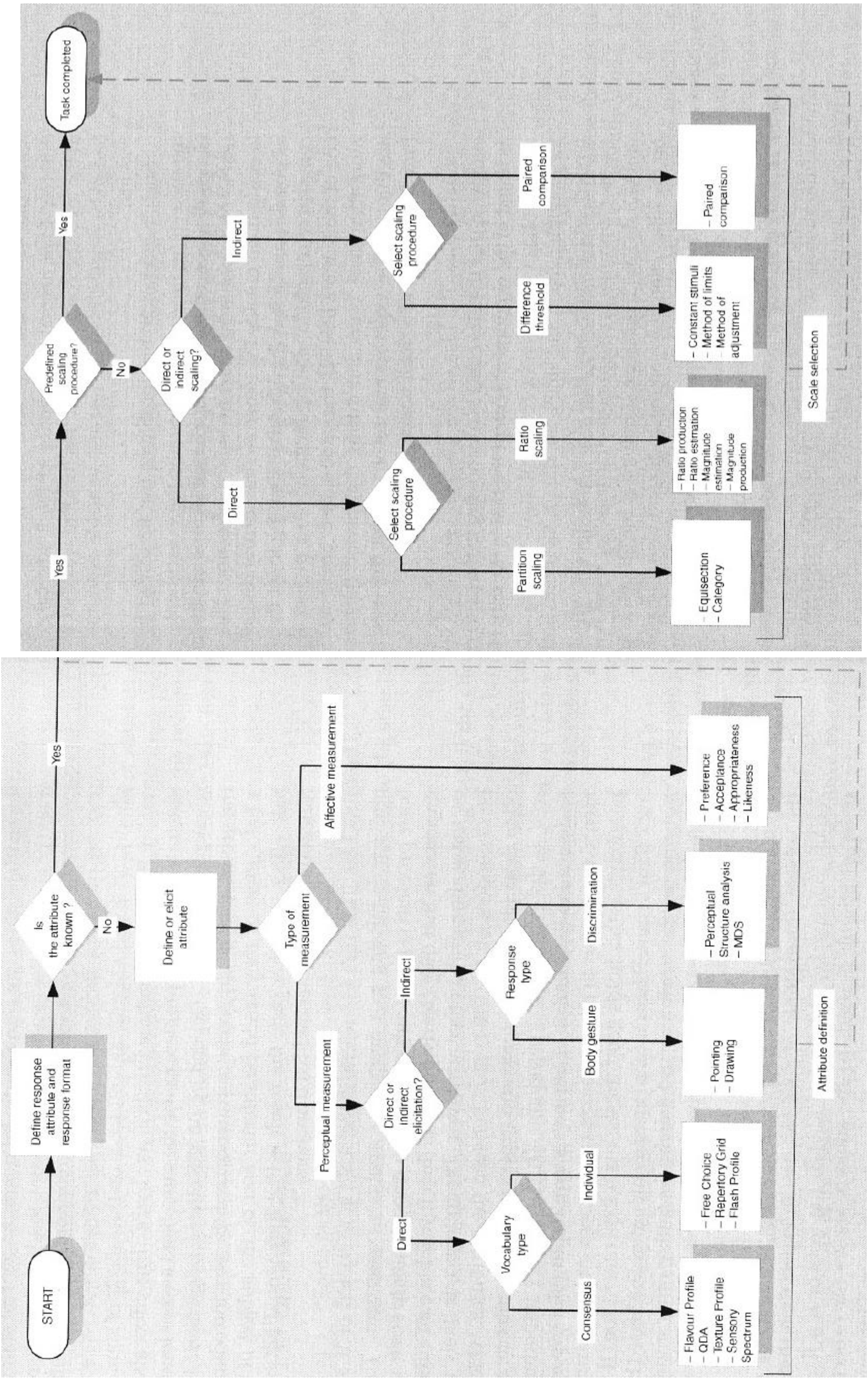


Figure 3.1.1: Process of Identifying the Response Attribute and Selecting the Appropriate Scaling Procedure According to Bech & Zacharov (2006, pp. 94 - 95)

3.1.1 Selection of Attributes

By following the first steps displayed in Figure 3.1.1 Bech & Zacharov state that

> *"if more complex stimuli, like music, are used instead of designed signals, it means that several attributes are likely to be excited, and it thus becomes relevant to first identify the relevant attributes, than to define and formulate the description of the individual attributes and finally evaluate individual attributes in separate experiments."* (2006, p. 40)

In the context of attribute elicitation, Berg & Rumsey defined four main areas contributing to the overall perceived audio quality (also referred to as MOS; mean opinion score), as depicted in Table 3.1.1 (2003, p. 3):

Timbral Quality	Relates to the tone colour or describes *"the sensation whereby a listener can judge that two sounds are dissimilar using other criteria other than pitch, loudness or duration"*, as defined by Pratt and Doak (p. 317, 1976)
Spatial Quality	Relates to the three-dimensional nature of the sound sources and their environments
Technical Quality	Relates to distortion, hiss, hum, and similar
Miscellaneous Quality	Relates to the remaining properties

Table 3.1.1: Four Main Quality Categories Contributing to MOS According to Berg & Rumsey (2003, p. 3)

A schematic representation of this categorisation can be seen in Figure 3.1.2:

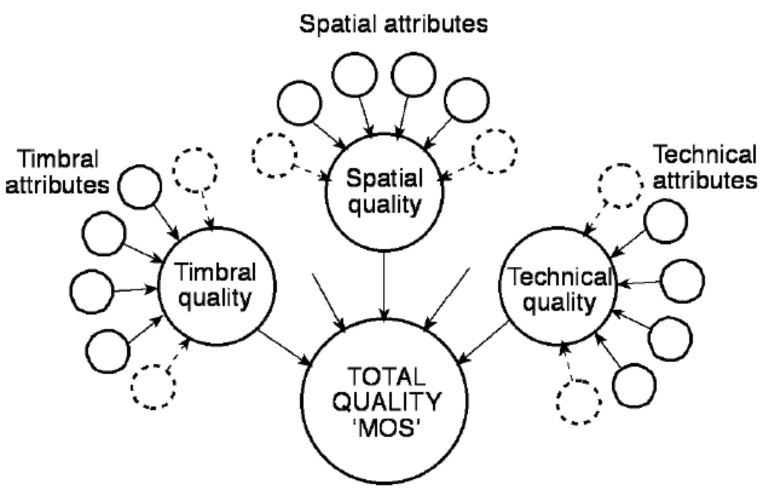


Figure 3.1.2: Relations between Total Audio Quality, its Subsets, and Attributes (Berg & Rumsey, 2003, p. 3)

Several studies have been conducted to identify the quality attributes having the most influence on each, spatial, timbral and technical perception through elicitation of verbal descriptors. This goes along with Berg & Zacharov's procedure, referring to Figure 3.1.1. Tables 3.1.2 and 3.1.3 depict the studies which have been compared against each other regarding their elicited attributes and definitions. A more frequent appearance of a specific attribute throughout these studies was considered to indicate higher importance to describe a spatial or timbral impression adequately.

Selected Studies for Identifying Attributes for Spatial Quality Evaluation	
Author(s) and Year of Publication	**Title of Paper or Journal Article**
Berg & Rumsey (2006)	*Identification of Quality Attributes of Spatial Audio by Repertory Grid Technique*
Berg & Rumsey (2003)	*Systematic Evaluation of Perceived Spatial Quality*
Francombe, Brookes & Mason (2015)	*Elicitation of the Differences between Real and Reproduced Audio*
Gerzon (1971)	*Whither Four Channels*
Rumsey (2002)	*Spatial Quality Evaluation for Reproduced Sound: Terminology, Meaning, and a Scene-Based Paradigm*
Zacharov & Pedersen (2015)	*Spatial Sound Attributes - Development of a Common Lexicon*

Table 3.1.2: Selected Studies for the Selection of Spatial Attributes

Selected Studies for Identifying Attributes for Timbral and Technical Quality Evaluation	
Author(s) and Year of Publication	**Title of Paper or Journal Article**
Howie *et al.* (2018)	*Subjective and Objective Evaluation of 9ch Three-Dimensional Acoustic Music Recording Techniques*
Pedersen & Zacharov (2015)	*The Development of a Sound Wheel for Reproduced Sound*
Robotham, Stephenson and Lee (2017)	*The Effect of a Vertical Reflection on the Relationship between Preference and Perceived Change in Timbral and Spatial Attributes*
Simurra & Queiroz (2017)	*Pilot Experiment on Verbal Attributes Classification of Orchestral Timbres*
Williams (2011)	*A Comparative Perceptual Evaluation of the Timbral Variations in Choral Location Recordings Created by four Common Stereo Microphone Techniques*

Table 3.1.3: Selected Studies for the Selection of Timbral and Technical Attributes

The attributes derived in *Systematic Evaluation of Perceived Spatial Quality* (Berg & Rumsey, 2003), displayed in Figure 3.1.3, have been taken as a reference for the selection of spatial attributes, as this study strongly relates to the context of the current investigation:

> *"The method described has been shown to produce statistically significant results in evaluation of different modes of spatial reproduction and different microphone techniques. Despite changes of subjects and stimuli, the attributes on which the scales are based seem valid and reliable in the context of evaluating the spatial quality of surround sound reproductions of stationary, naturally occurring sound sources in reverberant spaces, recorded acoustically without using artificial multitrack mixing. This reinforces the strength of attributes originating in constructs elicited from listeners. There are no data available for direct comparison to support the superiority of attributes generated this way, but the fact that all attributes have showed [sic] to be highly significant indicates the power of this method… the general validity of the attributes found could be confirmed by comparing them with attributes employed by other authors, like Zacharov and Koivuniemi, Toole and Gabrielsson et al."* (Berg & Rumsey, 2003, p. 11)

This means that these attributes are also in accord with Bech & Zacharov's proposed system displayed in Figure 3.1.1 (see next page):

Attribute	Description
Naturalness	How similar to a natural (i.e. not reproduced through e g loudspeakers) listening experience the sound as a whole sounds.
Presence	The experience of being in the same acoustical environment as the sound source, e g to be in the same room.
Preference	If the sound as a whole pleases you. If you think the sound as a whole sounds good. Try to disregard the *content* of the programme, i e do not assess genre of music or content of speech.
Low frequency content	The level of low frequencies (the bass register).
Ensemble width	The perceived width/broadness of the ensemble, from its left flank to its right flank. The angle occupied by the ensemble. The meaning of "the ensemble" is all of the individual sound sources considered together. Does not necessarily indicate the known size of the source, e g one knows the size of a string quartet in reality, but the task to assess is how wide the sound from the string quartet is perceived. Disregard sounds coming from the sound source's environment, e g reverberation – only assess the width of the sound source.
Individual source width	The perceived width of an individual sound source (an instrument or a voice). The angle occupied by this source. Does not necessarily indicate the known size of such a source, e g one knows the size of a piano in reality, but the task is to assess how wide the sound from the piano is perceived. Disregard sounds coming from the sound source's environment, e g reverberation – only assess the width of the sound source.
Localisation	How easy it is to perceive a distinct location of the source – how easy it is to pinpoint the direction of the sound source. Its opposite is when the source's position is hard to determine – a blurred position.
Source distance	The perceived distance from the listener to the sound source.
Source envelopment	The extent to which the sound source envelops/surrounds/exists around you. The feeling of being surrounded by the sound source. If several sound sources occur in the sound excerpt: assess the sound source perceived to be the most enveloping. Disregard sounds coming from the sound source's environment, e g reverberation – only assess the sound source.
Room width	The width/angle occupied by the sounds coming from the sound source's reflections in the room (the reverberation). Disregard the direct sound from the sound source.
Room size	In cases where you perceive a room/hall, this denotes the relative size of that room.
Room sound level	The level of sounds generated in the room as a result of the sound source's action, e g reverberation – i e not extraneous disturbing sounds. Disregard the direct sound from the sound source.
Room envelopment	The extent to which the sound coming from the sound source's reflections in the room (the reverberation) envelops/surrounds/exists around you – i e not the sound source itself. The feeling of being surrounded by the reflected sound. .

Figure 3.1.3: Attributes in the Final Evaluation Experiment Conducted by Berg & Rumsey (2003, p. 12)

However, some attributes elicited in the other studies were missing in Figure 3.1.3 and thus have been added to the spatial attribute selection. These include Spatial Balance, Vertical Image Shift, Vertical Image Spread, and Vertical Frequency Separation. The same procedure has been applied to the selection of significant timbral and technical attributes. Based on the extensive research of Pedersen & Zacharov (2015) in these fields, their *Sound Wheel for Reproduced Sound* (Figure 3.1.4) served as a primary reference. The definitions therein for the different timbral attributes have been taken for the current investigation:

Loudness		How loud the sound is perceived.
Dynamics	Attack	Transient response. Specifies whether the drum beats and percussion, etc. are accurate and clear i.e. if you can hear the actual strokes from drumstick, the plucking of the strings etc. it is also expressed as the ability to reproduce each audio source transients cleanly and separated from the rest of the sound image. Imprecise Attack is understood as unclear or a muted impact.
	Bass Precision	Are instrument impacts from the bass drum and bass precise, crisp and without distortion, are the impacts tight and well defined? Bass precision may be defined as Attack in the bass region. Imprecise means that the attack spread in time and the peak of the impact is softened.
	Punch	Specifies whether the strokes on drums and bass are reproduced with clout, almost as if you can feel the blow. The ability to effortlessly handle large volume excursions without compression (compression is heard as level variations that are smaller than one would expect from the perceived original sound).
	Powerful	The ability to handle high sound levels, especially when striking the drums and bass. Indicates whether the Punch, Attack and Bass precision are maintained at high volume.
Treble	Treble strength	The relative strength of the treble or high frequencies -Weak: Covered, unsharp. -A little under neutral: A soft sound without being dull. -Neutral: In the middle of the scale, where you can clearly distinguish instruments. A lot: Treble Raised. Sharp, hard sound.
	Brilliance	Treble or high frequency extension: -A little: As if you hear music through a door, muffled, blurred or dull. -A lot: Crystal-clear reproduction extended treble range with airy and open treble. Lightness, purity and clarity with space for instruments. Clarity in the upper frequencies without being sharp or shrill and without distortion.
	Tinny	Resonances or narrowband frequency prominence in the treble or high frequencies.
Midrange	Midrange strength	The relative level of the midrange, i.e. the middle frequencies, e.g. sopranos, trumpets, violins and xylophones. Not to be confused with Canny which represents prominent narrow frequency ranges (resonances) in the midrange.
	Nasal	A closed sound with pronounced midrange. Gives the impression corresponding to vocalists singing through the nose (nasal).
	Canny	The music sounds like it is being played in a can or tube. The sound is characterized by prominent and narrowband resonances in the midrange.
Bass	Bass strength	The relative level of bass, i.e. the low frequencies, for example male voices, bass guitar, bass drum, timpani and tuba. Should not be confused with bass depth that indicates the low frequency bass extension.
	Bass depth	Denotes how far the bass extends downwards. If it goes down in the low end of the spectrum, there is great depth. Should not be confused with Bass strength, which indicates the strength of the bass or Boomy which related to resonances in the lower bass region.
	Boomy	Resonances in the low bass, as sound in a large barrel, which gives a prominent bass resound resounding (reverberating) when bass and bass drums are heard. The representation tends to become muddy and imprecise.
	Boxy	Boxy denotes a hollow sound, as if the sound was played inside a small box. Represents resonances in the upper bass frequency range.
Timbral balance	Dark-Bright	Denotes the balance between bass and treble. -Dark: Excessive bass. Either loud bass or weak treble. -Neutral: Bass and treble are perceived equally loud, there is a balance in the reproduction. This also applies if both bass and treble are equally weak or if the bass and treble are both too loud. If it leads to prominent or soft midrange this is assessed by the Midrange strength. -Bright: Excessive treble. Either loud treble or weak bass. The cause for the sound being dark or light can deduced from the assessments of Bass strength and Treble strength.
	Full	If both low and high frequencies are well represented with good extension the sound is Full.
	Homogeneous	Denotes to which degree the different frequency ranges (bass, midrange and treble) are coherent, continuous, and balanced without gaps between them. That there are seamless transitions between the tone ranges.
Sound image	Distance	The perceived distance between the listener and the main sound sources (instruments / singer). Does it sound as the music is close or far away?
	Width	The width of the sound image (expressed as the perceived angle). - The width of the sound sources positions (soundscape width). The width of any reverberation should not be included in the assessment.
	Depth	The depth of the sound image (i.e., in the direction away from the listener). Not to be confused with distance.
	Balance	Is the soundstage skewed to one side (left or right) or is it centred in the middle?
Localization	Precise	Can the individual instruments and voices be clearly placed and separated in the spatial sound image? How precise are the individual sound sources positioned in the room? If the individual sound sources are inadvertently spread or broadened out the precision is low. Can the individual instruments and voices be clearly placed and separated in the spatial sound image? How precise are the individual sound sources positioned in the room? If the individual sound sources are inadvertently spread or broadened out the precision is low.
	Envelop ment	Are you surrounded by the reproduced sound and does it give a sense of space around you?
	External ization	When listening on headphones: To what extent do you perceive the sound sources outside of your head?
Transparency	Presence	Does it sound as if the sound sources are present and not distant or absent?
	Clean	It is easy to listen into the music, which is clear and distinct. Instruments and vocals are reproduced accurately and distinctly. The opposite of clean: dull, muddy.
	Detailed	A well-resolved sound rich in detail. Instruments, voices etc. can easily be separated. The music has many details, details that cannot be measured, details that give the music "soul". It may be small audible nuances: Breathing from a singer, fingers wandering across the guitar strings, the flaps from the clarinet, embouchure sound of the saxophone, the impact from the piano's hammers when they hit the strings.
	Natural	Sounds reproduced with high fidelity. Acoustic instruments, voices and sounds, sounds like in reality. The sound is similar to the listener's expectation to the original sound without any timbral or spatial coloration or distortion, "Nothing added - nothing missing." The soundstage is clear in space and brings you close to the perceived original sound experience.
Artefacts – Signal related	Shrill	Treble Distortion. Very sharp s-sounds, cymbals etc.
	Rub bing	As the sound of something scraping on a (rough) surface.
	Rough	A hoarse off-sound unintentionally accompanying the reproduced sound. Bass distortion.
	Buz zing	A zzz-like, undesireable sound typically in the low and midrange frequencies.
	Clip- ped	The harmonics are to pronounced and sharp.
	Distor- ted	Additional and undesired sounds that add a sharpness to the reproduction.
	Compres sed	Limited dynamic range leading to a lack of natural peaks. Dymanic compression may be heard as a pumping effect.
Noise	Fluctuating/ Intermittend	Noise with varying loudness and or pauses.
	Bubbling	Sound or noise with fast (<1 sec.) variations in frequency and/or loudness.
	Humming	Low frequency noise with tonal components.
	Hissing	A noise-like sizzling sound, like the sound of bacon in a frying pan.

Figure 3.1.4: Sound Wheel for Reproduced Sound by Pedersen & Zacharov (2015, p. 8)

After a comparison of all studies displayed in Table 3.1.3 the timbral and technical attributes presented in Chapter 3.1.2 have been considered repeatedly as significant throughout these publications, and were thus chosen as a focus of the *Sound Wheel*. This approach offered the possibility to include also other attributes from the *Sound Wheel* during category scaling if the attributes in focus would turn out inappropriate to describe the auditory perception.

3.1.2 Definition of Attributes

As a result of the explanations in Chapter 3.1.1. the following attributes and definitions have been derived (see Table 3.1.4). The definitions have been taken from Berg & Rumsey (2003) and Pedersen & Zacharov (2015), whereas their wording was slightly amended to fit into the context of the current experiment. As can be seen in Table 3.1.4, some attributes entailed a combination of sub-attributes. This was done to reduce the rather large number of attributes to be assessed and thus making the evaluation process more efficient. Although some of these sub-attributes would exclude each other by definition, it was yet considered essential to define them to reduce the possibility of ambiguity (see Width, e.g.).

Spatial Quality: Attitudinal Attributes	
Naturalness:	How similar to a natural listening experience the sound as a whole sounds (i.e. not reproduced through, e.g. loudspeakers).
Presence:	The experience of being in the same acoustical environment as the sound source, e.g. to be in the same room.
Preference:	The degree to which the sound as a whole is pleasing, disregarding the programme content.
Spatial Quality: Descriptive Attributes	
Width: Entailing both, individual source width and room width	**Individual source width:** The perceived width of an individual sound source, or the angle occupied by it. This does not necessarily indicate the known size of such a source, e.g. one knows the size of a piano in reality, but the task is to assess how wide the sound from the source is perceived. This disregards sounds coming from the sound source's environment, e.g. reverberation, as only the width of the sound source itself is assessed. **Room width:** The width or angle occupied by the sounds coming from the sound source's reflections in the room, disregarding the direct sound from the sound source.
Localisation Accuracy:	How easy it is to perceive a distinct location of the source or how easy it is to pinpoint the direction of the sound source. Its opposite is when the source's position is hard to determine - e.g. a blurred position.
Distance/Depth: Entailing source distance and environmental depth	**Source distance:** The perceived distance from the listener to the sound source.

	Environmental depth: Depth of the (reflective) environment within which the source is located.
Envelopment: Entailing source envelopment and room envelopment	**Source envelopment:** The extent to which the sound source envelops the listener or the feeling of being surrounded by the sound source. This disregards sounds coming from the sound source's environment, e.g. reverberation and only assesses the sound source. **Room envelopment:** The extent to which the sound coming from the sound source's reflections in the room (the reverberation) envelops or surrounds the listener.
Spatial Balance:	The distribution of energy at different points in space.
Room Perception: The ability to experience the characteristics of the room, **entailing the perceived room size and room sound level**	**Room size:** The ability to perceive the relative size of the room. **Room sound level:** The level of sounds generated in the room as a result of the sound source's action, e.g. reverberation. This disregards the direct sound from the sound source and extraneous disturbing sounds.
Vertical Image Shift:	The amount of perceived vertical image shift.
Vertical Image Spread:	The amount of perceived vertical spread of the sound source.
Vertical Frequency Separation:	The amount of perceived vertical frequency separation. E.g. if low frequencies are located lower than high frequencies.
Timbral Quality	
Full:	If both low and high frequencies are well represented with extension towards both ends of the spectrum.
Brilliance (also often referred to as **"clear"**)**:**	Crystal-clear reproduction through an extended treble range with an airy and open treble. Lightness, purity and clarity with space for instruments. Clarity in the upper frequencies without being sharp or shrill and without distortion.
Sharp / Bright:	An excessively raised treble content leading to a hard sound.
Nasal:	A closed sound with a pronounced midrange. Gives the impression corresponding to vocalists singing through the nose (nasal).
Bass depth (also referred to as **"thin"** if lacking)**:**	Denotes how far the bass extends downwards. If it goes down in the low end of the spectrum, there is great depth. This should not be confused with bass strength, which

	indicates the strength of the bass or boomy which relates to resonances in the lower bass region.
Boxy:	Denotes a hollow sound, as if the sound was played inside a small box and represents resonances in the upper bass frequency range.
Homogeneous:	Denotes to which degree the different frequency ranges (bass, midrange and treble) are coherent, continuous, and balanced without gaps between them. There are seamless transitions between the tone ranges.
Timbral realism (also often referred to as **"natural"** or **"coloured"**):	The extent to which the reproduced audio sounds like any real experience. This does not necessarily have to be the original sound the recording was trying to reproduce.
Treble content in reverberation:	Allows perceiving different aspects of a room, as according to Gerzon *"realism is lost if the rear treble response is poorer than the front."* (1971, p. 12)
Dynamics (only for guitar and djembe)	
Attack:	Denotes the transient response: The ability to reproduce transients cleanly and separated from the rest of the sound image. An imprecise attack is understood as unclear or as a muted impact. It denotes the degree to which one can hear the actual strokes on the djembe or the plucking of the strings of the guitar.
Punch:	Specifies whether the strokes on the percussion or the plucking of the strings of the guitar are reproduced with clout, almost as if one can feel the blow. It also denotes the ability to handle large volume excursions without compression effortlessly.
Powerful:	The ability to handle high sound levels, especially when striking the percussion. It indicates whether the punch and attack are maintained at high volume.

Table 3.1.4: Definitions of Attributes

3.1.3 Selection of Response Format and Scaling Procedure

After the selection and definition of the attributes (Figure 3.1.1, left) the next step was to choose the most appropriate response format and scaling procedure (Figure 3.1.1, right) when following the procedure of Bech & Zacharov (2006, pp. 94 - 95).

In short, the chosen response format for the evaluation of the spatial attributes was a direct scale magnitude estimation where the subject assigns a (numerical) value to one stimulus and then judges subsequent stimuli against the first. This response format was considered appropriate as it stands in accordance with the test design used in similar investigations such as Millns & Lee (2018), Howie *et al.* (2017), Howie *et al.* (2015), Gribben & Lee (2015), and Lee & Gribben (2014).

Direct scale magnitude estimation, however, was considered ineffective and inefficient for the grading of timbral and technical attributes. This led to the choice of category scaling for the evaluation of timbral and technical attributes. In category scaling, the subject is asked to assign a category (in this case a timbral or dynamic label) to each stimulus presented. The method of category scaling seemed justified as according to Bech & Zacharov *"this scaling method is often employed for audio evaluations."* (2006, p. 72)

Table 3.1.5 displays the requirements to be considered in the selection process of an appropriate response and scaling method according to Bech & Zacharov (*ibid.*, pp. 83-96) and how these have been met by the current choice (see Chapter 3.2). For a detailed description of these requirements, the reader is referred to Bech & Zacharov (*ibid.*).

Requirement of the Response Format and Scale	Implementation in the Test Design
It should be meaningful to the subject	The subject was provided with an unambiguous explanation of the scale and what it is intended to measure. The attributes have been clearly defined.
It should be uncomplicated to use	Using the HULTI-Gen (Huddersfield Universal Listening Test Interface Generator) allowed for a simple slider-based user interface control.
It should possess the ability to differentiate between the stimuli of interest	With a resolution from 0-100 the subject was provided with a continuous scale. Therefore intermediate scale values were guaranteed. Based on previous studies and the ITU-R Recommendation BS.1116 anchor labels other than the extremes were avoided to reduce bias.
It should be relevant for the task	The chosen attributes have been shown to produce statistically significant results in the evaluation (see Chapter 3.1.1). The subject understood the definition and the use of the attributes. The selected stimuli excited all the attributes.
It should avoid the endpoint effect	The subject was familiar with all stimuli before doing the listening test and was thus familiar with the range of auditory impressions.
It should reduce context effects	Using the following methods, context effects were limited as much as possible:

	<u>Randomisation:</u> Order effects and sequential dependencies have been eliminated by applying a random presentation of the trials and stimuli through HULT-Gen. <u>Stabilisation:</u> Anchors for the endpoints of the scale were introduced. As the GUI did not allow for randomisation of the reference, no separate auditory anchor reference was included. Instead, the stimulus with the highest attribute response was taken as a reference with a grading of 100, and the other two stimuli were rated accordingly, as proposed by Howie *et al.* (2017, p. 5) <u>Calibration:</u> The subject was familiar with the testing method prior to the test. <u>Interpretation:</u> The experimenter was aware of the experimental context and used that knowledge when analysing and interpreting the results.
It should be unbiased	<u>Minimising contraction bias:</u> Stimuli were randomised, and the response range was anchored using the stimulus with the highest attribute perception as a reference for the value of 100. <u>Minimising bias caused by unfamiliarity with units of magnitude:</u> The subject was familiar with all stimuli prior assessment. <u>Minimising bias caused by unfamiliarity with the mapping of the responses to the stimuli:</u> Logarithmic response bias was avoided by hiding the numerical display during grading. Range equalising bias was avoided by letting the subject choose its response ranges and avoiding category ratings within the scale. <u>Reducing other bias effects:</u> Perceptual sensitivity: The subject had previous experiences in critical listening of various spatial audio attributes and reported normal hearing. This had been tested by an otologist one year before the test. As the current listening test was not conducted in the field of low bit-rate coding systems, the bias of perceptual oversensitivity could be ignored.

	Halo bias: Stimuli and thus the reference were randomised. Two different response formats were chosen for spatial and timbral evaluation. Dumping bias: Neither was the scale restricted in the evaluation of the spatial attributes nor was a category missing when evaluating timbral and other qualities of the sound, as the subject was free to add comments in case the provided categories would not be sufficient to describe the auditory perception appropriately.

Table 3.1.5: Requirements of Response Formats and Scales According to Bech & Zacharov (2006, pp. 83-96)

3.2 Implementation of the Test Design

For the evaluation of the spatial attributes, double-blind multiple stimuli comparison tests were conducted using a GUI with the Huddersfield Universal Listening Test Interface Generator, a Cycling 74 Max-based tool (Gribben & Lee, 2015). The subject could freely turn its head if it stayed in the sweet spot. The task was to grade three stimuli against each other on a continuous rating scale. The scale ranged from 0 (labelled “lesser”) to 100 (labelled “greater”), whereas the stimulus with the “greatest” attribute impression was taken as a reference of 100 with the other two stimuli being graded accordingly. This procedure was proposed by Howie *et al.* (2017, p. 5) to reduce scaling bias. The presentation order of both, the stimuli and the trials was randomised to avoid the potential biases stated in Chapter 3.1.3. The stimuli were synced in playback, meaning the subject could switch between mixes at any point during the playback. For each test, the subject was to complete a total of five trials (corresponding to the five musical sources), each of which contained the stimuli of the different mixes (resulting in 3 stimuli per trial). Figures 3.2.1-3.2.5 display the configuration and the resulting GUI of the listening test according to the explanations given in Chapter 3.1.3 (see next page):

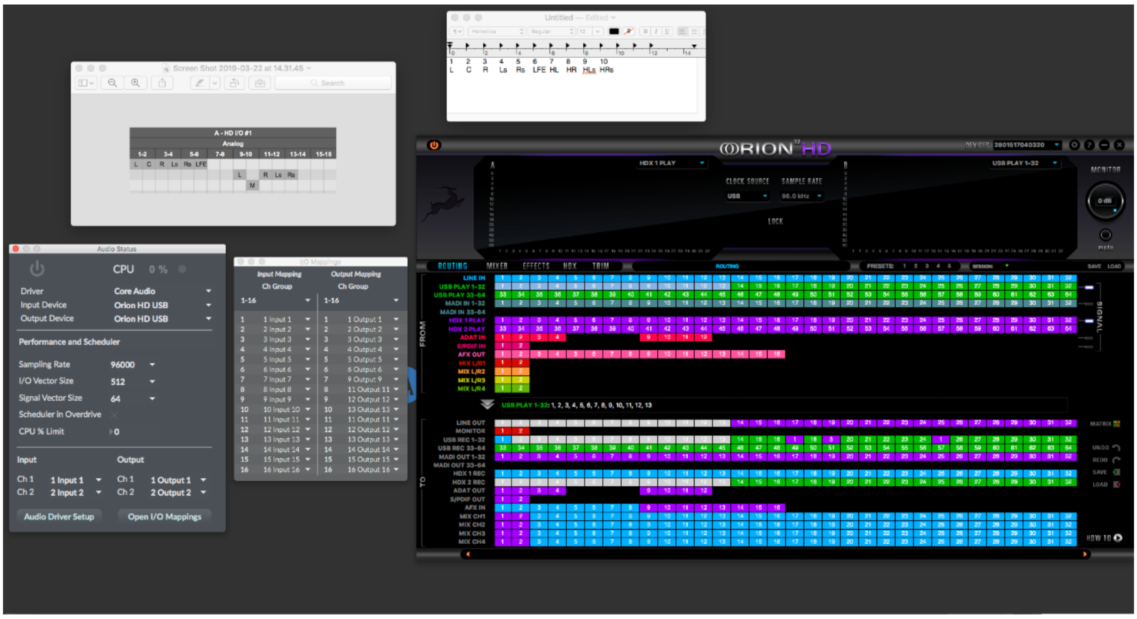


Figure 3.2.1: Routing of the Individual Channels of the 9.1 Stimulus Files to their Corresponding Outputs

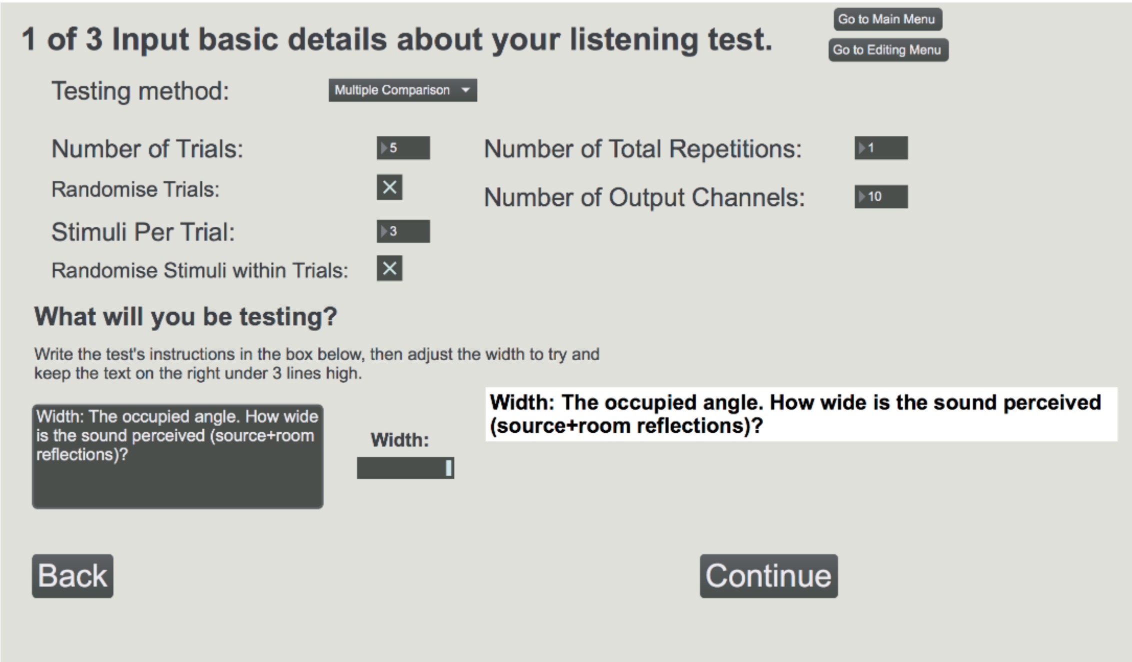


Figure 3.2.2: Listening Test Basic Settings

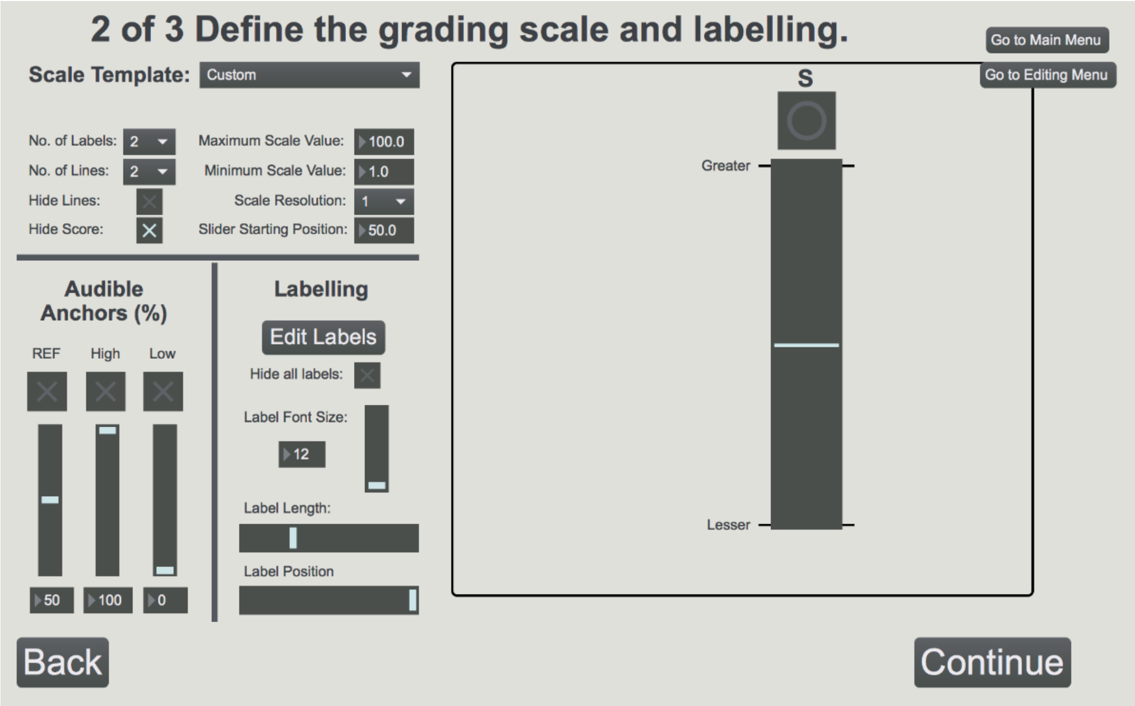


Figure 3.2.3: Listening Test Scale Design

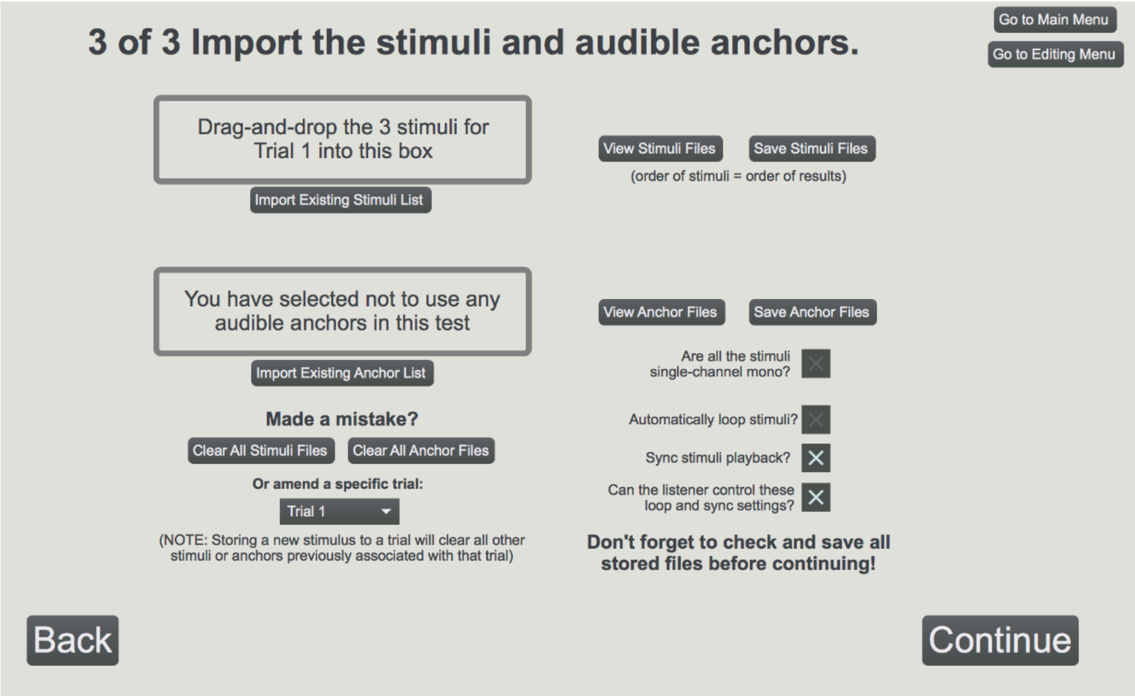


Figure 3.2.4: Listening Test Playback Settings

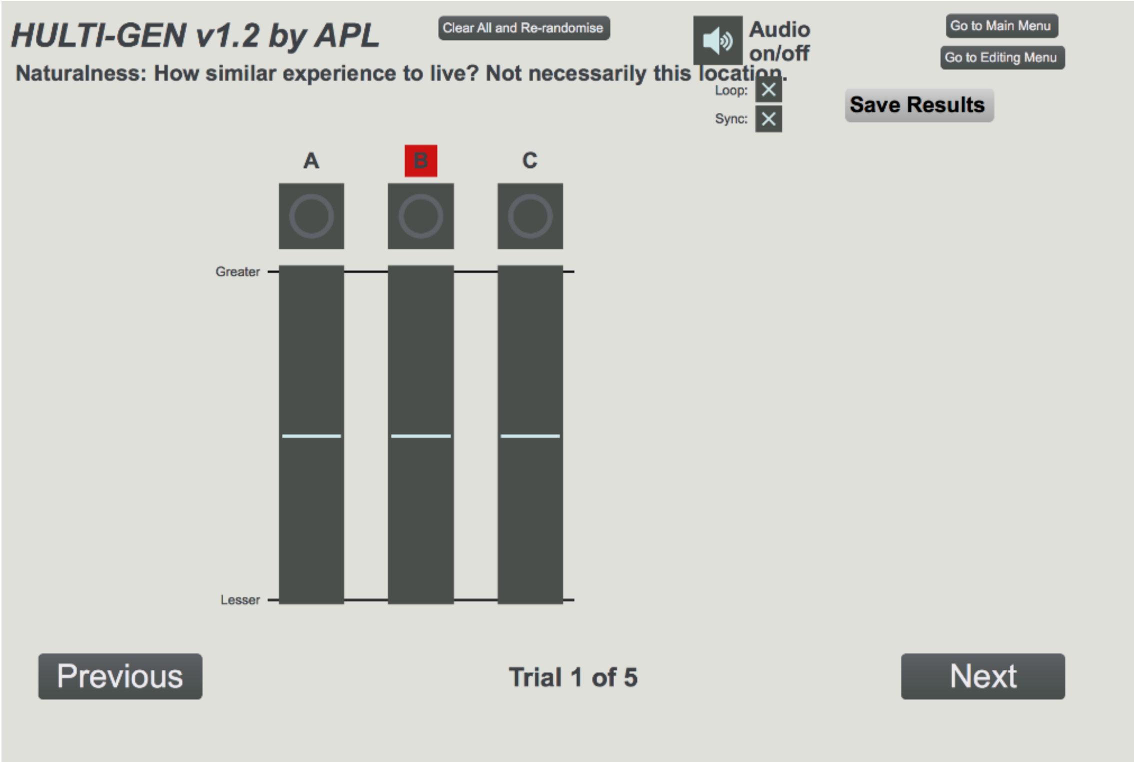


Figure 3.2.5: Listening Test GUI

When rating Preference, the subject was not given any specific subjective qualities or attributes to consider but was advised to write a comment on its decision on a paper. For the evaluation of all other attributes, the same GUI was used for playback to allow for a double-blind test and the randomisation of the stimuli and trials. The sliders, however, were not used as the task was to assign one or multiple labels to a stimulus by writing the assigned label on a paper. The labels could be chosen from the timbral and dynamic categories listed in Chapter 3.1.2. If, however, none of the labels provided in this list (see Table 3.1.4) seemed appropriate for describing the perception, the subject was free to choose another label of Pedersen & Zacharov's *Sound Wheel* (see Figure 3.1.4).

3.3 Reproduction Setup

All 3D audio playback including mixing took place in the Auro-3D studio at SAE Brussels, an acoustically treated listening room with a 10.1 Auro-3D setup with Sonodyne SM100AK speakers, as can be seen in Figure 3.3.1:

Figure 3.3.1: Auro-3D Studio at SAE Brussels (SAE Institute Brussels, 2017)

The sound pressure level of each loudspeaker was measured and calibrated for 79dB SPL (A-weighting, slow response) at the listening position using pink noise, as proposed by Holman (2008, p. 69). The monitor level was calibrated and kept constant to -18dB throughout the sessions.

3.4 Stimuli Creation

3.4.1 Selection of Stimuli

Based on previous similar studies in the field of psychoacoustics such as Howie *et al.* (2017; 2015), musical excerpts of 30 seconds were chosen. The selection was mostly aimed towards passages containing pauses of a certain length to ease the perception of room-related attributes, as proposed by Rumsey (2002, p. 659). Related to that the stimuli should exhibit as little background noise as possible to avoid any distractions from the assessment tasks as much as possible.

3.4.2 Procedure

Rumsey addressed the need for ecological validity and its function in the experimental design when comparing different recording techniques:

> *“Ecological validity describes the extent to which an experimental situation matches the real-world context and circumstances it is supposed to represent. For example, numerous psychological experiments take place under highly controlled laboratory conditions that may give rise to unrepresentative human responses. Such situations could be considered to have low ecological validity. Ecological validity is similar to external validity, which relates to the validity of experimental results outside the context of the individual experiment. In psychoacoustic experiments there is nearly always a tension between ecological validity and scientific control of variables - the more tightly one controls experimental variables in order to observe individual effects, the less ecologically valid the experiment becomes. There appears to be a form of uncertainty principle at work, in that one can obtain an unambiguous result with high certainty but low ecological validity, or a more uncertain result with higher ecological validity. The more like a real-world situation the experiment becomes, the less easy it is to control all the variables. This tension is strongly evident when one tries to undertake controlled experiments comparing recording techniques.”* (2002, p. 654)

Based on this, the method of a balanced mix for stimuli creation was chosen, as a balanced mix represents a scenario of high ecological validity. Besides, comparable studies have applied the same process, such as Howie *et al.* (2017, p. 5; 2016, p. 6; 2015, p. 6) or Luthar *et al.* (2015, p. 3). When balancing, the aim was

> *“to convey a sense of depth and realism to the instruments, using a 'direct sound/instrument in front, ambience to the sides, behind and above' approach. It was considered very important that mixes contain enough height channel information to be pleasant, realistic and enveloping, rather than exaggerating the differences between polar patterns. The goal was not to create an 'obvious' listening test, but one that mirrored the subtle mix differences that professional engineers discriminate between on a daily basis.”* (Howie *et al.,* 2015, p. 6)

This includes that *“careful attention was given to maintaining a similar balance of direct to reverberant sound for each technique. No filtering of any kind was applied to the microphone signals.”* (*ibid.,* 2018, p. 5)

The approach of intra-array level matching as in Martin *et al.* (2016) has been taken into account but was considered ecologically invalid. A pure channel-based routing as in Kassier *et al.* (2005), on the other hand, was not possible due to the recording procedure the recording engineer was forced to apply (see Appendix, Chapter 9.6.3) as it influenced the original ICLDs of some of the recorded signals.

Before starting to balance, phase relations have been checked aurally and visually, and the polarity was flipped where necessary. However, as the recording setup was measured, signals have been found mostly to be in phase (as shown in Figures 3.4.1-3.4.3).

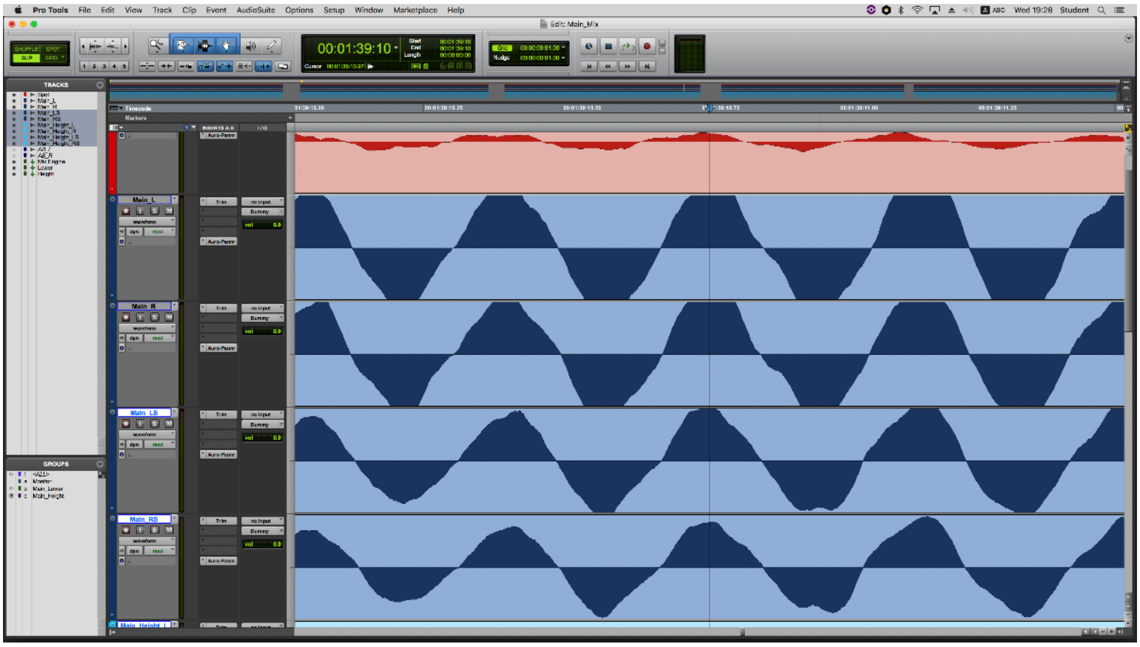


Figure 3.4.1: Main-Array Channels in Phase

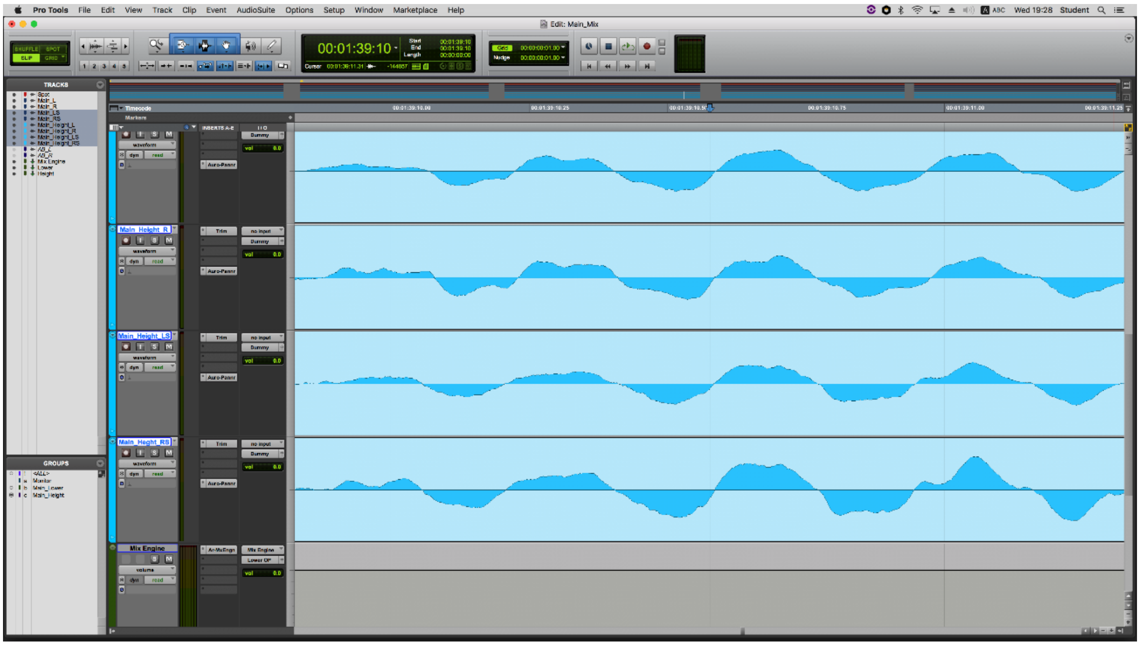


Figure 3.4.2: Main-Array Height Channels in Phase

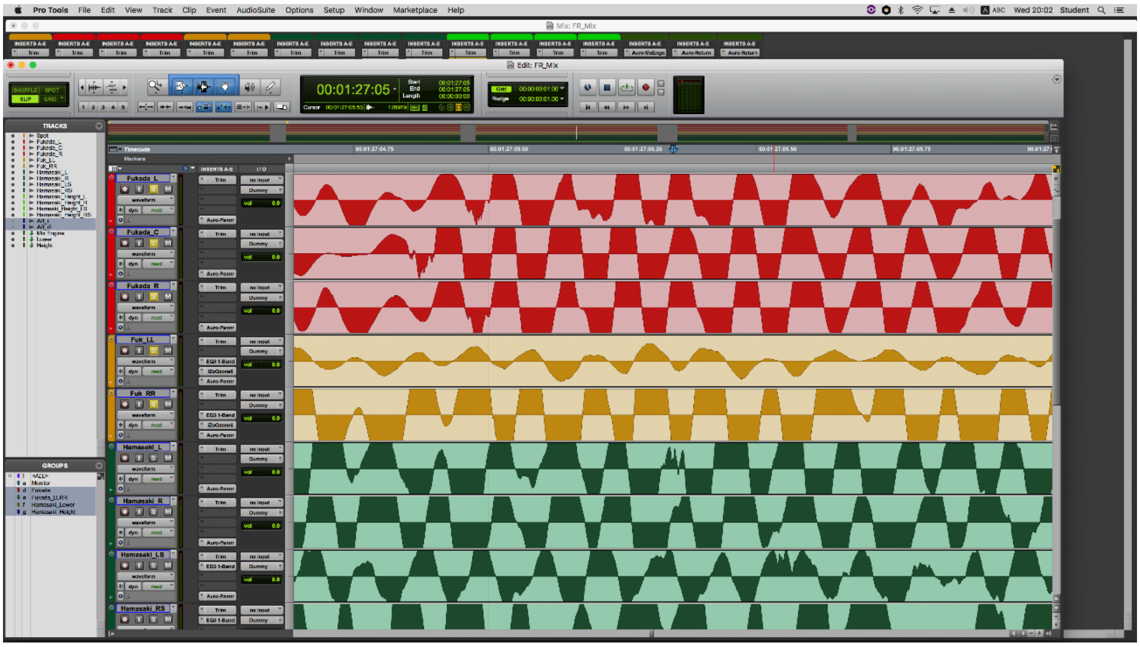


Figure 3.4.3: Some Channels out of Phase within the F/R-Approach

Furthermore, the differences in level on the channels where the preamp gain had to be turned down enforcedly during recording (see Appendix, Chapter 9.6.3) have been compensated by ear. The reference for the level adjustment was the microphone signals from the same part of the array in question. In addition, a LPF was applied to the Fukada LL and RR channels at 250 Hz, as proposed by Rumsey (2013, p. 197):

Figure 3.4.4: LPF at 250Hz for the Fukada LL and RR Channels

The stimuli have been mixed in Pro Tools with the Auro-3D Authoring Tools (version 2.5) through the Digidesign D-Control (ICON) control surface, as can be seen in Figures 3.4.5-3.4.8:

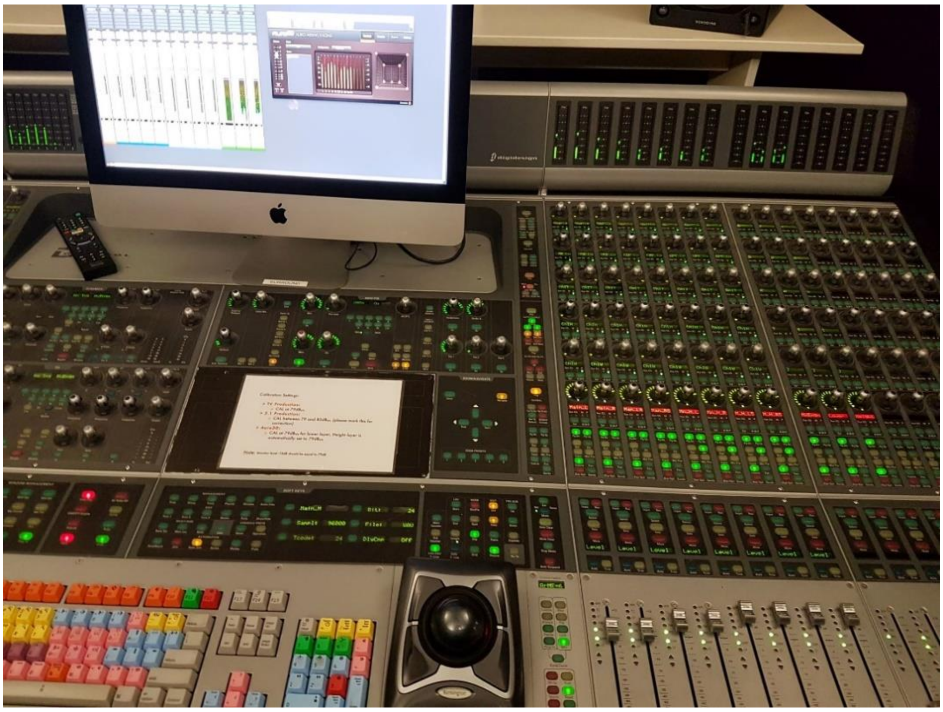


Figure 3.4.5: Auro-3D Authoring Software Controlled by the ICON

Figure 3.4.6: Auro-3D Panner

Figure 3.4.7: Auro-3D Mixing Engine Displaying the 9.1 Mix

Figure 3.4.8: Auro-3D Authoring Tools Overview

Due to the nature of the research questions no processing was applied: neither for removing preamp noise nor other artefacts, thus following the procedure in Howie *et al.* (2018, p. 5). The appropriateness of the chosen excerpt allowing to perceive attribute related differences between the mixes as good as possible was weighted higher than the removal of any such noise as this would imply a direct intervention into an experimental constant being the recorded signals entailing all spatial and timbral cues. Nevertheless, the artefacts in question have not been perceived as distracting during mixing or evaluating the intended attributes.

All stimuli have been level matched for ± 0.2dB using the following procedure: A 3Dio FS Pro II dummy head was placed at the listening position at ear level and used to record the playback of each stimulus as can be seen in Figures 3.4.9 and 3.4.10. The binaural input was monitored with an LUFS meter (iZotope Insight) with an integration window of 30 seconds (stimulus duration) which can be seen in Figures 3.4.11 and 3.4.12. This method for stimuli level matching has been applied previously in similar studies like Howie *et al.* (2017, p. 5) or Martin *et al.* (2016, p. 4).

Figure 3.4.9: Dummy Head Position Side View

Figure 3.4.10: Dummy Head Position Back View

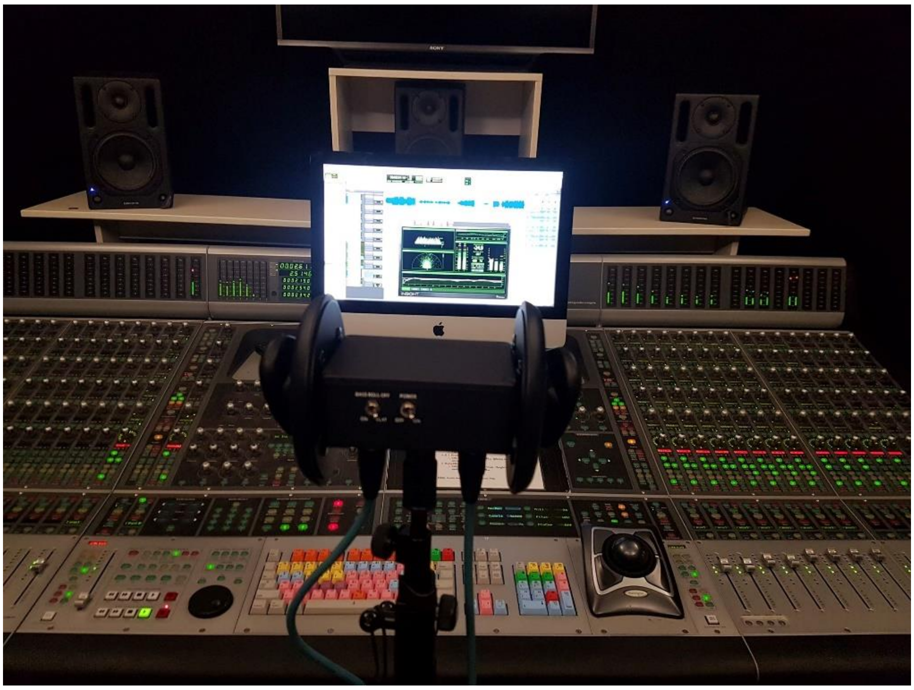


Figure 3.4.11: iZotope Insight Measuring Binaural Dummy Head Input LUFS

Figure 3.4.12: Stimuli Level Matching

Once this has been done, the mixes were bounced into multiple mono files and recombined in Wave Agent to obtain the 9.1 wav files demanded by the HULTI-GEN. All 9.1 stimuli files have been checked on their correctness before importing them into the GUI.

3.5 Limitations

Referring to Bech & Zacharov, several requirements for improving the validity of the results could not be met within the scope of the current research. Due to limited resources, the process of subject pre- and post-selection (2006, pp. 118-140) could not be accomplished in order to obtain a listening panel of at least five expert subjects to ensure a sufficient resolution in the test (*ibid.*, p. 113). However, a preliminary (partial) listening test was conducted one day before the listening test, and the results of the two tests showed a strong correlation, which indicates intra-subject consistency. Furthermore, other field-related investigations have been published with the experimenter being the subject simultaneously, such as Williams (2012), King *et al.* (2016), Riaz *et al.* (2017), Geluso (2012), Bowles (2015), Sawaguchi (2018), and Luthar *et al.* (2015). Similarly, the validity of the results can be seen reduced as only one balancing engineer was involved, and not several, to minimise the factor of subjectivity in stimuli creation (compared to Howie *et al.,* 2017; 2015).

Regarding the experimental design the limitation of transfer bias and expectation bias could not be accomplished according to Bech & Zacharov (2006, pp. 92 - 93), as the same subject was used for assessments of different attributes and different conditions (mixes) of the same attribute. However, the same authors stress the advantages of having permanent listeners throughout the tests and the magnitude of transfer bias in listening tests has not yet been researched to the author's knowledge. Also, as the subject was at the same time the experimenter, the subject was familiar with the experimental detail and thus more prone to expectation bias. The results throughout the listening tests for the different array-mixes, however, suggest that the magnitude of expectation bias was not dominating the subject's evaluation.

Further limitations include that the recording was conducted only in one space as the characteristics of the room can take influence on the perception of environment-related attributes (see Appendix Chapter 9.1.2). Similarly, as all sources were recorded solo, ensemble-related attributes could not be tested. In relation to that, the ecological validity of the experiment would be improved when involving ensembles as this enables *"the full range of problems and effects that can arise in spatial audio production."* (Rumsey, 2002, p. 657) The required evaluation time for this, however, would have exceeded the allocated time for the resources. In addition, based on the recording setup, the results are only applicable to the dry-wet scenario, and not a sound all-around approach.

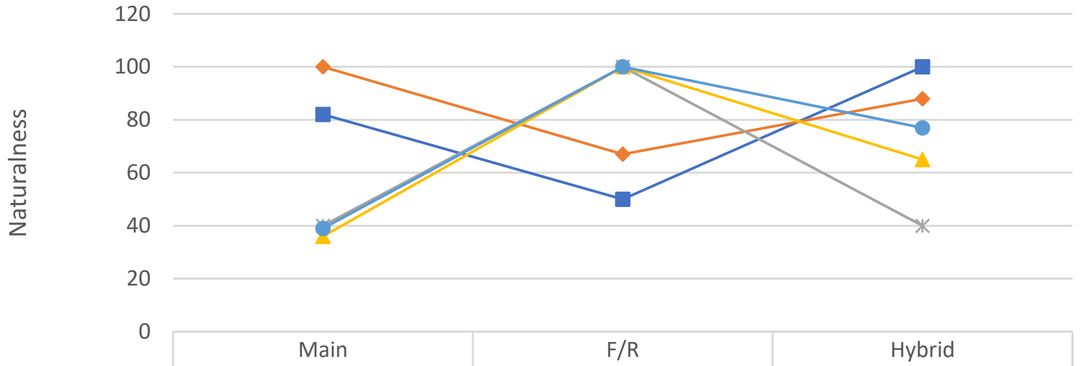
4. Results

Tables 4.1.1-4.1.16 visualise the values obtained in the listening test through direct scale magnitude estimation:

4.1 Graphical Representations of the Listening Test Results

Spatial Quality: Attitudinal Attributes

Naturalness

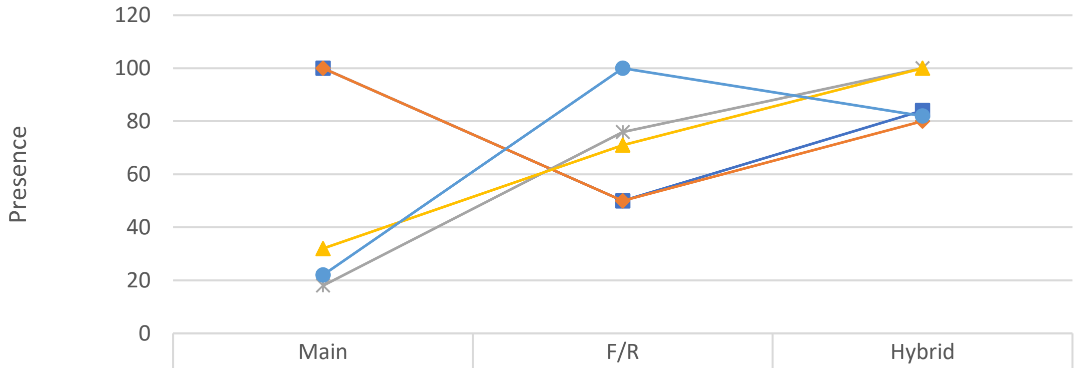


	Main	F/R	Hybrid
Cello	82	50	100
Violin	100	67	88
Djembe	40	100	40
Handpan	36	100	65
Guitar	39	100	77

Cello Violin Djembe Handpan Guitar

Table 4.1.1: Results for the Attribute Naturalness

Presence

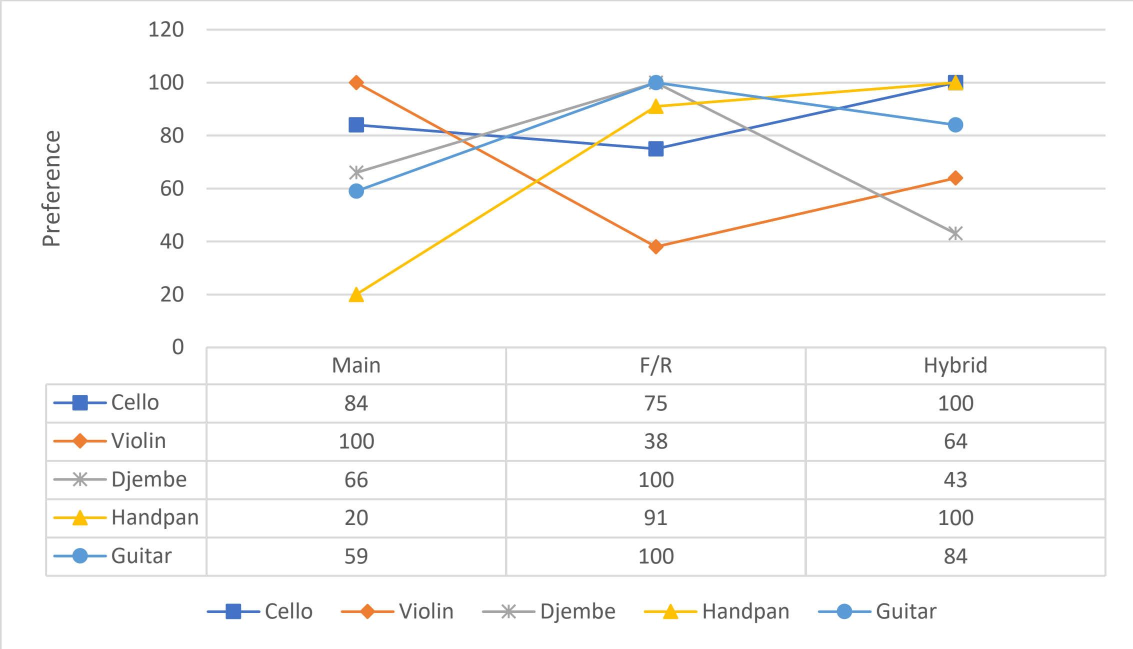


	Main	F/R	Hybrid
Cello	100	50	84
Violin	100	50	80
Djembe	18	76	100
Handpan	32	71	100
Guitar	22	100	82

Cello Violin Djembe Handpan Guitar

Table 4.1.2: Results for the Attribute Presence

Preference

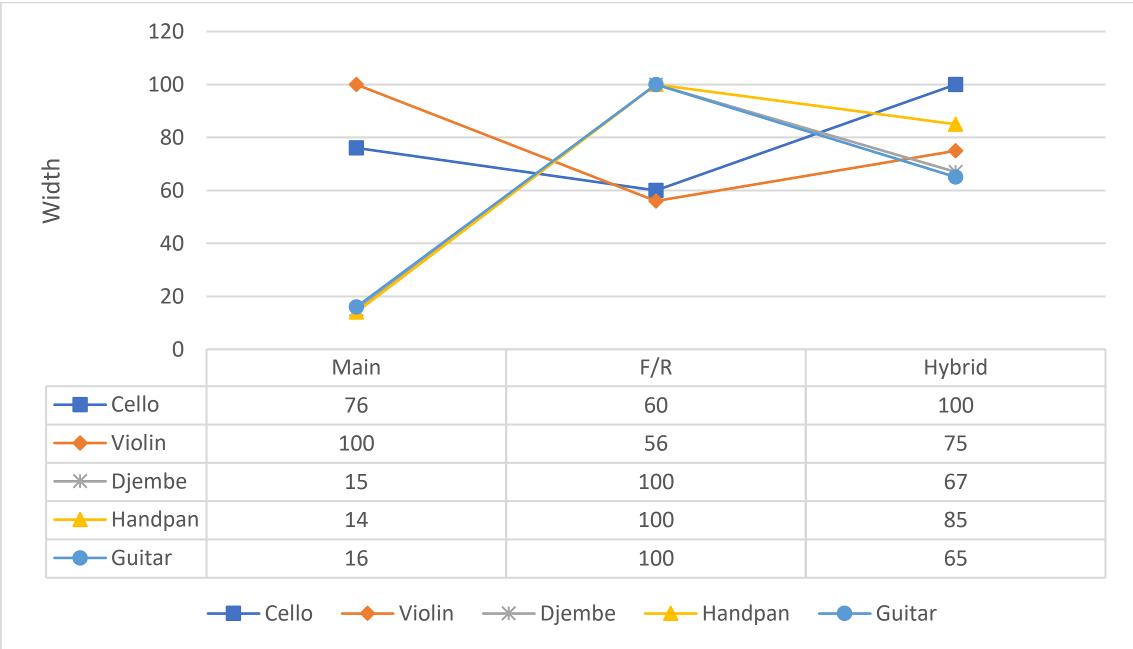


	Main	F/R	Hybrid
Cello	84	75	100
Violin	100	38	64
Djembe	66	100	43
Handpan	20	91	100
Guitar	59	100	84

Table 4.1.3: Results for the Attribute Preference

<u>Spatial Quality: Descriptive Attributes</u>

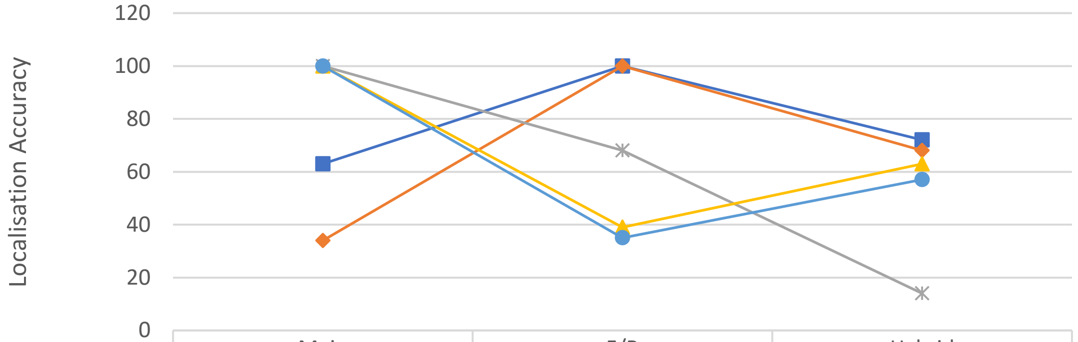
Width



	Main	F/R	Hybrid
Cello	76	60	100
Violin	100	56	75
Djembe	15	100	67
Handpan	14	100	85
Guitar	16	100	65

Table 4.1.4: Results for the Attribute Width

Localisation Accuracy

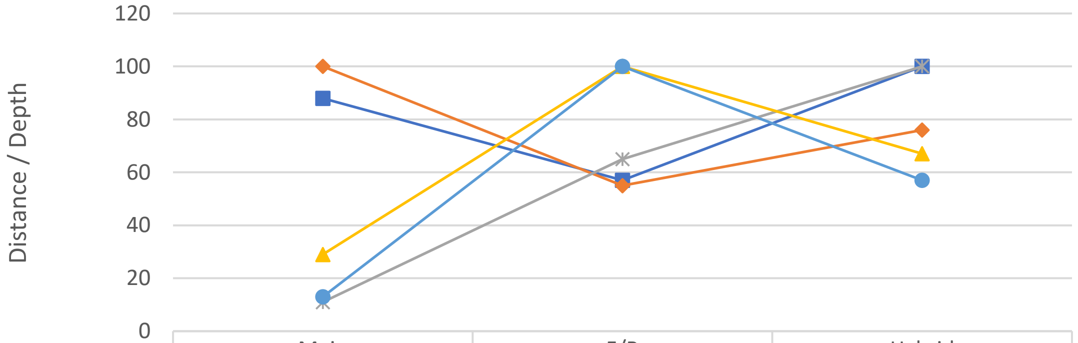


	Main	F/R	Hybrid
Cello	63	100	72
Violin	34	100	68
Djembe	100	68	14
Handpan	100	39	63
Guitar	100	35	57

Cello Violin Djembe Handpan Guitar

Table 4.1.5: Results for the Attribute Localisation Accuracy

Distance/Depth

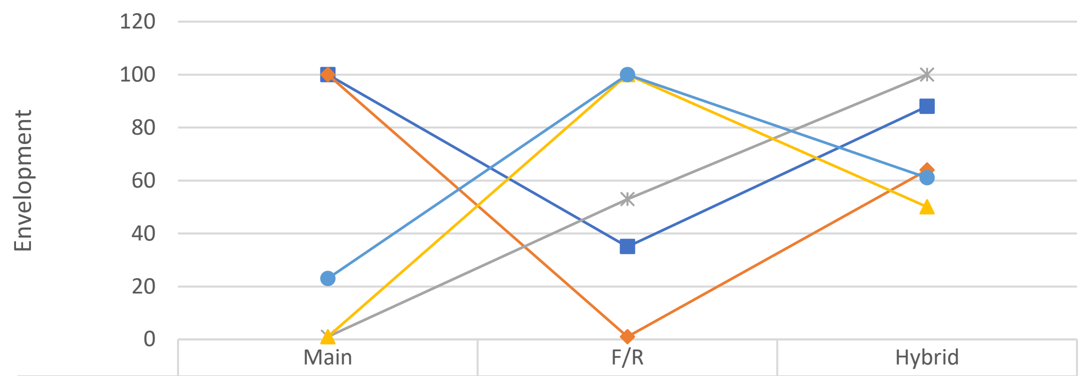


	Main	F/R	Hybrid
Cello	88	57	100
Violin	100	55	76
Djembe	11	65	100
Handpan	29	100	67
Guitar	13	100	57

Cello Violin Djembe Handpan Guitar

Table 4.1.6: Results for the Attribute Distance/Depth

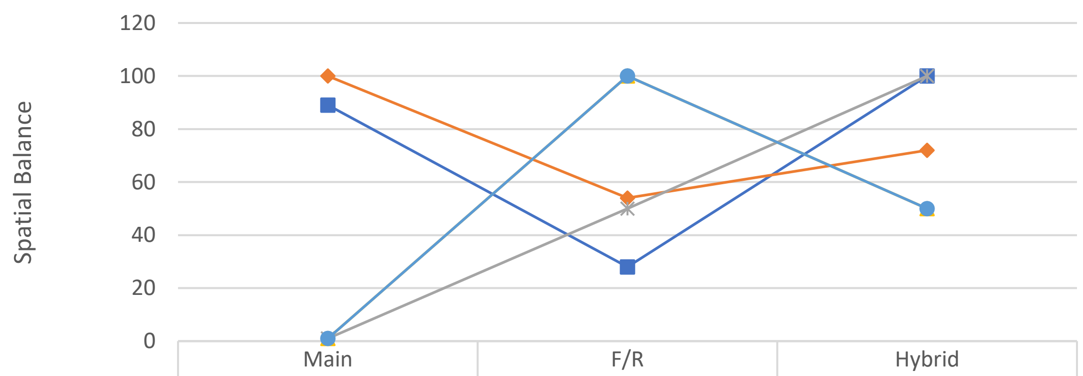
Envelopment



	Main	F/R	Hybrid
Cello	100	35	88
Violin	100	1	64
Djembe	1	53	100
Handpan	1	100	50
Guitar	23	100	61

Table 4.1.7: Results for the Attribute Envelopment

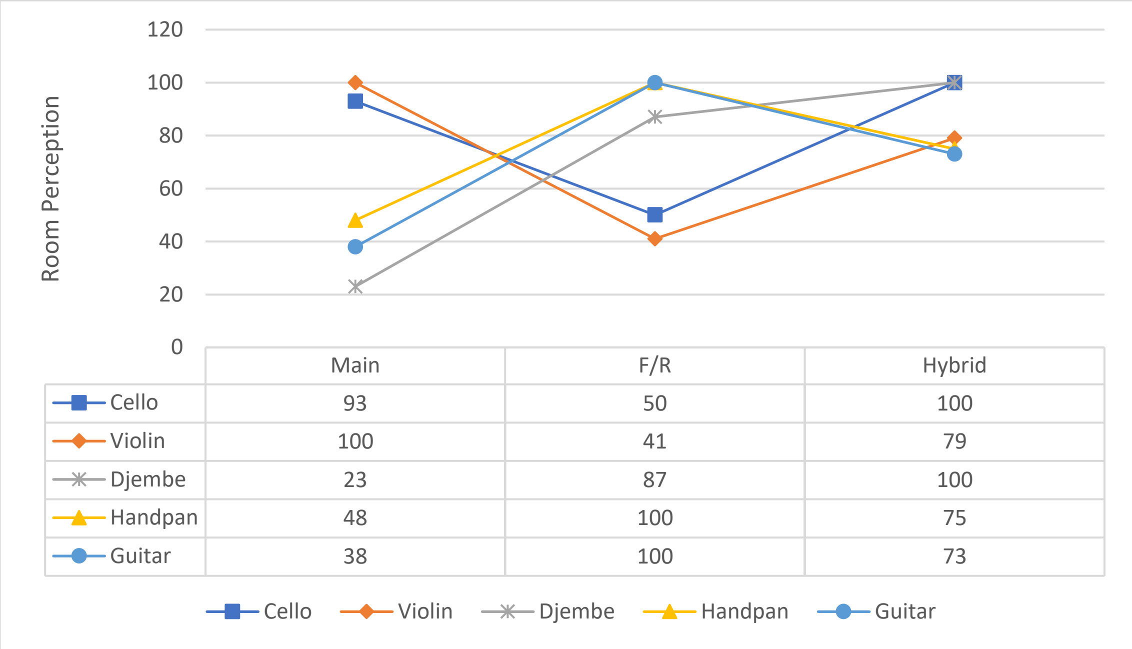
Spatial Balance



	Main	F/R	Hybrid
Cello	89	28	100
Violin	100	54	72
Djembe	1	50	100
Handpan	1	100	50
Guitar	1	100	50

Table 4.1.8: Results for the Attribute Spatial Balance

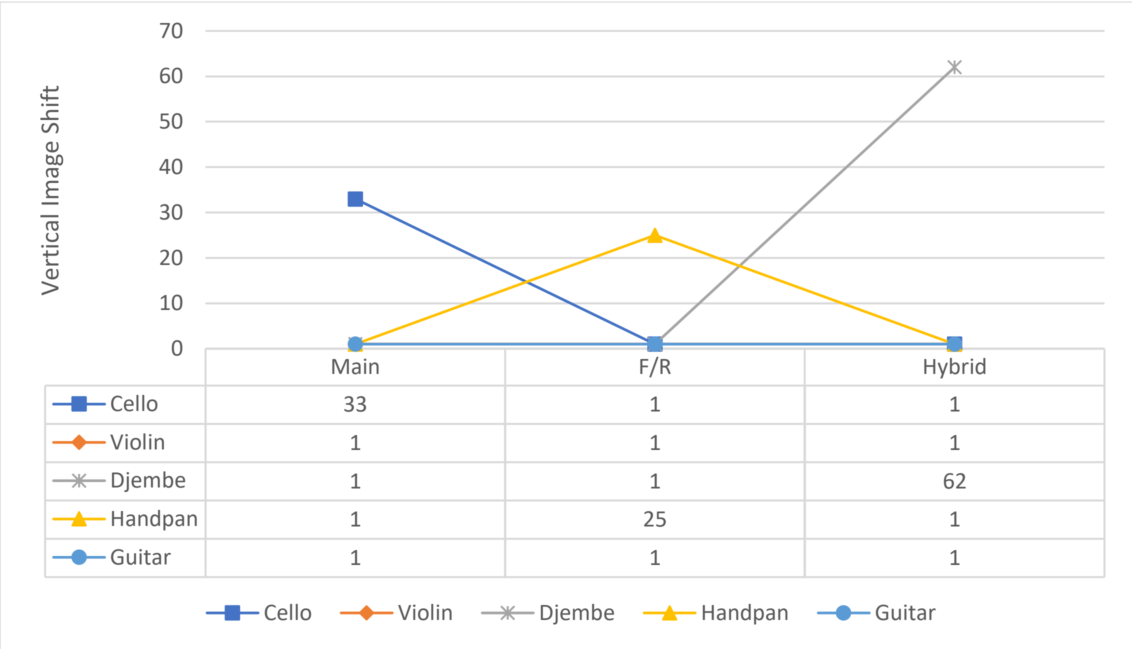
Room Perception



	Main	F/R	Hybrid
Cello	93	50	100
Violin	100	41	79
Djembe	23	87	100
Handpan	48	100	75
Guitar	38	100	73

Table 4.1.9: Results for the Attribute Room Perception

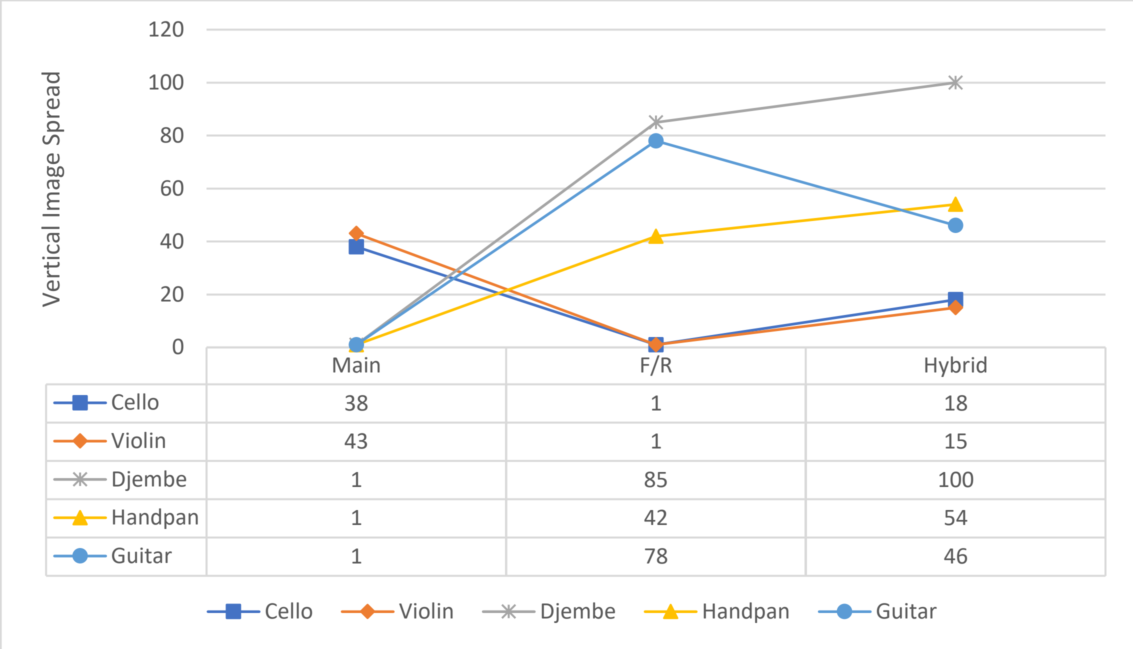
Vertical Image Shift



	Main	F/R	Hybrid
Cello	33	1	1
Violin	1	1	1
Djembe	1	1	62
Handpan	1	25	1
Guitar	1	1	1

Table 4.1.10: Results for the Attribute Vertical Image Shift

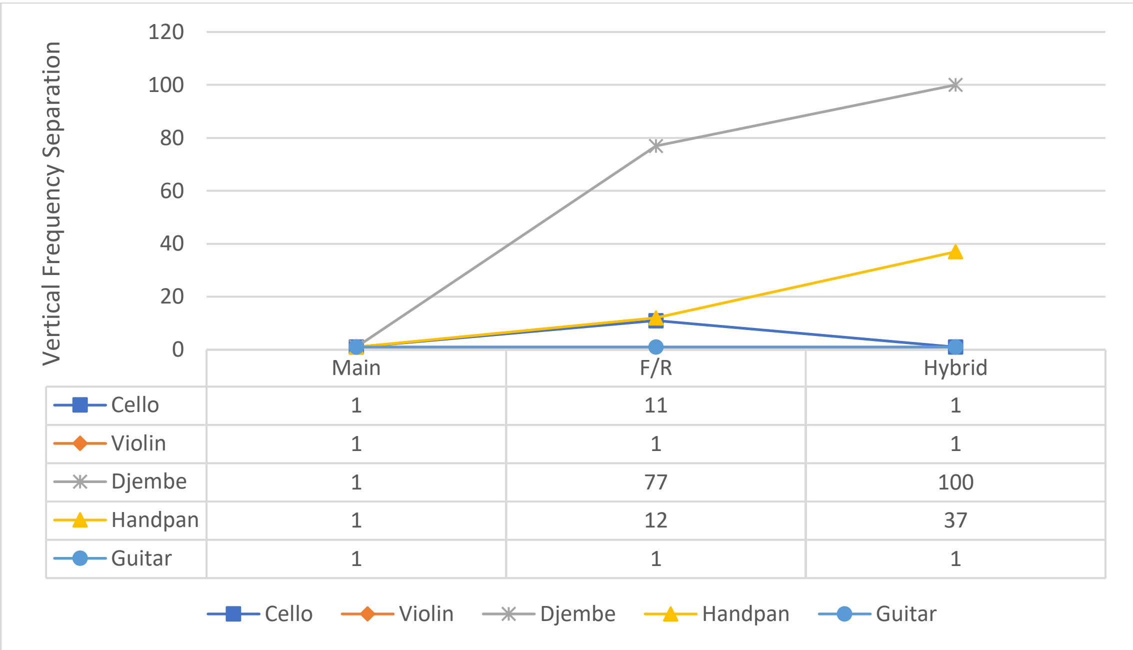
Vertical Image Spread



	Main	F/R	Hybrid
Cello	38	1	18
Violin	43	1	15
Djembe	1	85	100
Handpan	1	42	54
Guitar	1	78	46

Table 4.1.11: Results for the Attribute Vertical Image Spread

Vertical Frequency Separation

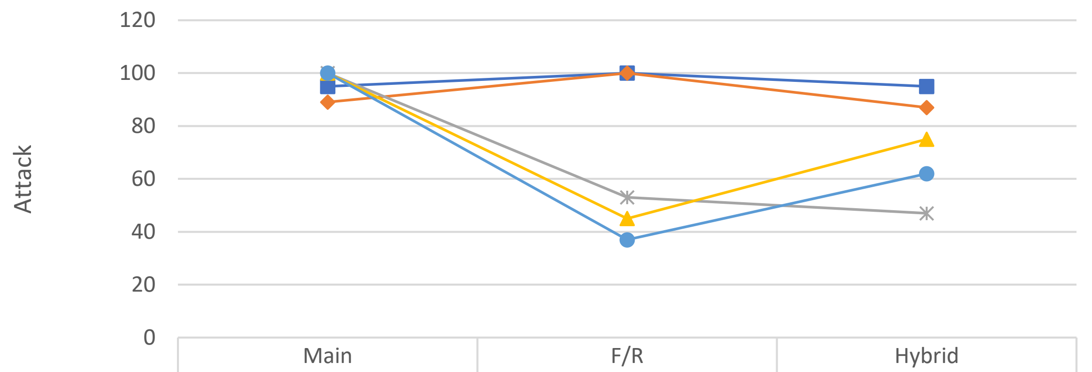


	Main	F/R	Hybrid
Cello	1	11	1
Violin	1	1	1
Djembe	1	77	100
Handpan	1	12	37
Guitar	1	1	1

Table 4.1.12: Results for the Attribute Vertical Frequency Separation

Dynamics

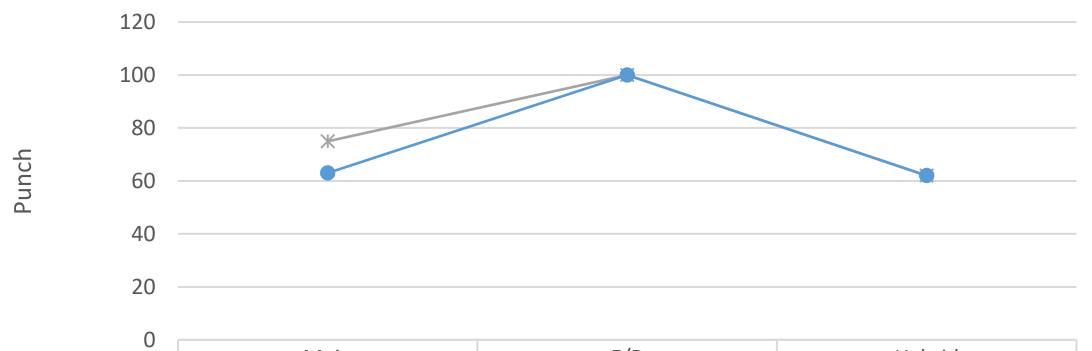
Attack



	Main	F/R	Hybrid
Cello	95	100	95
Violin	89	100	87
Djembe	100	53	47
Handpan	100	45	75
Guitar	100	37	62

Table 4.1.13: Results for the Attribute Dynamics

Punch



	Main	F/R	Hybrid
Cello			
Violin			
Djembe	75	100	62
Handpan			
Guitar	63	100	62

Table 4.1.14: Results for the Attribute Punch

Powerful

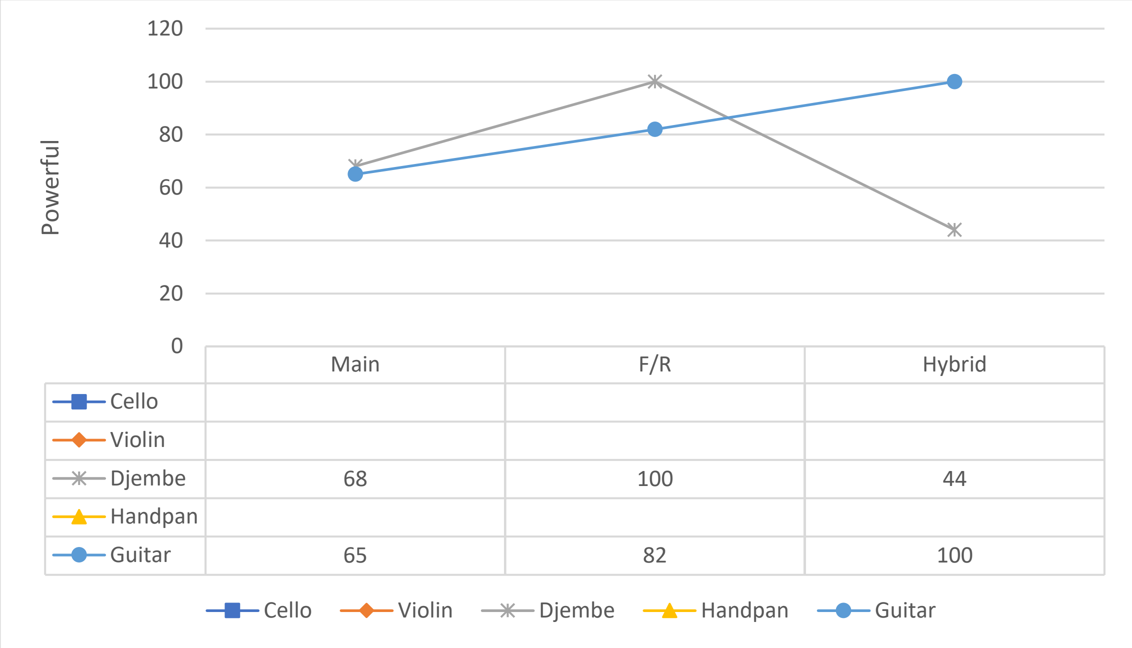


Table 4.1.15: Results for the Attribute Powerful

<u>Timbre</u>

The results of the auditory evaluation regarding Timbre through category scaling can be seen in Table 4.1.16. Figure 4.1.1 indicates the order in which the trials and stimuli were played back.

Source (Trial)	**Stimulus used for scale A**	**Stimulus used for scale B**	**Stimulus used for scale C**
Handpan (Trial 4)	thin	full, but lacking mids	• slightly nasal, • less full than B, but more homogeneous • natural • highest treble content in reverb
Cello (Trial 1)	full	• homogeneous • natural • highest treble content in reverb	• (very) thin • nasal
Djembe (Trial 3)	• homogeneous • natural • highest treble content in reverb	• full, but completely lacking mids • excessive bass	• canny/nasal • thin
Violin (Trial 2)	Treble strength neutral	• sharp • thin • nasal	• brilliance • highest treble content in reverb
Guitar (Trial 5)	• lacking mids • nasal • thin	• homogeneous • brilliant • full	• completely lacking midrange • very bright

Table 4.1.16: Results for the Timbral Attribute Category Scaling

```
1, "Trial 1" 50. 50. 50.;
2, "Trial 2" 50. 50. 50.;
3, "Trial 3" 50. 50. 50.;
4, "Trial 4" 50. 50. 50.;
5, "Trial 5" 50. 50. 50.;
100, Trial Order 4 1 3 2 5;
101, 2 3 1;
102, 3 2 1;
103, 2 1 3;
104, 2 1 3;
105, 3 2 1;
```

Figure 4.1.1: Order of Trial and Stimuli Randomisation during the Evaluation of Timbral Attributes

4.2 Correlations between Attributes and Subject Responses

Figure 4.2.1 depicts a schematic representation of the correlations found between the assessed attributes based on their response patterns through direct scale magnitude estimation (what array was graded how for what source regarding a specific attribute). Vertically coherent displayed attributes showed an identical response pattern regarding the ranking of the different arrays. Horizontally intended attributes showed an identical response pattern to the attributes above, except for one source. Similarly, the closer the attributes are horizontally to each other, the more similar are their response patterns. The attributes Vertical Frequency Separation, Vertical Image Shift, Powerful and Punch have not been included in this graphic as Powerful and Punch have only been tested on two sources, and Vertical Frequency Separation and Vertical Image Shift have been measured with a reference of 1 (“not existent”). Therefore, the obtained values for these attributes do not allow for a direct comparison with the other attributes.

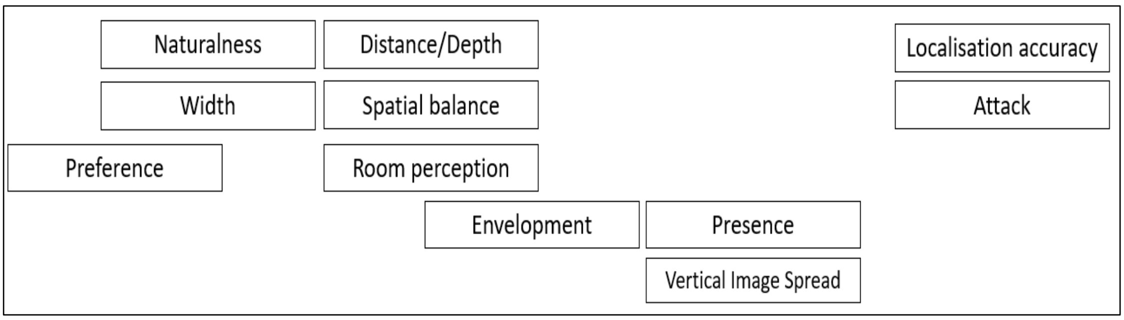


Figure 4.2.1: Correlations between Attributes based on their Response Patterns Part One

In summary, some significant observations can be made; firstly, the close relationship between the two attitudinal attributes Naturalness and Preference regarding their response patterns. They differ only slightly from each other in the rating of the handpan. Secondly, the almost identical response patterns of the section Naturalness and Width in relation to spatial descriptive attributes such as Distance/Depth, Spatial Balance and Room Perception. The two sections only differ from each other in the response for the djembe. Thirdly, the function of the attribute Envelopment as a link between the spatial descriptive attributes and the attitudinal attribute Presence. The attribute Envelopment only differs from each of the two by one source each (cello for the spatial descriptive attributes and handpan for Presence). Fourthly, although the response pattern for the attributes Localisation Accuracy and Attack is identical, they showed no similarity whatsoever to the responses of the other attribute-clusters. Regarding Timbre and Preference, no array seemed to dominate the highest rankings of all sources. Within these two attributes, no particular regularity could be found, as can be seen in Tables 4.2.1 and 4.2.2:

Source (Trial)	**Stimulus used for scale A**	**Stimulus used for scale B**	**Stimulus used for scale C**
Handpan (Trial 4)	thin	full, but lacking mids	• slightly nasal, • less full than B, but more homogeneous • natural • highest treble content in reverb
Cello (Trial 1)	full	• homogeneous • natural • highest treble content in reverb	• (very) thin • nasal
Djembe (Trial 3)	• homogeneous • natural • highest treble content in reverb	• full, but completely lacking mids • excessive bass	• canny/nasal • thin
Violin (Trial 2)	Treble strength neutral	• sharp • thin • nasal	• brilliance • highest treble content in reverb
Guitar (Trial 5)	• lacking mids • nasal • thin	• homogeneous • brilliant • full	• completely lacking midrange • very bright

Table 4.2.1: Results for the Timbral Attribute Category Scaling

Source (Trial)	**Stimulus used for scale A**	**Stimulus used for scale B**	**Stimulus used for scale C**
Guitar (Trial 5)	**Very open, yet still clear, wide pleasing sound**	Very precise but lacking space	Compromise between A and B
Violin (Trial 2)	**Very open sound, good D/R ratio, not sharp**	Not as sharp as C but not as wide and opened as A	Sharp, edgy, lacking space
Djembe (Trial 3)	Phase issues not as bad as B, but also not as realistic as C	Disturbing frequency, phase issues in the ambient sound with changing frequency	**Very realistic**
Cello (Trial 1)	Dislike timbre	**Balanced D/R ratio, further away**	Not as wide as B, lows too present, narrowing the picture unnaturally
Handpan (Trial 4)	Natural but less than C	No space	**Natural, real**

Table 4.2.2: Comments on the Grading for the Attribute Preference

The colour blue indicates the mix of the Main-Array, orange the mix of the F/R-Array and green the mix of the Hybrid-Array. Comments in bold indicate the comment of the preferred stimulus.

However, when comparing the comments on Preference (Table 4.2.2) with the timbral attribute category scaling (Table 4.2.1), a strong correlation between the highest preference ratings and positive timbral descriptors can be identified. The assigned timbral labels for the perceived auditory event in case of preference were *"homogeneous"*, *"natural"*, *"brilliant"* and containing *"highest treble content in reverb."*

This is a remarkable result as Preference was the first and Timbre the last attribute to be tested in the listening test. This indicates that expectation bias or sequential contraction bias can be excluded. Although according to Table 4.2.2 Timbre is suggested to have a strong influence on Preference, Room Perception and Naturalness also seemed of importance. Timbral factors were mentioned only in case of negative perception, whereas Naturalness appeared mostly as a positive descriptor. *"Space"* was noted in both, a positive and negative context.

When including the results of the timbral category scaling, the continuation of Figure 4.2.1 is depicted in Figure 4.2.2, and any implications will be discussed in Chapter 5:

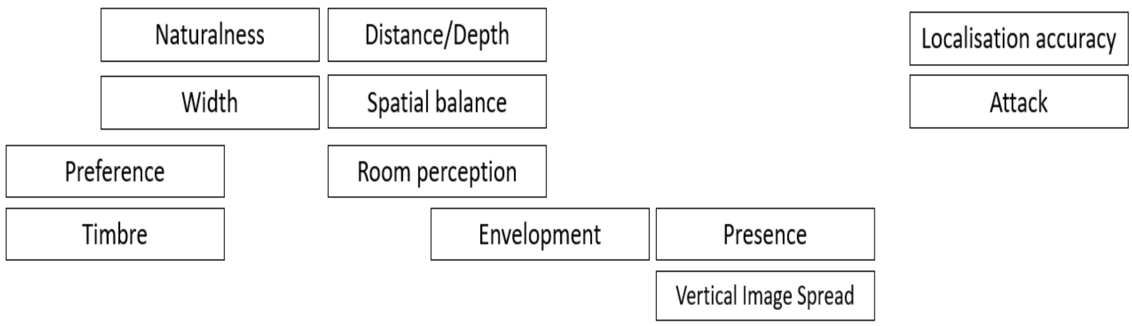


Figure 4.2.2: Correlations between Attributes based on their Response Patterns Part Two

4.3 Correlations between Array Gradings and Sources

Going one level deeper, Tables 4.3.1-4.3.3 give an overview of the different arrays being ranked highest for a specific instrument regarding a specific attribute:

Main-Array Highest Gradings:

Attribute	Source
Naturalness	Violin
Width	Violin
Preference	Violin
Distance/Depth	Violin
Spatial Balance	Violin
Room Perception	Violin
Envelopment	Violin, Cello
Presence	Violin, Cello
Vertical Image Spread	Violin, Cello
Localisation Accuracy	Djembe, Handpan, Guitar
Attack	Djembe, Handpan, Guitar

Table 4.3.1: Main-Array Highest Gradings

F/R-Array Highest Gradings:

Attribute	Source
Naturalness	Djembe, Handpan, Guitar
Width	Djembe, Handpan, Guitar
Preference	Djembe, Guitar
Distance/Depth	Handpan, Guitar
Spatial Balance	Handpan, Guitar
Room Perception	Handpan, Guitar
Envelopment	Handpan, Guitar
Presence	Guitar
Vertical Image Spread	Guitar
Localisation Accuracy	Violin, Cello
Attack	Violin, Cello

Table 4.3.2: F/R-Array Highest Gradings

Hybrid-Array Highest Gradings:

Attribute	Source
Naturalness	Cello
Width	Cello
Preference	Cello, Handpan
Distance/Depth	Cello, Djembe
Spatial Balance	Cello, Djembe
Room Perception	Cello, Djembe
Envelopment	Djembe
Presence	Djembe, Handpan
Vertical Image Spread	Djembe, Handpan
Localisation Accuracy	-
Attack	-

Table 4.3.3: Hybrid-Array Highest Gradings

At the same time, it was considered worthy of depicting results where ratings are critically below the values of the other arrays for the same attribute and source.

When doing so *"critically"* has been defined as a value below 40. Tables 4.3.4-4.3.6 show where such ratings occurred:

Main-Array Critical Gradings:

Attribute	Source
Naturalness	Djembe, Handpan, Guitar
Width	Djembe, Handpan, Guitar
Preference	Handpan
Distance/Depth	Djembe, Handpan, Guitar
Spatial Balance	Djembe, Handpan, Guitar (1!)
Room Perception	Djembe, Guitar
Envelopment	Djembe, Handpan, Guitar (1!)
Presence	Djembe, Handpan, Guitar
Vertical Image Spread	not applicable
Localisation Accuracy	Violin
Attack	-

Table 4.3.4: Main-Array Critical Gradings

F/R-Array Critical Gradings:

Attribute	Source
Naturalness	-
Width	-
Preference	Violin
Distance/Depth	-
Spatial Balance	Cello
Room Perception	-
Envelopment	Violin, Cello
Presence	-
Vertical Image Spread	not applicable
Localisation Accuracy	Handpan, Guitar
Attack	Guitar

Table 4.3.5: F/R-Array Critical Gradings

Hybrid-Array Critical Gradings:

Attribute	Source
Naturalness	Djembe
Width	-
Preference	-
Distance/Depth	-
Spatial Balance	-
Room Perception	-
Envelopment	-
Presence	-
Vertical Image Spread	not applicable
Localisation Accuracy	Djembe
Attack	-

Table 4.3.6: Hybrid-Array Critical Gradings

Based on these results a source-dependent array behaviour can be assumed, and several tendencies stick out when observing Tables 4.3.1-4.3.3:

First, the Main-Array dominates the high gradings for the violin regarding all attributes. Exceptions are Localisation Accuracy and Attack where the Main- and F/R-Array seemed to have swopped their behaviour (see Table 4.3.1 and 4.3.2). The Main-Array also prevails the ratings for the violin and cello regarding the attributes Envelopment, Presence and Vertical Image Spread. Similarly, the F/R-Array dominated positive responses for the guitar except for Localisation Accuracy and Attack. Almost to the same extent the F/R-Array had led to positive responses for the handpan. Analog to these examples, the Hybrid-Array was responsible for most high ratings for the cello and djembe.

Considering the sonic nature of the instruments when describing source-array dependencies, the following regularities were discovered (see Table 4.3.7): The Main-Array seems to dominate primarily the responses of the violin, which exhibits a sustained HF character. Besides, it also featured sustained sources with different frequency content (violin/cello). Analog, the F/R-Array seems to prevail sources mostly being active in the mid-frequency range and being of both, a sustained and percussive nature. The Hybrid-Array, on the other hand, seems to have the most influence on positive ratings of instruments with enhanced LF content.

Highest Gradings:

Array Type	Instrument
Main-Array	Violin, Cello / Violin
F/R-Array	1. Guitar, 2. Handpan
Hybrid-Array	Cello, Djembe (equal)

Table 4.3.7: Relation between Array Type and Highest Scored Instruments

Critical Gradings:

Array Type	Instrument
Main-Array	1. Djembe, Handpan, Guitar (all) 2. Violin (Localisation Accuracy)
F/R-Array	Violin (Preference), Cello (Spatial Balance), Envelopment: Violin/Cello Handpan, Guitar (Localisation Accuracy)
Hybrid-Array	Djembe (Naturalness, Localisation Accuracy)

Table 4.3.8: Relation between Array Type and Critical Gradings

The most obvious pattern within Table 4.3.8 is the Main-Array leading to negative responses mostly for instruments entailing a percussive element. An interesting note is also the negative rating for the violin for Localisation Accuracy, as for all other attributes the same array resulted in the highest ratings for the violin. Secondly, a minor tendency of the F/R-Array towards critical gradings on string instruments can be assumed. Last but not least, the only instrument where the Hybrid-Array resulted in critical responses was the djembe.

Regarding the attributes Vertical Image Shift and Vertical Frequency Separation the only pattern to be discovered was that these phenomena appear almost exclusively in the FR- and Hybrid-Array, which both contain the height layer of the Hamasaki-Cube. For the attributes Punch and Powerful, the F/R-Array seemed to dominate positive ratings.

4.4 Objective Measures

When comparing the outcomes of Chapter 4.2 and 4.3 with the spectral graphs derived from the dummy head and ambient microphone signals no deductions could be drawn. The same applied to the IRs and RT60 measurements. The only regularity to be discovered was an enhanced HF content in the signals of the Hamasaki-Cube compared to the ambient signals of the Main-Array and will be discussed in the next chapter.

5. Discussion

5.1 Correlation between Attributes and Subject Responses

The relationships between attributes as seen in Table 4.2.2 are confirmed by previous research and will be discussed in this chapter. Since the attribute clusters in Table 4.2.2 are based on the rankings of the arrays for the different sources (see Chapter 4.1), a close correlation between the attributes implicates an identical or strongly correlated response pattern for the different arrays. Hence, when the attribute correlations can be backed up by previous research, the validity of the array gradings can be seen increased. Naturalness, for example, was proofed in experiments

> *"to be by far the most important factor in determining overall preference in sound quality. Possibly it is mainly an evaluative or emotive judgement, it may consist of an optimum combination of other sub-factors, and it may have a strong timbral component and be highly context dependent."* (Rumsey, 2013, p. 39)

Hence, this statement confirms the close relation between Naturalness, Preference and Timbre as displayed in Table 4.2.2. It also explains the frequent appearance of the descriptor *"natural"* in the comments on Preference (see Table 4.2.2). Related to that, if one equals the attribute Preference with the highest overall judgement of audio quality, Preference can be seen as "*integrative evaluation that takes into account all of the lower-level attributes and weighs up their contribution."* (Rumsey as in Roginska & Geluso, 2018, p. 215) As it was shown that timbral fidelity contributes strongly more to the overall quality judgment than spatial fidelity (Rumsey, 2013, p. 39; Rumsey *et al.*, 2005, p. 968), the identical response pattern for Preference and positive timbral descriptors leading to a closer correlation of Preference to Timbre than Preference to any spatial descriptive attributes (see Table 4.2.2) can be explained this way.

The positive timbral descriptors used when describing the auditory perception for stimuli which independently had been graded highest in Preference were *"homogeneous"*, *"natural"*, *"brilliant"* and *"highest treble content in reverb."* Their definitions as used during the listening test have been taken and contrasted with the spectral graphs of the dummy head and ambient microphone signals. However, no conclusions could be drawn as the graphs where visually too similar to identify significant differences in the frequency areas of concern.

When considering that Naturalness is one of the most critical factors for Preference and has been defined as *“how similar to a natural listening experience the sound as a whole sounds,”* a possible explanation for the negative correlation between negative timbral perception and Preference (see Table 4.2.2) could be seen in the following statement by Rumsey:

> *“There are, nonetheless, occasional phenomena that might be considered as specifically associated with reproduced sound, being rarely or never encountered in natural environments. The one that springs most readily to mind is the 'out of phase' phenomenon…Anomalies in signal processing or microphone technique can create such effects and they are unique to reproduced sound, so there is in effect no natural anchor or reference point against which to compare these experiences.”* (2013, p. 39)

Unlike Timbre, Naturalness appeared mostly as a positive descriptor in the comments for Preference (see Table 4.2.2), and could be explained through the concept that *“Naturalness can be taken as a comparison between the stimulus under evaluation and an internal reference that relates to memories of the characteristics of natural environments.”* (Rumsey, 2002, p. 654)

Furthermore, the perception of space as a further influential factor for Preference was confirmed by Toole who found that increased quality of spatial ratings can significantly influence the overall sound quality rating (1985, p. 2). In that regard, the descriptor *“open”*, which frequently appears in the comments on Preference, contributes to *“a feeling of space”* (see Table 4.2.2) and leads to higher ratings for Naturalness and Preference (Rumsey, 2013, p. 42). Hence, the rather close relation of spatial descriptive attributes to the cluster of Naturalness and Preference in Table 4.2.2 can be seen confirmed.

Having backed up the strong correlation of the array ratings between Preference, Naturalness and Timbre, their correlation to the attribute Width must be included in the discussion when following the structure in Table 4.2.2. In concert hall acoustics ASW has been associated with positive listener responses (Rumsey, 2002, p. 659). According to the definitions of the current study ASW could be considered equal to “individual source width”, which is a sub-attribute of Width. Therefore, the concept of ASW could be seen to back up the correlation between Width and Preference. In relation to that, the stronger correlation between Width and Preference compared to Depth and Preference was previously investigated by Rumsey:

“By far the stronger perception seems to be environment width [compared to environmental depth], for reasons not yet explained, although it may have to do with the concept of construct masking, in which strong perceptual constructs have a tendency to dominate the overall judgment, thereby hiding weaker ones.” (2002, p. 660)

Although environmental depth and source distance seem to be dominated by the perception of environmental width, Depth is still crucial to the appreciation of sound quality (Rumsey, 2013, p. 35). This explains why the array gradings for Depth are not as correlated to the array gradings of Preference as are the gradings for Width, but can still be found in the same cluster area, referring to Table 4.2.2.

Another phenomenon to explain is the correlation between the spatial descriptive attributes, Envelopment and Presence. The link of these attributes to Preference, Naturalness and so on can be established in the connection between Width and Room Perception (Rumsey 2002, p. 659). From there, on one side, Rumsey found a correlation between Envelopment, Room Perception and Spatial Balance: *“Environmental envelopment appears to be related to the background information stream, in reproductions of natural spaces being dependent on the level and directional distribution of late, diffuse reverberant energy, similar to the concert hall LEV.”* (*ibid.*, p. 663) On the other side, he declares that *“presence and environmental envelopment are not necessarily the same, although they may be closely related.”* (*ibid.*) Combining these statements, it can be argued that the function of Envelopment as a link between spatial descriptive attributes and Presence in Table 4.2.2 can be seen confirmed.

Last but not least the isolated cluster of Localisation Accuracy and Attack requires further explanation. Presumably, a *“clear transient response”* as by definition in the listening test would lead to a better Localisation Accuracy when thinking about the high portion of direct sound this conditions (Rumsey, 2013, p. 34). This might explain why these attributes achieved an identical response pattern. The outcome that their response pattern seems uncorrelated to all other attributes could be confirmed in the findings of Berg & Rumsey having proved *“that localisation in itself is not the attribute closest to naturalness and positive sensations.”* (2000, p.12)

Therefore, by backing up the correlations found between the attributes as depicted in Table 4.2.2 the nature of the response patterns for the different arrays could be validated.

5.2 Expected and Unexpected Results

5.2.1 Expected Results

Based on the explanations given in Chapter 2 and the psychoacoustic principles the Bowles-Array and Fukada-Tree/Hamasaki-Cube configuration operate on (see Appendix, Chapters 9.1.1, 9.1.2, 9.2.1, 9.2.2 and 9.5) some assumptions could be made on the outcomes of the listening test regarding the behaviour of the different arrays. These are summarised in Table 5.2.1:

Main-Array (Bowles-Array)	**FR-Array (Fukada-Tree/Hamasaki-Cube)**	**Hybrid (Bowles Array Lower Layer/ Hamasaki-Cube Height -Layer)**
Enhanced perception of Naturalness	Enhanced Envelopment	
Enhanced perception of Presence	Enhanced Perception of (environmental) Width	
Reduced Spatial Balance	Enhanced Spatial Balance	Risk of perceived separation between the main and the height layer
Preferred perception for Timbre	Enhanced Room Perception	
Enhanced potential of VIS	Reduced possibility of VIS	Reduced possibility of VIS
Enhanced risk of vertical phantom image shift if the localisation threshold is exceeded during mixing	Reduced risk of vertical phantom image shift	Reduced risk of vertical phantom image shift
Limited perception of Distance/Depth	More stable sound source localisation in centre	Limited perception of Distance/Depth

Table 5.2.1: Assumptions on the Outcomes of the Listening Test for the Different Array Types

Besides, there was a possibility of tonal colouration for all three arrays although the nature of these colourations was unknown by the time to the authors' knowledge (see Chapter 2.3).

5.2.2 Source-Dependent Array Behaviour

As outlined in the hypothesis and depicted in Table 5.2.1, it was assumed that each array would dominate the high ratings for all sources for specific attributes. This assumption is based on preliminary research proving that these arrays operate based on different stereophonic and psychoacoustic principles (see Chapter 2.3 and Appendix Chapters 9.1.1, 9.1.2, 9.2.1, 9.2.2 and 9.5). Therefore, the outcome of the listening test indicating that no array dominates the highest scores for all sources for a specific attribute (or at least four of the five sources) and therefore giving the results a source-dependent character, was not expected and disproved the hypothesis. Although some regularities regarding frequency content and acoustic envelopes could be identified amongst the sources, a thorough explanation of the source-dependent results requires further experiments with a more controlled experimental design. Such experiments would most likely be of a lower *"ecological validity"* (Rumsey, 2002, p. 654) but may give some indications about this phenomenon on a psychoacoustic level. As the listening test results could be explained to some extent by means of previous psychoacoustic findings, this may be a direction worth trying.

When investigating the source-dependence in the psychoacoustic realm, research indicates that the radiation pattern of the different sources could be a factor which could have influenced the current results. Martin *et al.* proved that the radiation pattern impacts the instrument's perceived audio image whereas the non-coincident arrays featured the most irregular source image perceptions (2016, p. 5). This is worth mentioning as in the current study only non-coincident arrays have been applied. Although the research of Martin *et al.* was only concerned about imaging, the diverse perception of the source images within the same arrays indicates that also other attributes could be affected by radiation patterns. Even if this approach should prove to be unable to explain the current results, the insight gained therein could provide a better understanding of 3D recording:

"Our traditional understanding of how musical instruments radiate sound is still based on a paradigm of a two-dimensional plane; re-learning sound radiation in a three-dimensional paradigm will help understand height channel recording better." (Bowles, 2015, p. 3)

5.2.3 Fulfilled and Unfulfilled Expectations

Although the patterns of the listening test were found to be consistent, confirmed by research (see Chapter 5.1), and intra-subject consistency is assumed (see Chapter 3.5), there is a rather distinctive deviation from the expectations of the different arrays for specific attributes, as can be seen in Table 5.2.2. When creating Table 5.2.2, the average rating of a specific array for all sources of a chosen attribute was taken to label the expectation as fulfilled or unfulfilled. The numerical proximity to the second highest average score was not considered.

Main-Array (Bowles-Array)	FR-Array (Fukada-Tree/Hamasaki-Cube)	Hybrid (Bowles Array Lower Layer/ Hamasaki-Cube Height -Layer)
Enhanced perception of Naturalness	Enhanced Envelopment	
Enhanced perception of Presence	Enhanced Perception of (environmental) Width	
Reduced Spatial Balance	Enhanced Spatial Balance	Risk of perceived separation between the main and the height layer
Preferred perception for Timbre	Enhanced Room Perception	
Enhanced potential of VIS	Reduced possibility of VIS	Reduced possibility of VIS
Enhanced risk of vertical phantom image shift if the localisation threshold is exceeded during mixing	Reduced risk of vertical phantom image shift	Reduced risk of vertical phantom image shift
Limited perception of Distance/Depth	More stable sound source localisation in centre	Limited perception of Distance/Depth

Table 5.2.2: Fulfilled and Unfulfilled Expectations. The colour green indicates fulfilled, and the colour red unfulfilled expectations.

However, when having a closer look at the listening test results, including the proximity of the array values to each other, the situation is more complicated than shown in Table 5.2.2. Some peculiarities could be discovered for specific attributes and are worth discussing in the next chapters.

5.2.4 Vertical Image Spread, Vertical Image Shift and Vertical Frequency Separation

The unexpected result of the FR- and Hybrid-Arrays dominating the perception of Vertical Image Spread, Vertical Image Shift and Vertical Frequency separation, whereas all these attributes have been assigned initially exclusively to the Main-Array, could be explained when having a look at psychoacoustics. The spectral graphs of the Hamasaki-Cube signals indicate a slightly enhanced HF-content compared to the ambient signals of the Main-Array. As the FR- and Hybrid-Array both contained the signals of the Hamasaki-Cube for their height layer, this could have caused a pitch-height effect (see Appendix, Chapter 9.5.7), which may have led to the perception of Vertical Image Spread, Shift or Frequency Separation.

The scenario of an exceeded localisation threshold during mixing leading to vertical ICCT and thus to these effects (see Appendix Chapter 9.5.3 and 9.5.5) can be considered unlikely, as the Hamasaki-Cube is optimised to capture mainly ambient sound (see Appendix, Chapter 9.4.2) and vertical ICCT by definition conditions a certain amount of direct sound in the height layer (see Appendix, Chapter 9.5.3).

A remaining question here would be why some of these effects have been observed in the Main-Array for string instruments, but no other sources. As the ICCT leading to these effects depends on the ICLDs between the main and height layer signals (see Appendix

Chapter, 9.5.3 and 9.5.5), this would indicate the option that the ICLDs for string instruments were smaller than for other sources. However, the reason for this would be currently unknown to the author.

5.2.5 Naturalness and Width

Opposed to Chapter 2.2.2 where the Main-Array was claimed to convey an enhanced perception of Naturalness, in fact, only one source scored highest for this attribute for the Main-Array (see Chapter 4.1). This implies that the FR- and Hybrid-Array dominated the attribute Naturalness. Referring again to Table 4.2.2, a possible reason could be that the highly decorrelated signals of the Hamasaki-Cube (see Appendix, 9.1.1 and 9.2.2), being a part of both, the FR- and Hybrid-Array, lead to a decreased IAC (see Appendix Chapter 9.5.1) and thus to an increased ASW (Gribben & Lee, 2018, p. 537). ASW can be considered being part of the attribute Width as per current definition, which in Table 4.2.2 resulted in the same response pattern as Naturalness. Having said this, the expected result of the F/R-Array achieving the highest scores for Width may be explained the same way.

5.2.6 Room Perception, Spatial Balance and Distance/Depth

Although it was the Hybrid-Array scoring highest for Room Perception, Spatial Balance and Distance/Depth, and not the F/R-Array, as expected, it can be said that the Main-Array scored considerably lower than any of the other two arrays for these attributes, even if this was expected. It could be assumed, that since both, the FR- and the Hybrid-Array share the common ground of the Hamasaki-Cube height layer, that a possible explanation for this outcome could be the way early reflections have been captured. Early reflections play a crucial role in spatial hearing:

> *"When it comes to recording, this portion of reflected sound [early reflections] deserves special attention as it critically affects attributes such as distance, depth, and spatial impression. The hearing takes spatial information from early reflections and converts it to a spatial event…With natural sound, the human ear performs this conversion spontaneously and with amazing robustness because that type of sound contains all properties of a reflection pattern in their original form. Key parameters include the timing structure in relation to direct sound, levels and spectrum, and horizontal and vertical incidence directions. Imaging a spatial environment is realistic when the ear is able to recognize and interpret the features of the reflected sound – that is, when it 'understands' the reflection pattern… Perception is particularly stable when the reflections come in from the original directions of the upper half space. Reproducing depth requires careful handling of early reflections."* (Theile & Wittek, 2012, p. 7)

The argument that "*perception is particularly stable when the reflections come in from the original directions of the upper half space"* got confirmed by Hamasaki & Van Baelen declaring that "*it is necessary to situate the rear and upper microphones in order to catch the reflections coming from behind and above a listener."* (2015, p. 3) Therefore, one could think that the increased height, distance and directivity of the Hamasaki-Cube microphones compared to the directivity and placement of the Main-Array height layer microphones (see Appendix Chapters 9.1.2, 9.2.2 and 9.6.4) would lead to a more distinctive capture of early reflection key parameters such as "*the timing structure in relation to direct sound, levels and spectrum, and horizontal and vertical incidence directions."* (Theile & Wittek, 2012, p. 7) As a consequence, this would ease (and thus probably increase) the perception of Room Perception, Spatial Balance and Distance/Depth. For a more detailed explanation of the influence of early reflections on the auditory perception of a sound event, the reader is referred to the Appendix (Chapter 9.1.2).

So far, the previous sections could give some possible explanations for the difference in behaviour between the Main-Array and the FR- / Hybrid-Arrays regarding these attributes. They, however, do not explain why by trend the Hybrid-Array scored higher than the F/R-Array. In a broader context this leads to the question of how the Hybrid-Array entailing omnidirectional microphones in the main layer and thus having no directivity in its capture could score higher than the F/R-Array containing the Hamasaki-Cube main layer which is optimised for capturing lateral early reflections. This is an unexpected outcome as the increased importance of lateral early reflections compared to ceiling reflections for the perception of Room Perception was originally found by Barron:

> *"'Spatial impression' [as by definition in the study equal to Room Perception in the current investigation] was produced for reflection delays between 10 and 80msec by lateral rather than ceiling reflections…It was concluded that the degree of spatial impression is probably related to the ratio of lateral to non-lateral sound arriving within 80msec of direct sound."* (1971, p. 475)

It has to be mentioned, however, that the experimental design of Barron entailed *"simulated reflections in an anechoic chamber in an attempt to understand the importance of early reflections in a concert hall"* (*ibid.*) and thus differs from the current context, where the recording took place in diffuse field conditions.

5.2.7 Envelopment

Based on the same arguments as in Chapter 5.2.6 (and further explanations in the Appendix Chapters 9.1.1, 9.1.2, 9.2.2 and 9.4.2) no explanation could be found why the Hybrid-Array scored higher in average than the F/R-Array for Envelopment. Although it is not early lateral reflections influencing the perception of Envelopment, but late lateral reflected energy (Theile & Wittek, 2012, p. 6; Rumsey, 2002, p. 661) the parameter of directional lateral capture remains the same. What could be explained instead is the positive correlation of average scores between Envelopment and Spatial Balance: Hanyu *et al.* found that Envelopment *"increases if there is adequate spatial balance in the direction of arriving reflections."* (1999, cited in Howie *et al.*, 2016, p. 2) Consequently, as the Hybrid-Array was perceived to have the highest degree of Spatial Balance, it thus might also have been perceived as most enveloping.

5.2.8 Presence

Like in the case of Envelopment (see Chapter 5.2.7) the dominance of the Hybrid-Array for Presence could indirectly be explained employing the perception of Spatial Balance: *"An important criterion for presence is hypothesized to be an awareness of background-stream sound energy arriving from many directions."* (Rumsey, 2002, p. 663) Therefore it follows that the increased perception of Spatial Balance of the Hybrid-Array might have led to an increased perception of Presence compared to the other arrays. Based on the rather high ratings for Presence of both, the Hybrid- and F/R-Array compared to the Main-Array, it can only be hypothesised that the increased treble content in the Hamasaki-Cube height layer has contributed to the perception of *"realism"* (Gerzon, 1971, p. 659) and thus to Presence. If this would be the case, however, the question would arise why the Hybrid-Array was scored higher than the F/R-Array as the F/R-Array also contains the Hamasaki-Cube main layer signals, which have been shown to exhibit a slight increase of HF-content compared to the Hybrid-Array main layer signals.In any case, the results regarding this attribute stand in contradiction with the claim that one of the main advantages of the Main-Array is its ability to convey Presence (see Chapter 2.2).

5.2.9 Timbral Colouration

Last but not least, no direct indications could be derived from the spectral graphs of the dummy head recordings about the nature of the possible timbral colourations of the different arrays as outlined in Chapter 2.3. Similar to the suggested further research in Chapter 5.2.2, it is proposed that an experimental design with less uncontrolled variables should be applied to approach this complex matter, similar as in Robotham *et al.* (2016, p. 2).

6. Conclusion

In summary, the hypothesis that the arrays in concern will produce recordings that shall lead each to an increased perception of specific attributes for all sources tested was disproved.

Instead, the outcomes indicate a source-dependent character. When having a look at the source-dependencies, the instruments featuring the highest and lowest scores of each array could be described based on their frequency content or acoustic envelope. As the objective measures gathered in the current experiment did not allow for any conclusions regarding this matter and previous research suggests an influence of radiation patterns on attribute perception further experiments focussing on this phenomenon were suggested. In addition, the following observations have been made:

When collating the listening test results based on their similarity, the obtained pattern of their corresponding attributes was either identical or highly correlated with the findings of psychoacoustic theories dealing with the correlation of these attributes with regards to their auditory perception. This was interpreted as an increase in the validity of the listening test results. Furthermore, when having a closer look at possible reasons for the deviation of the expected results several hypotheses have been derived:

Firstly, the slightly enhanced HF-content captured by the Hamasaki-Cube could have caused a pitch-height effect in the Hybrid and F/R-Array which in turn led to an enhanced perception of the attributes Vertical Image Spread, Vertical Image Shift or Vertical Frequency Separation. At the same time, the highly decorrelated signals of the Hamasaki-Cube might have contributed to the increased perception of Width and Naturalness for both, the FR- and Hybrid-Array compared to the Main-Array. This was unexpected as the strength of the Main-Array has been claimed to be the conveyance of Naturalness. As the main layer of the F/R-Array entails the signals of the Hamasaki-Cube main layer, the dominance of the F/R-Array compared to the Hybrid-Array for these attributes could be explained through the lower decorrelation of its signals in the main layer.

Besides, it could be argued that the increased height, distance and directivity of the Hamasaki-Cube could have led to a more distinctive capture of early reflections which in turn might have caused the high scores for Room Perception, Spatial Balance and Distance/Depth for the FR- and Hybrid-Array. In relation to that, it could be assumed that the increased directional capture of early reflections by the Hamasaki-Cube presumably leading to a higher Spatial Balance might explain the higher scores of the same arrays for Envelopment and Presence, as these attributes have been found depending on the perception of Spatial Balance. Another reason for the enhanced perception of Presence in the Hybrid and F/R-Array compared to the Main-Array could be traced back to the enhanced HF-content in their height signals, which has been claimed to enhance the perception of *"realism."* Either way, the outcomes for Presence deviated from the expectation of the Main-Array conveying the highest perception for this attribute.

Opposed to that the averaged higher scores of the Hybrid-Array compared to the F/R-Array for Room Perception, Spatial Balance, Distance/Depth, Envelopment and Presence could not be explained, as they contradict previous psychoacoustic research stressing the importance of the directive capture of lateral reflections for the perception of these attributes. In addition to the further research proposed in Chapter 5, an in-depth investigation on the influence of lateral early and late reflections in a diffuse field 3D recording scenario is therefore suggested.

On the other side, the high importance of Timbre has been confirmed throughout this investigation, as positive timbral descriptors have been found strongly linked to Preference. Based on previous research indicating the influence of vertical reflections on Timbre and Preference, an auditory evaluation using category scaling with more expert subjects could give further indications of a possible correlation between the individual array parts and their influence on Timbre.

This work has been useful in gaining an understanding of the spatial and timbral perception of a Bowles-Array, a Fukada-Tree/Hamasaki-Cube configuration and their hybrid version when compared against each other. The report of the different experimental techniques and the discussion of their outcomes gave further indications on how individual array parts might have contributed to the perception of specific attributes. This insight could be seen as a valuable basis for recording engineers experimenting with 3D recording techniques for informing some of their decisions. Although objective measures could explain only a few results at this stage, the contribution of this research can be seen as one of many to be done in order to derive a *"perceptual 'handbook'"* for the manipulation of specific attributes in the realm of 3D audio recording techniques.

7. Reference List

Auro-3D® Home Theater Setup Installation Guidelines. (2015). [pdf] Mol: Auro Technologies. Available at: https://www.auro-3d.com/wp-content/uploads/documents/Auro-3D-Home-Theater-Setup-Guidelines_lores.pdf [Accessed 10 Feb. 2019].

Barron, M. (1971). The subjective effects of first reflections in concert halls—The need for lateral reflections. *Journal of Sound and Vibration*, [online] 15(4), pp.475-494. Available at: https://www.researchgate.net/profile/Mike_Barron2/publication/222627645_The_subjective_effects_of_first_reflections_in_concert_halls-The_need_for_lateral_reflections/links/5a0b23e2458515e482747bf3/The-subjective-effects-of-first-reflections-in-concert-halls-The-need-for-lateral-reflections.pdf?origin=publication_detail [Accessed 9 Feb. 2019].

Bartlett, B. (2017). *Size Matters: The Differences In Large- And Small-Diaphragm Condenser Microphones - ProSoundWeb*. [online] ProSoundWeb. Available at: https://www.prosoundweb.com/channels/recording/size_matters_the_differences_in_large-_and_small-diaphragm_microphones/ [Accessed 6 Feb. 2019].

Bech, S. and Zacharov, N. (2006). *Perceptual Audio Evaluation - Theory, Method and Application*. Chichester: Wiley.

Berg, J. and Rumsey, F. (2006). Identification of Quality Attributes of Spatial Audio by Repertory Grid Technique. *Journal of the Audio Engineering Society*, [online] 54(5), pp.365-379. Available at: http://www.aes.org.saeezproxy.idm.oclc.org/e-lib/browse.cfm?elib=13681 [Accessed 17 Feb. 2019].

Berg, J. and Rumsey, F. (2003). Systematic Evaluation of Perceived Spatial Quality. In: *AES 24th International Conference*. [online] Banff. Available at: http://www.aes.org/e-lib/browse.cfm?elib=12272 [Accessed 25 Mar. 2019].

Berg, J. and Rumsey, F. (2000). Correlation between Emotive, Descriptive and Naturalness Attributes in Subjective Data Relating to Spatial Sound Reproduction. In: *109th International AES Convention*. [online] New York. Preprint No. 5206. Available at: http://www.aes.org/e-lib/browse.cfm?elib=9132 [Accessed 3 Apr. 2019].

Bowles, D. (2015). A microphone array for recording music in surround-sound with height channels. In: *139th International AES Convention*. [online] New York. Preprint No. 9430. Available at: http://www.aes.org/e-lib/browse.cfm?elib=17986 [Accessed 12 Jan. 2019].

Conetta, R., Brookes, T., Rumsey, F., Zielinski, S., Dewhirst, M., Jackson, P., Bech, S., Meares, D. and George, S. (2014). Spatial Audio Quality Perception (Part 1): Impact of Commonly Encountered Processes. *Journal of the Audio Engineering Society*, [online] 62(12), pp.831-846. Available at: https://secure.aes.org/forum/pubs/journal/?elib=17557 [Accessed 17 Feb. 2019].

Ellis-Geiger, R. (2016). Music Production for Dolby Atmos and Auro-3D. In: *141st International AES Convention*. [online] Los Angeles. Preprint No. 9675. Available at: http://www.aes.org/e-lib/browse.cfm?elib=18479 [Accessed 12 Jan. 2019].

Faulkner, T. (2019a). *All Saints Tooting for String Quartet?*. [email].

Faulkner, T. (2019b). *LCR Stereo Recording?*. [email].

Francombe, J., Brookes, T. and Mason, R. (2015). Elicitation of the Differences between Real and Reproduced Audio. In: *138th International AES Convention*. [online] Warsaw. Preprint No. 9307. Available at: http://www.aes.org/e-lib/online/browse.cfm?elib=17731 [Accessed 25 Mar. 2019].

Geluso, P. (2012). Capturing Height: The Addition of Z Microphones to Stereo and Surround Microphone Arrays. In: *132nd International AES Convention*. [online] Budapest. Preprint No. 8595. Available at: http://www.aes.org/e-lib/browse.cfm?elib=16233 [Accessed 12 Jan. 2019].

Gerzon, M. (1971). *Whither Four Channels?*. [pdf] Croydon: Link House Publications. Available at: https://www.michaelgerzonphotos.org.uk/articles/Whither.pdf [Accessed 25 Mar. 2019].

Gribben, C. and Lee, H. (2018). The Frequency and Loudspeaker-Azimuth Dependencies of Vertical Interchannel Decorrelation on the Vertical Spread of an Auditory Image. *Journal of the Audio Engineering Society*, [online] 66(7/8), pp.537-555. Available at: http://www.aes.org/e-lib/browse.cfm?elib=19705 [Accessed 13 Jan. 2019].

Gribben, C. and Lee, H. (2015). Towards the Development of a Universal Listening Test Interface Generator in Max. In: *138th International AES Convention*. [online] Warsaw. E-Brief No. 187. Available at: http://www.aes.org/e-lib/browse.cfm?elib=17622 [Accessed 13 Jan. 2019].

Griesinger, D. (1999). *The Science of Surround*. [pdf] Available at: http://www.davidgriesinger.com/aes99.pdf [Accessed 9 Feb. 2019].

Griesinger, D. (1998). General Overview of Spatial Impression, Envelopment, Localization, and Externalization. In: *15th International AES Conference*. [online]. Preprint No. 15-013. Available at: http://www.aes.org.saeezproxy.idm.oclc.org/e-lib/inst/search.cfm [Accessed 8 Feb. 2019].

Hamasaki, K., Shinmura, T., Akita, S. and Hiyama,, K. (2001). Approach and Mixing Technique for Natural Sound Recording of Multichannel Audio. In: *19th International AES Conference*. [online] Preprint No. 1878. Available at: http://www.aes.org/e-lib/browse.cfm?elib=10111 [Accessed 9 Feb. 2019].

Hamasaki, K. and Van Baelen, W. (2015). Natural Sound Recording of an Orchestra with Three-dimensional Sound. In: *138th International AES Convention*. [online] Warsaw. Preprint No. 9348. Available at: http://www.aes.org/e-lib/browse.cfm?elib=17772 [Accessed 12 Jan. 2019].

Hanyu, T., Kimura, S. and Chiba, S. (1999). A New Objective Measure For Evaluation of Listener Envelopment Focusing on the Spatial Balance of Reflections. *Journal of Architecture and Planning (Transactions of AIJ)*, [online] 64(520), pp.9-16. Available at: https://www.researchgate.net/publication/313326738_A_NEW_OBJECTIVE_MEASURE_FOR_EVALUATION_OF_LISTENER_ENVELOPMENT_FOCUSING_ON_THE_SPATIAL_BALANCE_OF_REFLECTIONS [Accessed 9 Feb. 2019].

Holman, T. (2008). *Surround Sound: Up and Running*. 2nd ed. Oxon: Focal Press.

Howie, W. and King, R. (2015). Exploratory microphone techniques for three-dimensional classical music recording. In: *138th International AES Convention*. [online] Warsaw. E-Brief No. 196. Available at: http://www.aes.org/e-lib/browse.cfm?elib=17631 [Accessed 12 Jan. 2019].

Howie, W., King, R., Boerum, M., Benson, D. and Han, A. (2015). Listener preference for height channel microphone polar patterns in three-dimensional recording. In: *139th International AES Convention*. [online] New York. Preprint No. 9419. Available at: http://www.aes.org/e-lib/browse.cfm?elib=17975 [Accessed 12 Jan. 2019].

Howie, W., King, R. and Martin, D. (2016). A Three-Dimensional Orchestral Music Recording Technique, Optimized for 22.2 Multichannel Sound. In: *141st International AES Convention*. [online] Los Angeles. Preprint No. 9612. Available at: http://www.aes.org/e-lib/browse.cfm?elib=18416 [Accessed 12 Jan. 2019].

Howie, W., King, R., Martin, D. and Grond, F. (2017). Subjective Evaluation of Orchestral Music Recording Techniques for Three-Dimensional Audio. In: *142nd International AES Convention*. [online] Berlin. Preprint No. 9797. Available at: http://www.aes.org/e-lib/browse.cfm?elib=18671 [Accessed 12 Jan. 2019].

Howie, W., Martin, D., Benson, D., Kelly, J. and King, R. (2018). Subjective and objective evaluation of 9ch three-dimensional acoustic music recording techniques. In: *International Conference on Spatial Reproduction – Aesthetics and Science*. [online] Tokyo. Available at: http://www.aes.org/e-lib/browse.cfm?elib=19630 [Accessed 25 Mar. 2019].

Izhaki, R. (2018). *Mixing Audio*. 3rd ed. New York: Focal Press.

Kassier, R., Lee, H., Brookes, T. and Rumsey, F. (2005). An Informal Comparison Between Surround-Sound Microphone Techniques. In: *118th International AES Convention*. [online] Barcelona. Preprint No. 6429. Available at: http://www.aes.org/e-lib/browse.cfm?elib=13145 [Accessed 5 Feb. 2019].

King, R. (2016). *Recording Orchestra and Other Classical Music Ensembles*. New York: Routledge.

King, R., Howie, W. and Kelly, J. (2016). A Survey of Suggested Techniques for Height Channel Capture in Multi-channel Recording. In: *140th International AES Convention*. [online] Paris. E-Brief No. 266. Available at: http://www.aes.org/e-lib/browse.cfm?elib=18170 [Accessed 12 Jan. 2019].

Koivuniemi, K. and Zacharov, N. (2001). Unravelling the Perception of Spatial Sound Reproduction: Language Development, Verbal Protocol Analysis and Listener Training.

In: *111th International AES Convention*. [online] New York. Preprint No. 5424. Available at: http://www.aes.org/e-lib/browse.cfm?elib=9815 [Accessed 3 Apr. 2019].

Lee, H. (2019). *Microphone Angle Calculations/ICLD*. [email].

Lee, H. (2018a). *Hyunkook Lee: Psychoacoustics of 3D sound recording techniques: Research and Practice*. [online] YouTube. Available at: https://www.youtube.com/watch?v=5SsQdfKAhVA [Accessed 14 Jan. 2019].

Lee, H. (2018b). *Psychoacoustics of 3D Sound Recording: Research and Practice*. [pdf] University of Huddersfield. Available at: https://www.researchgate.net/publication/325429307_Psychoacoustics_of_3D_Sound_Recording_Research_and_Practice [Accessed 10 Feb. 2019].

Lee, H. (2011). The Relationship between Interchannel Time and Level Differences in Vertical Sound Localisation and Masking. In: *131st International AES Convention*. [online] New York. Preprint No. 8556. Available at: http://www.aes.org/e-lib/browse.cfm?elib=16082 [Accessed 12 Jan. 2019].

Lee, H. and Gribben, C. (2014). Effect of Vertical Microphone Layer Spacing for a 3D Microphone Array. *Journal of the Audio Engineering Society*, [online] 62(12), pp.870-884. Available at: http://www.aes.org/e-lib/browse.cfm?elib=17560 [Accessed 12 Jan. 2019].

Lee, H. and Gribben, C. (2013). On the optimum microphone array configuration for height channels. In: *134th International AES Convention*. [online] Rome. E-Brief No. 93. Available at: http://www.aes.org/e-lib/browse.cfm?elib=16694 [Accessed 12 Jan. 2019].

Lee, H., Gribben, C. and Wallis, R. (2014). Psychoacoustic Considerations in Surround Sound with Height. In: *28th Tonmeistertagung VDT International Convention*. [online] Cologne. Available at: http://eprints.hud.ac.uk/id/eprint/23151/ [Accessed 12 Jan. 2019].

Lee, H., Johnson, D. and Mironovs, M. (2017a). An Interactive and Intelligent Tool for Microphone Array Design. In: *143rd International AES Convention*. [online] New York. E-Brief No. 390. Available at: http://www.aes.org/e-lib/browse.cfm?elib=19338 [Accessed 12 Jan. 2019].

Lee, H., Johnson, D. and Mironovs, M. (2017b). *MARRS Quick Start Guide*. [pdf] Huddersfield: Applied Psychoacoustics Lab (APL), University of Huddersfield. Available at: http://eprints.hud.ac.uk/id/eprint/33675/8/MARRS%20Quick%20Start%20Guide.pdf [Accessed 20 Jan. 2019].

Letowski, T. (1989). Sound Quality Assessment: Concepts and Criteria. In: *87th International AES Convention*. [online] New York. Preprint No. 2825. Available at: http://www.aes.org.saeezproxy.idm.oclc.org/e-lib/browse.cfm?elib=5869 [Accessed 17 Feb. 2019].

Lindberg, M. (2018a). [image] Available at: http://www.2l.no/media/2L-145_recording-sessions/2L-145_recording-sessions-21.jpg [Accessed 10 Feb. 2019].

Lindberg, M. (2018b). *Sennheiser PRO TALK | Morten Lindberg - Part 3 of 3 I Sennheiser*. [online] YouTube. Available at: https://www.youtube.com/watch?v=F88b8VRqeQg [Accessed 14 Jan. 2019].

Lindberg, M. (2015). 3D Recording with the “2L-cube”. *VDT*, [online] (3/2015), p.15. Available at: http://www.2l.no/artikler/2L-VDT.pdf [Accessed 12 Jan. 2019].

Luthar, M. and Maltezos, E. (2015). Polar Pattern Comparisons for the Left, Centre, and Right Channels in a 3-D Microphone Array. In: *139th International AES Convention*. [online] New York. E-Brief No. 206. Available at: http://www.aes.org/e-lib/browse.cfm?elib=17882 [Accessed 12 Jan. 2019].

Martin, B., King, R., Leonard, B., Benson, D. and Howie, W. (2015). Immersive Content in Three Dimensional Recording Techniques for Single Instruments in Popular Music. In: *138th International AES Convention*. [online] Warsaw. Available at: http://www.aes.org.saeezproxy.idm.oclc.org/e-lib/browse.cfm?elib=17675 [Accessed 18 Jan. 2019].

Martin, B., King, R. and Woszcyk, W. (2016). Subjective Graphical Representation of Microphone Arrays for Vertical Imaging and Three Dimensional Capture of Acoustic Instruments, Part I. In: *141st International AES Convention*. [online] Los Angeles. Preprint No. 9613. Available at: http://www.aes.org/e-lib/browse.cfm?elib=18417 [Accessed 12 Jan. 2019].

Millns, C. and Lee, H. (2018). An Investigation into Spatial Attributes of 360° Microphone Techniques for Virtual Reality. In: *144th International AES Convention*. [online] Milan. Preprint No. 10005. Available at: http://www.aes.org/e-lib/browse.cfm?elib=19401 [Accessed 13 Jan. 2019].

Nakayama, T., Tanetoshi, M., Osamu, K., Okamoto, M. and Shiga, T. (1971). Subjective Assessment of Multichannel Reproduction. *Journal of the Audio Engineering Society*, [online] 19(9), pp.744-751. Available at: http://www.aes.org/e-lib/browse.cfm?elib=2135 [Accessed 17 Feb. 2019].

Pedersen, T. and Zacharov, N. (2015). The development of a Sound Wheel for Reproduced Sound. In: *138th International AES Convention*. [online] Warsaw. Preprint No. 9310. Available at: http://www.aes.org/e-lib/browse.cfm?elib=17734 [Accessed 25 Mar. 2019].

Pratt, R. and Doak, P. (1976). A subjective rating scale for timbre. *Journal of Sound and Vibration*, 45(3), pp.317-328.
Proceedings of the ICSA 2011. (2011). [DVD] Directed by G. Zielinsky, P. Lemmens, W. Dabringhaus, H. Wittek and J. Nettingsmeier. Berlin: VDT.

Riaz, H., Stiles, M., Armstrong, C., Chadwick, A., Lee, H. and Kearney, G. (2017). Multichannel Microphone Array Recording for Popular Music Production in Virtual Reality. In: *143rd International AES Convention*. [online] New York, E-Brief No. 384. Available at: http://www.aes.org/e-lib/browse.cfm?elib=19333 [Accessed 12 Jan. 2019].

Riitano, L. and Medina Victoria, J. (2018). Comparison between different Microphone-Arrays for 3D-Audio. In: *144th International AES Convention*. [online] Milan. Preprint No. 9980. Available at: http://www.aes.org/e-lib/browse.cfm?elib=19497 [Accessed 12 Jan. 2019].

Robotham, T., Stephenson, M. and Lee, H. (2016). The Effect of a Vertical Reflection on the Relationship between Preference and Perceived Change in Timbral and Spatial Attributes. In: *140th International AES Convention*. [online] Paris. Preprint No. 9547. Available at: http://www.aes.org/e-lib/online/browse.cfm?elib=18245 [Accessed 25 Mar. 2019].

Roginska, A. and Geluso, P. ed., (2018). *Immersive Sound: The Art and Science of Binaural and Multi-Channel Audio*. Oxon: Routledge.

Rumsey, F. (2013). *Spatial Audio*. Oxon: Focal Press.

Rumsey, F. (2002). Spatial Quality Evaluation for Reproduced Sound: Terminology, Meaning, and a Scene-Based Paradigm. *Journal of the Audio Engineering Society*, [online] 50(9), pp.651-666. Available at: http://www.aes.org/e-lib/browse.cfm?elib=11067 [Accessed 12 Jan. 2019].

Rumsey, F. and Lewis, W. (2002). Effect of rear microphone spacing on spatial impression for omnidirectional surround sound microphone arrays. [online] Preprint No. 5563. Available at: http://www.aes.org/e-lib/browse.cfm?elib=11421 [Accessed 9 Feb. 2019].

Rumsey, F., Zieliński, S., Kassier, R. and Bech, S. (2005). On the relative importance of spatial and timbral fidelities in judgments of degraded multichannel audio quality. *The Journal of the Acoustical Society of America*, [online] 118(2), pp.968-976. Available at: https://www.researchgate.net/publication/7603541_On_the_relative_importance_of_spatial_and_timbral_fidelities_in_judgments_of_degraded_multichannel_audio_quality [Accessed 17 Feb. 2019].

Ryaboy, A. (2015). Exploring 3D: A subjective evaluation of surround microphone arrays catered for Auro-3D reproduction system. In: *139th International AES Convention*. [online] New York. Preprint No. 9431. Available at: http://www.aes.org/e-lib/browse.cfm?elib=17987 [Accessed 12 Jan. 2019].

SAE Institute Brussels (2017). [image] Available at: https://www.facebook.com/saebrussels/photos/a.475828862738/10154895014007739/?type=1&theater [Accessed 27 Mar. 2019].

Sawaguchi, M. (2018). *Consideration of Height ch[1] miking along with music style*. [pdf] Tokyo: Mich Sawaguchi. Available at: https://www.academia.edu/37216050/2018_AES_INT_Conference_Immmersive_surround_.pdf [Accessed 12 Jan. 2019].

Schoeps.de. (2018a). *Image Assistant | SCHOEPS Mikrofone*. [online] Available at: https://schoeps.de/en/knowledge/image-assistant.html [Accessed 12 Feb. 2019].

[1] The published article includes this typo.

Schoeps.de. (2018b). *Omnis | SCHOEPS Mikrofone*. [online] Available at: https://schoeps.de/en/products/colette/capsules/omnis.html [Accessed 8 Feb. 2019].

Simmons, G. and Faulkner, T. (2016). *Stereo Masterclass*. [online] Audiotechnology.com. Available at: http://www.audiotechnology.com/stereo-masterclass/ [Accessed 12 Feb. 2019].

Simurra, I. and Queiroz, M. (2017). Pilot Experiment On Verbal Attributes Classification Of Orchestral Timbres. In: *143rd International AES Convention*. [online] New York. Preprint No. 9821. Available at: http://www.aes.org/e-lib/browse.cfm?elib=19218 [Accessed 25 Mar. 2019].

Theile, G. (2001). Natural 5.1 Music Recording Based on Psychoacoustic Principles. In: *19th International AES Conference*. [online] Elmau. Preprint 1904. Available at: http://www.aes.org/e-lib/browse.cfm?elib=10085 [Accessed 12 Jan. 2019].

Theile, G. and Wittek, H. (2012). 3D Audio Natural Recording. In: *27th Tonmeistertagung VDT International Convention*. [online] Cologne. Available at: https://www.hauptmikrofon.de/HW/TMT2012_3DNaturalRecording_Theile_Wittek_2012_11.pdf [Accessed 12 Jan. 2019].

Theile, G. and Wittek, H. (2011). Principles in Surround Recordings with Height. *130th International AES Convention*. [online] London. Preprint No. 8403. Available at: http://www.aes.org/e-lib/browse.cfm?elib=15870 [Accessed 12 Jan. 2019].

Toole, F. (1985). Subjective Measurements of Loudspeaker Sound Quality and Listener Performance. *Journal of the Audio Engineering Society*, [online] 33(1/2), pp.2-32. Available at: http://www.aes.org/e-lib/browse.cfm?elib=4465 [Accessed 7 Apr. 2019].

Transparency Market Research, (2018). *3D Audio Market*. [online] Available at: https://www.transparencymarketresearch.com/3d-audio-market.html [Accessed 29 Mar. 2019].

Wallis, R. and Lee, H. (2017). The Reduction of Vertical Interchannel Crosstalk: The Analysis of Localisation Thresholds for Natural Sound Sources. *Applied Sciences*, 7(3), pp.278-297.

Wallis, R. and Lee, H. (2016a). The Reduction of Vertical Interchannel Crosstalk: The Analysis of Localisation Thresholds for Musical Sources. In: *140th International AES Convention*. [online] Paris. Preprint No. 9513. Available at: http://www.aes.org/e-lib/browse.cfm?elib=18212 [Accessed 12 Jan. 2019].

Wallis, R. and Lee, H. (2016b). Vertical Stereophonic Localization in the Presence of Interchannel Crosstalk: The Analysis of Frequency-Dependent Localization Thresholds. *Journal of the Audio Engineering Society*, [online] 64(10), pp.762-770. Available at: https://secure.aes.org/forum/pubs/journal/?elib=18518 [Accessed 12 Jan. 2019].

Wallis, R. and Lee, H. (2015). The Effect of Interchannel Time Difference on Localization in Vertical Stereophony. *Journal of the Audio Engineering Society*, [online] 63(10), pp.767-776. Available at: http://www.aes.org/e-lib/browse.cfm?elib=18040 [Accessed 12 Jan. 2019].

Wallis, R. and Lee, H. (2014). Investigation into Vertical Stereophonic Localisation in the Presence of Interchannel Crosstalk. In: *136th International AES Convention*. [online] Berlin. Preprint No. 9026. Available at: http://www.aes.org/e-lib/browse.cfm?elib=17173 [Accessed 12 Jan. 2019].

Werner, A., Neukom, M., Bollinger, D., Dölle, L. and Könz, S. (2014). *3D Audio: aesthetical and technical aspects of recording and mixing in 3D*. [pdf] Zurich: Institute for Computer Music and Sound Technology: Zurich University of the Arts. Available at: https://www.zhdk.ch/file/live/b9/b9d9893ca5be0f1433df97d7d8e9c8905a88f1e5/20160811151104_57ac7968c1d24.pdf [Accessed 19 Jan. 2019].

Williams, D. (2011). A comparative perceptual evaluation of the timbral variations in choral location recordings created by four common stereo microphone techniques. In: *130th International AES Convention*. [online] London. Preprint No. 8371. Available at: http://www.aes.org/e-lib/browse.cfm?elib=15838 [Accessed 30 Mar. 2019].

Williams, M. (2012). Microphone Array Design for localisation with elevation cues. In: *132nd International AES Convention*. [online] Budapest. Preprint No. 8601. Available at: http://www.aes.org/e-lib/browse.cfm?elib=16239 [Accessed 12 Jan. 2019].

Williams, M. (2010). *The stereophonic zoom: A variable dual microphone system for stereophonic sound recording*. [pdf] Microphone Data Ltd. Available at:

https://microphone-data.com/media/filestore/articles/Stereo%20zoom-10.pdf [Accessed 12 Jan. 2019].

Williams, M. (2009). *Microphone Array Design*. [online] Mmad.info. Available at: http://www.mmad.info/MAD/Ch_n_cov.htm [Accessed 13 Feb. 2019].

Williams, M. and Le Dû, G. (2010). *Multichannel Microphone Array Design (MMAD)*. [pdf] Microphone Data Ltd. Available at: https://microphone-data.com/media/filestore/articles/MMAD-10.pdf [Accessed 12 Jan. 2019].

Wittek, H. (2018). *Microphone Techniques for VR and 360° Videos - Part 1*. [online] YouTube. Available at: https://www.youtube.com/watch?v=K-ktLM0dQeA&list=PLRqzOEeUQ2l91Gt-d7a2-wDpp6JTwWD4W&index=1 [Accessed 14 Jan. 2019].

Wittek, H. and Theile, G. (2017). Development and application of a stereophonic multichannel recording technique for 3D Audio and VR. In: *143rd International AES Convention*. [online] New York. Preprint No. 9869. Available at: http://www.aes.org/e-lib/browse.cfm?elib=19266 [Accessed 12 Jan. 2019].

Zacharov, N. and Pedersen, T. (2015). Spatial sound attributes - development of a common lexicon. In: *139th International AES Convention*. [online] New York. Preprint No. 9436. Available at: http://www.aes.org/e-lib/browse.cfm?elib=17992 [Accessed 25 Mar. 2019].

Zielinsky, G. (2018). *ICSA Documents*. [email].

8. Bibliography

Abbey Road Studios (2019). *Spatial Audio*. [online] Abbey Road. Available at: https://www.abbeyroad.com/spatial-audio [Accessed 14 Jan. 2019].

Allen, J. (2017). *Mixing & Broadcasting From The Royal Opera House*. [online] Sound on Sound. Available at: https://www.soundonsound.com/techniques/immersive-audio-opera [Accessed 12 Jan. 2019].

Andriessens, F. (2018). *3D-Audio Ambience Recording Techniques*. [online] Hauptmikrofon.de. Available at: http://www.hauptmikrofon.de/de/3d/tonundmeister [Accessed 14 Jan. 2019].

Auro-3D® Home Theater Setup Installation Guidelines. (2015). [pdf] Mol: Auro Technologies. Available at: https://www.auro-3d.com/wp-content/uploads/documents/Auro-3D-Home-Theater-Setup-Guidelines_lores.pdf [Accessed 10 Feb. 2019].

Barbour, J. (2003). Elevation Perception: Phantom Images in the Vertical Hemi-spere. In: *24th International AES Conference*. [online] Banff. Preprint No. 14. Available at: http://www.aes.org/e-lib/browse.cfm?elib=12301 [Accessed 12 Jan. 2019].

Barron, M. (1971). The subjective effects of first reflections in concert halls—The need for lateral reflections. *Journal of Sound and Vibration*, [online] 15(4), pp.475-494. Available at: https://www.researchgate.net/profile/Mike_Barron2/publication/222627645_The_subjective_effects_of_first_reflections_in_concert_halls-The_need_for_lateral_reflections/links/5a0b23e2458515e482747bf3/The-subjective-effects-of-first-reflections-in-concert-halls-The-need-for-lateral-reflections.pdf?origin=publication_detail [Accessed 9 Feb. 2019].

Bartlett, B. (2017). *Size Matters: The Differences In Large- And Small-Diaphragm Condenser Microphones - ProSoundWeb*. [online] ProSoundWeb. Available at: https://www.prosoundweb.com/channels/recording/size_matters_the_differences_in_large-_and_small-diaphragm_microphones/ [Accessed 6 Feb. 2019].

Bech, S. and Zacharov, N. (2006). *Perceptual Audio Evaluation - Theory, Method and Application*. Chichester: Wiley.

Berg, J. and Rumsey, F. (2006). Identification of Quality Attributes of Spatial Audio by Repertory Grid Technique. *Journal of the Audio Engineering Society*, [online] 54(5), pp.365-379. Available at: http://www.aes.org.saeezproxy.idm.oclc.org/e-lib/browse.cfm?elib=13681 [Accessed 17 Feb. 2019].

Berg, J. and Rumsey, F. (2003). Systematic Evaluation of Perceived Spatial Quality. In: *AES 24th International Conference*. [online] Banff. Available at: http://www.aes.org/e-lib/browse.cfm?elib=12272 [Accessed 25 Mar. 2019].

Berg, J. and Rumsey, F. (2000). Correlation between Emotive, Descriptive and Naturalness Attributes in Subjective Data Relating to Spatial Sound Reproduction. In: *109th International AES Convention*. [online] New York. Preprint No. 5206. Available at: http://www.aes.org/e-lib/browse.cfm?elib=9132 [Accessed 3 Apr. 2019].

Bowles, D. (2015). A microphone array for recording music in surround-sound with height channels. In: *139th International AES Convention*. [online] New York. Preprint No. 9430. Available at: http://www.aes.org/e-lib/browse.cfm?elib=17986 [Accessed 12 Jan. 2019].

Bruil, R. (2001). *Mercury Records Living Presence - Wilma Cozart Fine and 50 Years Mercury Recordings*. [online] Soundfountain.com. Available at: http://www.soundfountain.com/amb/mercury.html [Accessed 4 Feb. 2019].

Cengale, G. (2012). *Audio Technologies: applications to sound capture, post-production and listener perception*. Ph.D. Universitat Pompeu Fabra, Barcelona.

Conetta, R., Brookes, T., Rumsey, F., Zielinski, S., Dewhirst, M., Jackson, P., Bech, S., Meares, D. and George, S. (2014). Spatial Audio Quality Perception (Part 1): Impact of Commonly Encountered Processes. *Journal of the Audio Engineering Society*, [online] 62(12), pp.831-846. Available at: https://secure.aes.org/forum/pubs/journal/?elib=17557 [Accessed 17 Feb. 2019].

Dieterle, H. (2018). *Microphone Techniques for VR and 360° Videos - Part 3*. [online] YouTube. Available at: https://www.youtube.com/watch?v=ohE5I60vy6o&list=PLRqzOEeUQ2l91Gt-d7a2-wDpp6JTwWD4W&index=3 [Accessed 14 Jan. 2019].

DPA. (2018). *Immersive sound / Object-based Audio – and Microphones*. [online] Available at: https://www.dpamicrophones.com/mic-university/immersive-sound-object-based-audio-and-microphones [Accessed 14 Jan. 2019].

Ellis-Geiger, R. (2016). Music Production for Dolby Atmos and Auro-3D. In: *141st International AES Convention*. [online] Los Angeles. Preprint No. 9675. Available at: http://www.aes.org/e-lib/browse.cfm?elib=18479 [Accessed 12 Jan. 2019].

Faulkner, T. (2019). *All Saints Tooting for String Quartet?*. [email].

Faulkner, T. (2019). *LCR Stereo Recording?*. [email].

Fine, T. (2012). *Wilma Cozart Fine & C. Robert Fine: Mercury Living Presence Recordings*. [online] Tapeop.com. Available at: https://tapeop.com/tutorials/90/wilma-cozart-fine-c-robert-fine/ [Accessed 4 Feb. 2019].

Francombe, J., Brookes, T. and Mason, R. (2015). Elicitation of the Differences between Real and Reproduced Audio. In: *138th International AES Convention*. [online] Warsaw. Preprint No. 9307. Available at: http://www.aes.org/e-lib/online/browse.cfm?elib=17731 [Accessed 25 Mar. 2019].

Geluso, P. (2012). Capturing Height: The Addition of Z Microphones to Stereo and Surround Microphone Arrays. In: *132nd International AES Convention*. [online] Budapest. Preprint No. 8595. Available at: http://www.aes.org/e-lib/browse.cfm?elib=16233 [Accessed 12 Jan. 2019].

Gerzon, M. (1971). *Whither Four Channels?*. [pdf] Croydon: Link House Publications. Available at: https://www.michaelgerzonphotos.org.uk/articles/Whither.pdf [Accessed 25 Mar. 2019].

Gribben, C. and Lee, H. (2018). The Frequency and Loudspeaker-Azimuth Dependencies of Vertical Interchannel Decorrelation on the Vertical Spread of an Auditory Image. *Journal of the Audio Engineering Society*, [online] 66(7/8), pp.537-555. Available at: http://www.aes.org/e-lib/browse.cfm?elib=19705 [Accessed 13 Jan. 2019].

Gribben, C. and Lee, H. (2016). The Perception of Vertical Image Spread by Interchannel Decorrelation. In: *140th International AES Convention*. [online] Paris.

Preprint No. 9514. Available at: http://www.aes.org/e-lib/browse.cfm?elib=18213 [Accessed 12 Jan. 2019].

Gribben, C. and Lee, H. (2015). Towards the Development of a Universal Listening Test Interface Generator in Max. In: *138th International AES Convention*. [online] Warsaw. E-Brief No. 187. Available at: http://www.aes.org/e-lib/browse.cfm?elib=17622 [Accessed 13 Jan. 2019].

Griesinger, D. (1999). *The Science of Surround*. [pdf] Available at: http://www.davidgriesinger.com/aes99.pdf [Accessed 9 Feb. 2019].

Griesinger, D. (1998). General Overview of Spatial Impression, Envelopment, Localization, and Externalization. In: *15th International AES Conference*. [online]. Preprint No. 15-013. Available at: http://www.aes.org.saeezproxy.idm.oclc.org/e-lib/inst/search.cfm [Accessed 8 Feb. 2019].

Hamasaki, K., Hiyama, K., Nishiguchi, T. and Okumura, R. (2006). Effectiveness for height information for reproducing presence and reality in multichannel audio system. In: *120th International AES Convention*. [online] Paris. Preprint No. 6679. Available at: http://www.aes.org.saeezproxy.idm.oclc.org/e-lib/browse.cfm?elib=13483 [Accessed 19 Jan. 2019].

Hamasaki, K., Shinmura, T., Akita, S. and Hiyama,, K. (2001). Approach and Mixing Technique for Natural Sound Recording of Multichannel Audio. In: *19th International AES Conference*. [online] Preprint No. 1878. Available at: http://www.aes.org/e-lib/browse.cfm?elib=10111 [Accessed 9 Feb. 2019].

Hamasaki, K. and Van Baelen, W. (2015). Natural Sound Recording of an Orchestra with Three-dimensional Sound. In: *138th International AES Convention*. [online] Warsaw. Preprint No. 9348. Available at: http://www.aes.org/e-lib/browse.cfm?elib=17772 [Accessed 12 Jan. 2019].

Hanelt, M. and Ehret, A. (2017). Ambience recording for 3D audio. In: *142nd International AES Convention*. [online] Berlin. E-Brief No. 312. Available at: http://www.aes.org/e-lib/browse.cfm?elib=18688 [Accessed 12 Jan. 2019].

Hanyu, T., Kimura, S. and Chiba, S. (1999). A New Objective Measure For Evaluation of Listener Envelopment Focusing on the Spatial Balance of Reflections. *Journal of*

Architecture and Planning (Transactions of AIJ), [online] 64(520), pp.9-16. Available at: https://www.researchgate.net/publication/313326738_A_NEW_OBJECTIVE_MEASURE_FOR_EVALUATION_OF_LISTENER_ENVELOPMENT_FOCUSING_ON_THE_SPATIAL_BALANCE_OF_REFLECTIONS [Accessed 9 Feb. 2019].

Holman, T. (2008). *Surround Sound: Up and Running*. 2nd ed. Oxon: Focal Press.

Howie, W. (2018). *Capturing Orchestral Music for Three-Dimensional Audio Playback*. Ph.D. McGill University, Montreal.

Howie, W. and King, R. (2015). Exploratory microphone techniques for three-dimensional classical music recording. In: *138th International AES Convention*. [online] Warsaw. E-Brief No. 196. Available at: http://www.aes.org/e-lib/browse.cfm?elib=17631 [Accessed 12 Jan. 2019].

Howie, W., King, R., Boerum, M., Benson, D. and Han, A. (2015). Listener preference for height channel microphone polar patterns in three-dimensional recording. In: *139th International AES Convention*. [online] New York. Preprint No. 9419. Available at: http://www.aes.org/e-lib/browse.cfm?elib=17975 [Accessed 12 Jan. 2019].

Howie, W., King, R. and Martin, D. (2016). A Three-Dimensional Orchestral Music Recording Technique, Optimized for 22.2 Multichannel Sound. In: *141st International AES Convention*. [online] Los Angeles. Preprint No. 9612. Available at: http://www.aes.org/e-lib/browse.cfm?elib=18416 [Accessed 12 Jan. 2019].

Howie, W., King, R., Martin, D. and Grond, F. (2017). Subjective Evaluation of Orchestral Music Recording Techniques for Three-Dimensional Audio. In: *142nd International AES Convention*. [online] Berlin. Preprint No. 9797. Available at: http://www.aes.org/e-lib/browse.cfm?elib=18671 [Accessed 12 Jan. 2019].

Howie, W., Martin, D., Benson, D., Kelly, J. and King, R. (2018). Subjective and objective evaluation of 9ch three-dimensional acoustic music recording techniques. In: *International Conference on Spatial Reproduction – Aesthetics and Science*. [online] Tokyo. Available at: http://www.aes.org/e-lib/browse.cfm?elib=19630 [Accessed 25 Mar. 2019].

Izhaki, R. (2018). *Mixing Audio*. 3rd ed. New York: Focal Press.

Kamekawa, T., Marui, A., Date, T. and Enatsu, M. (2011). Evaluation of spatial impression comparing 2ch stereo, 5ch surround, and 7ch surround with height for 3D imagery. In: *130th International AES Convention*. [online] London. Preprint No. 8334. Available at: http://www.aes.org/e-lib/browse.cfm?elib=15801 [Accessed 19 Jan. 2019].

Kassier, R., Lee, H., Brookes, T. and Rumsey, F. (2005). An Informal Comparison Between Surround-Sound Microphone Techniques. In: *118th International AES Convention*. [online] Barcelona. Preprint No. 6429. Available at: http://www.aes.org/e-lib/browse.cfm?elib=13145 [Accessed 5 Feb. 2019].

Kim, S., Ko, D., Nagendra, A. and Woszczyk, W. (2013). Subjective evaluation of multichannel Sound with Surround-Height. In: *135th International AES Convention*. [online] New York. Preprint No. 9003. Available at: http://www.aes.org/e-lib/browse.cfm?elib=17051 [Accessed 19 Jan. 2019].

King, R. (2016). *Recording Orchestra and Other Classical Music Ensembles*. New York: Routledge.

King, R., Howie, W. and Kelly, J. (2016). A Survey of Suggested Techniques for Height Channel Capture in Multi-channel Recording. In: *140th International AES Convention*. [online] Paris. E-Brief No. 266. Available at: http://www.aes.org/e-lib/browse.cfm?elib=18170 [Accessed 12 Jan. 2019].

Koivuniemi, K. and Zacharov, N. (2001). Unravelling the Perception of Spatial Sound Reproduction: Language Development, Verbal Protocol Analysis and Listener Training. In: *111th International AES Convention*. [online] New York. Preprint No. 5424. Available at: http://www.aes.org/e-lib/browse.cfm?elib=9815 [Accessed 3 Apr. 2019].

Lee, H. (2019). *Microphone Angle Calculations/ICLD*. [email].

Lee, H. (2019). Capturing 360◦ Audio Using an Equal Segment Microphone Array (ESMA). *Journal of the Audio Engineering Society*, 67(1/2), pp.1-14.

Lee, H. (2018). *Hyunkook Lee: Psychoacoustics of 3D sound recording techniques: Research and Practice*. [online] YouTube. Available at: https://www.youtube.com/watch?v=5SsQdfKAhVA [Accessed 14 Jan. 2019].

Lee, H. (2018). *Psychoacoustics of 3D Sound Recording: Research and Practice*. [pdf] University of Huddersfield. Available at: https://www.researchgate.net/publication/325429307_Psychoacoustics_of_3D_Sound_Recording_Research_and_Practice [Accessed 10 Feb. 2019].

Lee, H. (2011). The Relationship between Interchannel Time and Level Differences in Vertical Sound Localisation and Masking. In: *131st International AES Convention*. [online] New York. Preprint No. 8556. Available at: http://www.aes.org/e-lib/browse.cfm?elib=16082 [Accessed 12 Jan. 2019].

Lee, H. and Gribben, C. (2014). Effect of Vertical Microphone Layer Spacing for a 3D Microphone Array. *Journal of the Audio Engineering Society*, [online] 62(12), pp.870-884. Available at: http://www.aes.org/e-lib/browse.cfm?elib=17560 [Accessed 12 Jan. 2019].

Lee, H. and Gribben, C. (2013). On the optimum microphone array configuration for height channels. In: *134th International AES Convention*. [online] Rome. E-Brief No. 93. Available at: http://www.aes.org/e-lib/browse.cfm?elib=16694 [Accessed 12 Jan. 2019].

Lee, H., Gribben, C. and Wallis, R. (2014). Psychoacoustic Considerations in Surround Sound with Height. In: *28th Tonmeistertagung VDT International Convention*. [online] Cologne. Available at: http://eprints.hud.ac.uk/id/eprint/23151/ [Accessed 12 Jan. 2019].

Lee, H., Johnson, D. and Mironovs, M. (2017). An Interactive and Intelligent Tool for Microphone Array Design. In: *143rd International AES Convention*. [online] New York. E-Brief No. 390. Available at: http://www.aes.org/e-lib/browse.cfm?elib=19338 [Accessed 12 Jan. 2019].

Lee, H., Johnson, D. and Mironovs, M. (2017). *MARRS Quick Start Guide*. [pdf] Huddersfield: Applied Psychoacoustics Lab (APL), University of Huddersfield. Available at: http://eprints.hud.ac.uk/id/eprint/33675/8/MARRS%20Quick%20Start%20Guide.pdf [Accessed 20 Jan. 2019].

Lee, H. and Millns, C. (2017). *Microphone Array Impulse Response (MAIR) Library and Renderer*. [pdf] Huddersfield: Hyunkook Lee/ University of Huddersfield. Available at: https://github.com/APL-Huddersfield/MAIR-Library-and-

Renderer/blob/master/MAIR%20Technical%20Descriptions%20and%20User%20Guide.pdf [Accessed 20 Jan. 2019].

Lee, H. and Millns, C. (2017). Microphone Array Impulse Response (MAIR) Library for Spatial Audio Research. In: *143rd International AES Convention*. [online] New York. E-Brief No. 356. Available at: http://www.aes.org/e-lib/browse.cfm?elib=19307 [Accessed 12 Jan. 2019].

Lemmens, P. (2014). *Auro-3D: Working With The Auro-3D Format In Music Recording & Production*. [online] YouTube. Available at: https://www.youtube.com/watch?v=CQi_bnO0tas [Accessed 14 Jan. 2019].

Letowski, T. (1989). Sound Quality Assessment: Concepts and Criteria. In: *87th International AES Convention*. [online] New York. Preprint No. 2825. Available at: http://www.aes.org.saeezproxy.idm.oclc.org/e-lib/browse.cfm?elib=5869 [Accessed 17 Feb. 2019].

Lindberg, M. (2018). [image] Available at: http://www.2l.no/media/2L-145_recording-sessions/2L-145_recording-sessions-21.jpg [Accessed 10 Feb. 2019].

Lindberg, M. (2018). *Sennheiser PRO TALK | Morten Lindberg - Part 3 of 3 I Sennheiser*. [online] YouTube. Available at: https://www.youtube.com/watch?v=F88b8VRqeQg [Accessed 14 Jan. 2019].

Lindberg, M. (2015). 3D Recording with the “2L-cube”. *VDT*, [online] (3/2015), p.15. Available at: http://www.2l.no/artikler/2L-VDT.pdf [Accessed 12 Jan. 2019].

Lindberg, M. (2009). *About 2L (Lindberg Lyd)*. [online] 2l.no. Available at: http://www.2l.no/pages/about.html [Accessed 14 Jan. 2019].

Lossenderosstudio.com. (2018). *Surround Sound Microphone Techniques*. [online] Available at: http://lossenderosstudio.com/article.php?subject=17 [Accessed 14 Jan. 2019].

Luthar, M. and Maltezos, E. (2015). Polar Pattern Comparisons for the Left, Center, and Right Channels in a 3-D Microphone Array. In: *139th International AES Convention*. [online] New York. E-Brief No. 206. Available at: http://www.aes.org/e-lib/browse.cfm?elib=17882 [Accessed 12 Jan. 2019].

Martin, B., King, R., Leonard, B., Benson, D. and Howie, W. (2015). Immersive Content in Three Dimensional Recording Techniques for Single Instruments in Popular Music. In: *138th International AES Convention*. [online] Warsaw. Available at: http://www.aes.org.saeezproxy.idm.oclc.org/e-lib/browse.cfm?elib=17675 [Accessed 18 Jan. 2019].

Martin, B., King, R. and Woszcyk, W. (2016). Subjective Graphical Representation of Microphone Arrays for Vertical Imaging and Three Dimensional Capture of Acoustic Instruments, Part I. In: *141st International AES Convention*. [online] Los Angeles. Preprint No. 9613. Available at: http://www.aes.org/e-lib/browse.cfm?elib=18417 [Accessed 12 Jan. 2019].

Meyerson, A. (2016). *Film Score mixing with Alan Meyerson*. [online] Film Mixing and Sound design. Available at: https://film-mixing.com/2016/07/28/film-score-mixing-with-alan-meyerson/ [Accessed 14 Jan. 2019].

Millns, C. and Lee, H. (2018). An Investigation into Spatial Attributes of 360° Microphone Techniques for Virtual Reality. In: *144th International AES Convention*. [online] Milan. Preprint No. 10005. Available at: http://www.aes.org/e-lib/browse.cfm?elib=19401 [Accessed 13 Jan. 2019].

Moylan, W. (2012). Considering Space in Recorded Music. In: S. Frith and S. Zagorski-Thomas, ed., *The Art of Record Production: An Introductory Readerfor a New Academic Field*, 1st ed. Surrey: Ashgate Publishing Limited, pp.163-188.

Nakayama, T., Tanetoshi, M., Osamu, K., Okamoto, M. and Shiga, T. (1971). Subjective Assessment of Multichannel Reproduction. *Journal of the Audio Engineering Society*, [online] 19(9), pp.744-751. Available at: http://www.aes.org/e-lib/browse.cfm?elib=2135 [Accessed 17 Feb. 2019].

Parnell, T. (2018). *Tom Parnell (BBC R&D): Spatial Audio at the BBC*. [online] YouTube. Available at: https://www.youtube.com/watch?v=nSJoSBNDKMs&list=PLRqzOEeUQ2l-aWeza-XNqArCbfKGvOGpt&t=86s&index=7 [Accessed 14 Jan. 2019].

Pedersen, T. and Zacharov, N. (2015). The development of a Sound Wheel for Reproduced Sound. In: *138th International AES Convention*. [online] Warsaw. Preprint

No. 9310. Available at: http://www.aes.org/e-lib/browse.cfm?elib=17734 [Accessed 25 Mar. 2019].

Pratt, R. and Doak, P. (1976). A subjective rating scale for timbre. *Journal of Sound and Vibration*, 45(3), pp.317-328.

Proceedings of the ICSA 2011. (2011). [DVD] Directed by G. Zielinsky, P. Lemmens, W. Dabringhaus, H. Wittek and J. Nettingsmeier. Berlin: VDT.

Riaz, H., Stiles, M., Armstrong, C., Chadwick, A., Lee, H. and Kearney, G. (2017). Multichannel Microphone Array Recording for Popular Music Production in Virtual Reality. In: *143rd International AES Convention*. [online] New York, E-Brief No. 384. Available at: http://www.aes.org/e-lib/browse.cfm?elib=19333 [Accessed 12 Jan. 2019].

Riitano, L. and Medina Victoria, J. (2018). Comparison between different Microphone-Arrays for 3D-Audio. In: *144th International AES Convention*. [online] Milan. Preprint No. 9980. Available at: http://www.aes.org/e-lib/browse.cfm?elib=19497 [Accessed 12 Jan. 2019].

Robotham, T., Stephenson, M. and Lee, H. (2016). The Effect of a Vertical Reflection on the Relationship between Preference and Perceived Change in Timbral and Spatial Attributes. In: *140th International AES Convention*. [online] Paris. Preprint No. 9547. Available at: http://www.aes.org/e-lib/online/browse.cfm?elib=18245 [Accessed 25 Mar. 2019].

Roginska, A. and Geluso, P. ed., (2018). *Immersive Sound: The Art and Science of Binaural and Multi-Channel Audio*. Oxon: Routledge.

Rudzki, T., Murphy, D. and Kearney, G. (2018). A DAW-based interactive tool for perceptual spatial audio evaluation. [online] New York. E-Brief No. 465. Available at: http://www.aes.org/e-lib/browse.cfm?elib=19730 [Accessed 13 Jan. 2019].

Rumsey, F. (2013). *Spatial Audio*. Oxon: Focal Press.

Rumsey, F. (2002). Spatial Quality Evaluation for Reproduced Sound: Terminology, Meaning, and a Scene-Based Paradigm. *Journal of the Audio Engineering Society*, [online] 50(9), pp.651-666. Available at: http://www.aes.org/e-lib/browse.cfm?elib=11067 [Accessed 12 Jan. 2019].

Rumsey, F. and Lewis, W. (2002). Effect of rear microphone spacing on spatial impression for omnidirectional surround sound microphone arrays. [online] Preprint No. 5563. Available at: http://www.aes.org/e-lib/browse.cfm?elib=11421 [Accessed 9 Feb. 2019].

Rumsey, F., Zieliński, S., Kassier, R. and Bech, S. (2005). On the relative importance of spatial and timbral fidelities in judgments of degraded multichannel audio quality. *The Journal of the Acoustical Society of America*, [online] 118(2), pp.968-976. Available at: https://www.researchgate.net/publication/7603541_On_the_relative_importance_of_spatial_and_timbral_fidelities_in_judgments_of_degraded_multichannel_audio_quality [Accessed 17 Feb. 2019].

SAE Institute Brussels (2017). [image] Available at: https://www.facebook.com/saebrussels/photos/a.475828862738/10154895014007739/?type=1&theater [Accessed 27 Mar. 2019].

Sawaguchi, M. (2018). *Consideration of Height ch[2] miking along with music style*. [pdf] Tokyo: Mich Sawaguchi. Available at: https://www.academia.edu/37216050/2018_AES_INT_Conference_Immmersive_surround_.pdf [Accessed 12 Jan. 2019].

Schoeps.de. (2018). *Image Assistant | SCHOEPS Mikrofone*. [online] Available at: https://schoeps.de/en/knowledge/image-assistant.html [Accessed 12 Feb. 2019].

Schoeps.de. (2018). *Omnis | SCHOEPS Mikrofone*. [online] Available at: https://schoeps.de/en/products/colette/capsules/omnis.html [Accessed 8 Feb. 2019].

Schoeps Mikroforum (2017). *Tonmeistertagung 2016: Roundtable 'Mikrofonierung für 3D-Audio'*. [online] YouTube. Available at: https://www.youtube.com/watch?v=2bbCAjUDvlI&t=0s&list=PLA6S4AUYu_D4S_yZq4Rm0LgMTuI_RqX36&index=3 [Accessed 14 Jan. 2019].

Simmons, G. and Faulkner, T. (2016). *Stereo Masterclass*. [online] Audiotechnology.com. Available at: http://www.audiotechnology.com/stereo-masterclass/ [Accessed 12 Feb. 2019].

[2] The published article includes this typo.

Simurra, I. and Queiroz, M. (2017). Pilot Experiment On Verbal Attributes Classification Of Orchestral Timbres. In: *143rd International AES Convention*. [online] New York. Preprint No. 9821. Available at: http://www.aes.org/e-lib/browse.cfm?elib=19218 [Accessed 25 Mar. 2019].

Stiles, M. (2018). *In conversation with Stephen Barton - The Future of Spatial Audio*. [online] Abbey Road. Available at: https://www.abbeyroad.com/news/in-conversation-with-stephen-barton-the-future-of-spatial-audio-2409 [Accessed 14 Jan. 2019].

Theile, G. (2001). Natural 5.1 Music Recording Based on Psychoacoustic Principles. In: *19th International AES Conference*. [online] Elmau. Preprint 1904. Available at: http://www.aes.org/e-lib/browse.cfm?elib=10085 [Accessed 12 Jan. 2019].

Theile, G. and Wittek, H. (2012). 3D Audio Natural Recording. In: *27th Tonmeistertagung VDT International Convention*. [online] Cologne. Available at: https://www.hauptmikrofon.de/HW/TMT2012_3DNaturalRecording_Theile_Wittek_2012_11.pdf [Accessed 12 Jan. 2019].

Theile, G. and Wittek, H. (2011). Principles in Surround Recordings with Height. *130th International AES Convention*. [online] London. Preprint No. 8403. Available at: http://www.aes.org/e-lib/browse.cfm?elib=15870 [Accessed 12 Jan. 2019].

Toole, F. (1985). Subjective Measurements of Loudspeaker Sound Quality and Listener Performance. *Journal of the Audio Engineering Society*, [online] 33(1/2), pp.2-32. Available at: http://www.aes.org/e-lib/browse.cfm?elib=4465 [Accessed 7 Apr. 2019].

Transparency Market Research, (2018). *3D Audio Market*. [online] Available at: https://www.transparencymarketresearch.com/3d-audio-market.html [Accessed 29 Mar. 2019].

University of York (2017). *SADIE : Spatial Audio For Domestic Interactive Entertainment*. [online] York.ac.uk. Available at: https://www.york.ac.uk/sadie-project/ [Accessed 14 Jan. 2019].

Wallis, R. and Lee, H. (2017). The Reduction of Vertical Interchannel Crosstalk: The Analysis of Localisation Thresholds for Natural Sound Sources. *Applied Sciences*, 7(3), pp.278-297.

Wallis, R. and Lee, H. (2016). The Reduction of Vertical Interchannel Crosstalk: The Analysis of Localisation Thresholds for Musical Sources. In: *140th International AES Convention*. [online] Paris. Preprint No. 9513. Available at: http://www.aes.org/e-lib/browse.cfm?elib=18212 [Accessed 12 Jan. 2019].

Wallis, R. and Lee, H. (2016). Vertical Stereophonic Localization in the Presence of Interchannel Crosstalk: The Analysis of Frequency-Dependent Localization Thresholds. *Journal of the Audio Engineering Society*, [online] 64(10), pp.762-770. Available at: https://secure.aes.org/forum/pubs/journal/?elib=18518 [Accessed 12 Jan. 2019].

Wallis, R. and Lee, H. (2015). The Effect of Interchannel Time Difference on Localization in Vertical Stereophony. *Journal of the Audio Engineering Society*, [online] 63(10), pp.767-776. Available at: http://www.aes.org/e-lib/browse.cfm?elib=18040 [Accessed 12 Jan. 2019].

Wallis, R. and Lee, H. (2014). Investigation into Vertical Stereophonic Localisation in the Presence of Interchannel Crosstalk. In: *136th International AES Convention*. [online] Berlin. Preprint No. 9026. Available at: http://www.aes.org/e-lib/browse.cfm?elib=17173 [Accessed 12 Jan. 2019].

Werner, A., Neukom, M., Bollinger, D., Dölle, L. and Könz, S. (2014). *3D Audio: aesthetical and technical aspects of recording and mixing in 3D*. [pdf] Zurich: Institute for Computer Music and Sound Technology: Zurich University of the Arts. Available at: https://www.zhdk.ch/file/live/b9/b9d9893ca5be0f1433df97d7d8e9c8905a88f1e5/20160811151104_57ac7968c1d24.pdf [Accessed 19 Jan. 2019].

Williams, D. (2011). A comparative perceptual evaluation of the timbral variations in choral location recordings created by four common stereo microphone techniques. In: *130th International AES Convention*. [online] London. Preprint No. 8371. Available at: http://www.aes.org/e-lib/browse.cfm?elib=15838 [Accessed 30 Mar. 2019].

Williams, M. (2012). Microphone Array Design for localisation with elevation cues. In: *132nd International AES Convention*. [online] Budapest. Preprint No. 8601. Available at: http://www.aes.org/e-lib/browse.cfm?elib=16239 [Accessed 12 Jan. 2019].

Williams, M. (2010). *The stereophonic zoom: A variable dual microphone system for stereophonic sound recording*. [pdf] Microphone Data Ltd. Available at:

https://microphone-data.com/media/filestore/articles/Stereo%20zoom-10.pdf [Accessed 12 Jan. 2019].

Williams, M. (2009). *Microphone Array Design*. [online] Mmad.info. Available at: http://www.mmad.info/MAD/Ch_n_cov.htm [Accessed 13 Feb. 2019].

Williams, M. (2003). Multichannel Sound Recording Practice Using Microphone Arrays. In: *24th International AES Conference*. [online] Banff. Preprint No. 28. Available at: http://www.aes.org/e-lib/browse.cfm?elib=12287 [Accessed 12 Jan. 2019].

Williams, M. and Le Dû, G. (2010). *Multichannel Microphone Array Design (MMAD)*. [pdf] Microphone Data Ltd. Available at: https://microphone-data.com/media/filestore/articles/MMAD-10.pdf [Accessed 12 Jan. 2019].

Wittek, H. (2018). *Microphone Techniques for VR and 360° Videos - Part 1*. [online] YouTube. Available at: https://www.youtube.com/watch?v=K-ktLM0dQeA&list=PLRqzOEeUQ2l91Gt-d7a2-wDpp6JTwWD4W&index=1 [Accessed 14 Jan. 2019].

Wittek, H. and Theile, G. (2017). Development and application of a stereophonic multichannel recording technique for 3D Audio and VR. In: *143rd International AES Convention*. [online] New York. Preprint No. 9869. Available at: http://www.aes.org/e-lib/browse.cfm?elib=19266 [Accessed 12 Jan. 2019].

Zacharov, N. and Pedersen, T. (2015). Spatial sound attributes - development of a common lexicon. In: *139th International AES Convention*. [online] New York. Preprint No. 9436. Available at: http://www.aes.org/e-lib/browse.cfm?elib=17992 [Accessed 25 Mar. 2019].

Zielinsky, G. (2018). *ICSA Documents*. [email].

9. Appendices

9.1 Conditions for Stereophonic Arrays

9.1.1 General Overview over the Basic Principles of Stereophonic Arrays

According to Theile & Wittek the requirements for 3D arrays are the same as in two- and five-channel stereophony (2011, pp. 2 and 4). These are:

- **Signal separation** among all channels in order to avoid comb filtering: No signal should be present at significant levels in more than two channels
- **Level and/or arrival time differences** between adjacent channels to achieve the desired imaging characteristics
- **Decorrelation of diffuse-field sound** for optimal envelopment and sound quality

The first requirement to provide enough signal separation is strongly related to the effect of crosstalk. Crosstalk refers to the signal presence in the third channel, which should be avoided as it blurs the image and introduces colouration (Theile & Wittek, 2011; Rumsey, 2008, p. 193). Decorrelation of the DF (a reduction of ICC) can be achieved with the help of three parameters

a) Increasing the angle between microphones
b) Increasing the distance between microphones
c) Choosing a polar pattern with increased directivity

This seems of particular importance as Griesinger found that ambient signals should be decorrelated over the entire frequency range, especially in the low-frequency range as it is the basis of envelopment perception (1998, p. 140).

In the decision-making process of choosing suitable arrays and their specific setup it was therefore considered important to fulfil these requirements as much as possible. This was done mainly through studying the results of previous researches, standard literature explaining the behaviours of the arrays in question, as well as through the use of the softwares MARRS (Lee *et al.*, 2017b) and the Image Assistant by Schoeps (Schoeps.de, 2018a), which allow the creation of ICTD/ICLD and ICCC graphs and localisation curves for custom stereophonic arrays.

ICTD and ICLD graphs convey an idea of the sonic behaviour of stereophonic arrays. A good example here fore is the AB technique, which will not show any level differences (being omni microphones). Instead, (almost) only time differences are visible between the two microphones which leads to a more spacious (more ICTD) but less accurate source localisation (lack of ICLD). This can be seen in Figure 9.1.1:

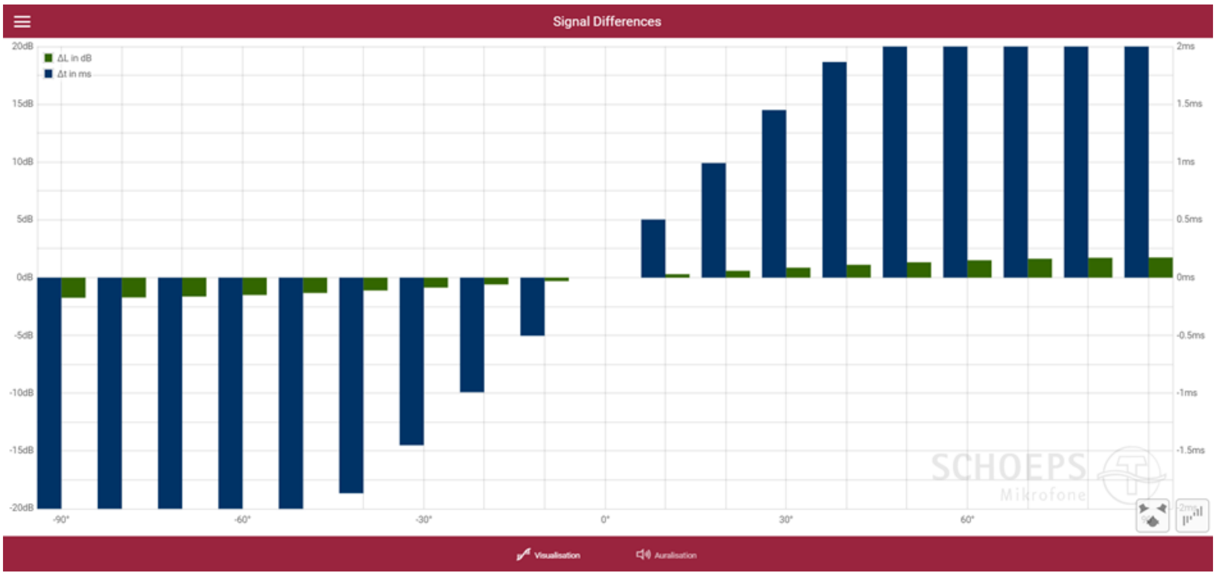


Figure 9.1.1: ICTD/ICLD Differences for an 1m Spaced AB Array (Schoeps.de, 2018a)

The XY technique can be taken as an example for the opposite behaviour. It is characterised by its superior localisation accuracy but at the same time it exhibits a limited spatial impression due to the lack of ICTDs (see Figure 9.1.2).

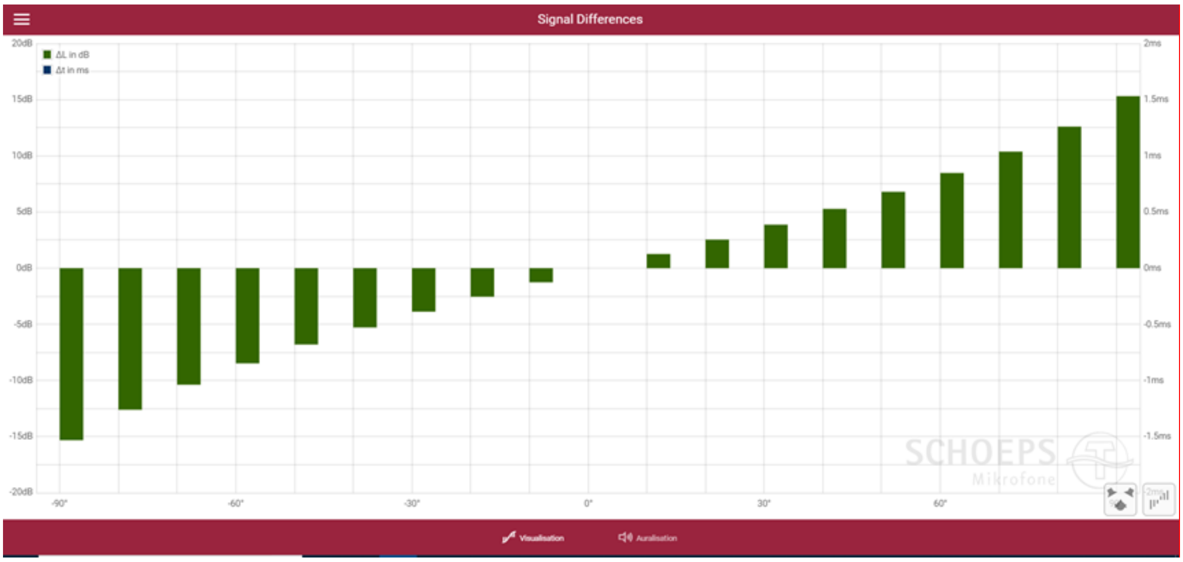


Figure 9.1.2: ICTD/ICLD Differences for a Coincident XY Array (Schoeps.de, 2018a)

ICCC graphs depict the behaviour of the correlation between the two microphone signals, whereas 1 means fully correlated and 0 fully decorrelated. Localisation curves, on the other hand, visualise the array's behaviour of producing phantom images between the loudspeakers (based on the SRA of the array). Also, for these examples, the two opposite cases of an AB and XY technique shall be taken as a comparison in Figures 9.1.3-9.1.6.

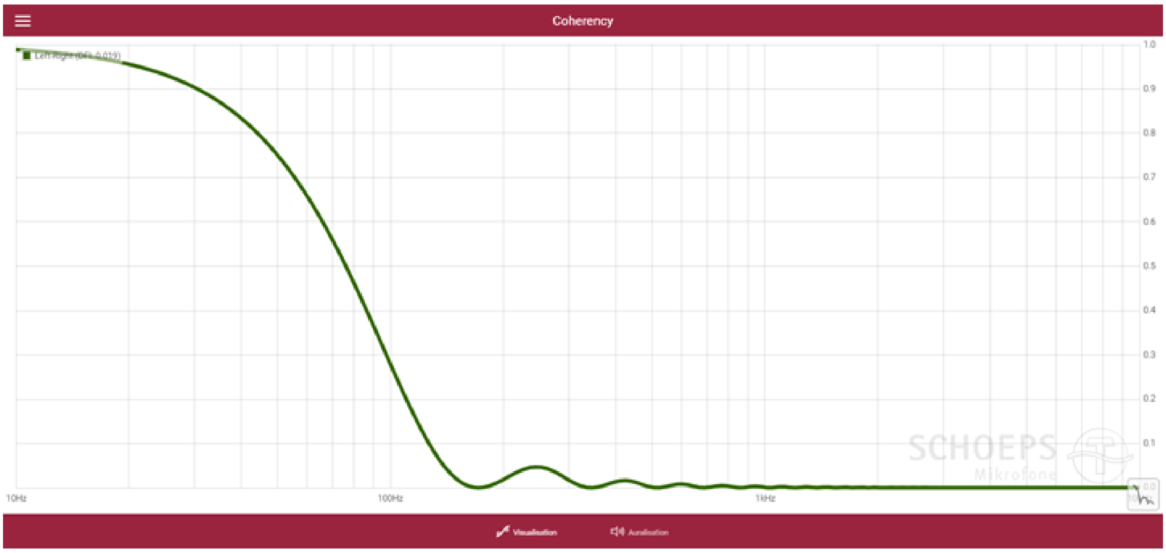


Figure 9.1.3: (De)-Correlation of an 1m Spaced AB Array (Schoeps.de, 2018a)

Frequency content above 100Hz is fully Decorrelated (ICCC=0)

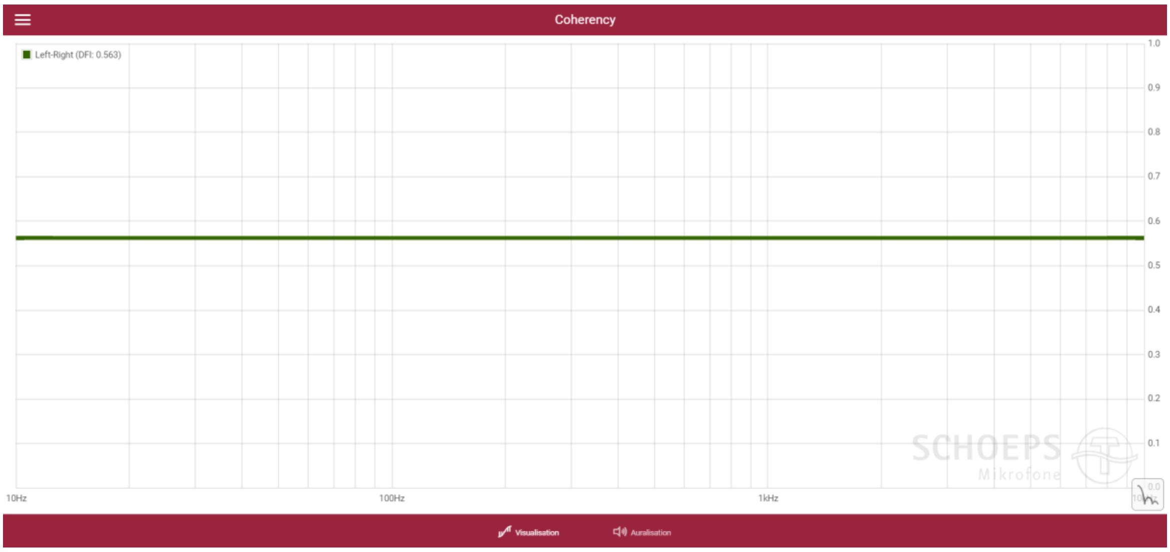


Figure 9.1.4: (De)-Correlation of an XY Array (Schoeps.de, 2018a)

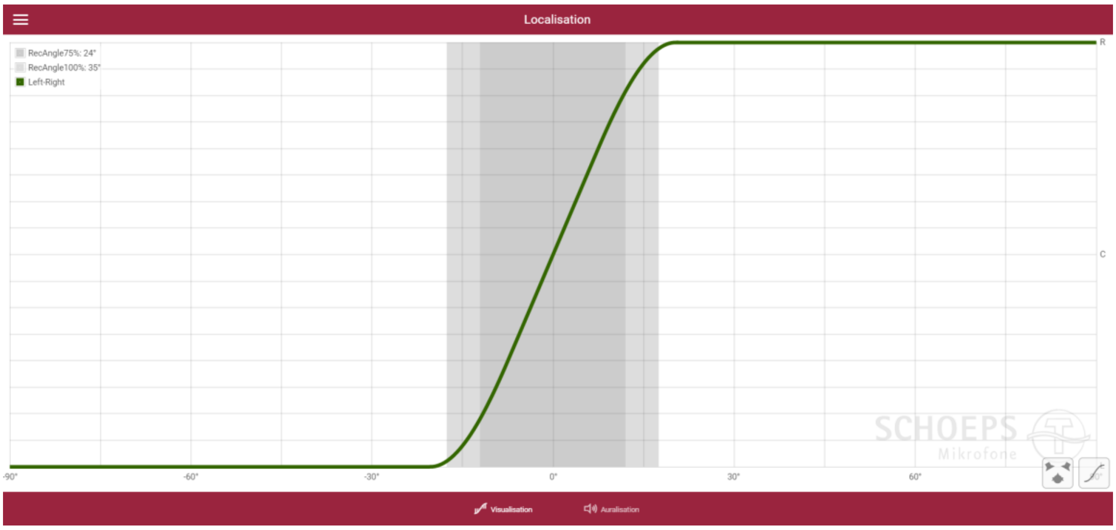


Figure 9.1.5: Localisation Curve of the XY Configuration (Schoeps.de, 2018a)

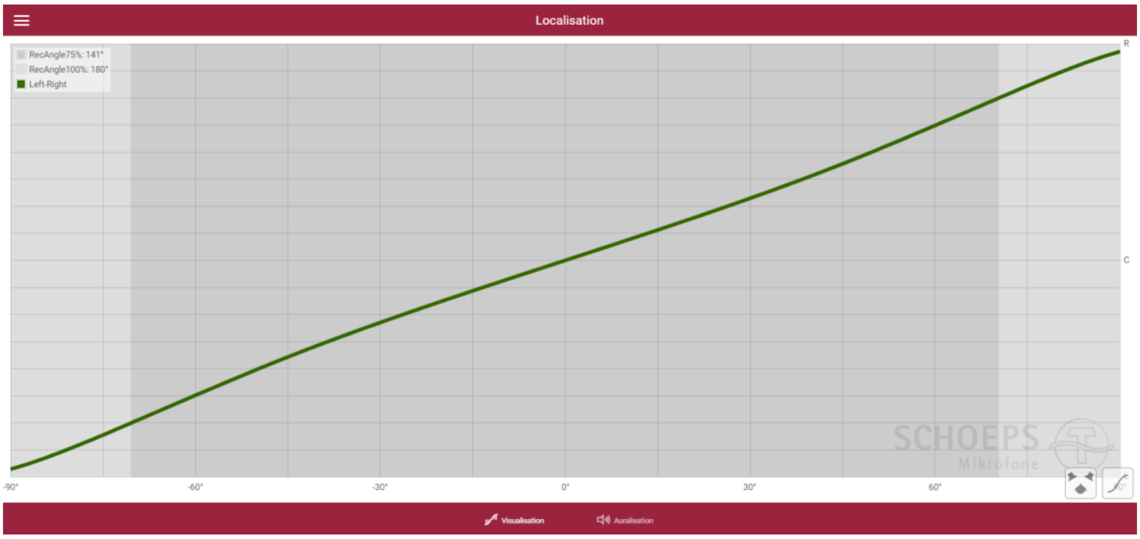


Figure 9.1.6: Localisation Curve of the XY Configuration (Schoeps.de, 2018a)

Both, the Image Assistant and MARRS have been used to get some ballpark comparison of the obtained graphs. A general overview of the MARRS application can be seen in Figure 9.1.7:

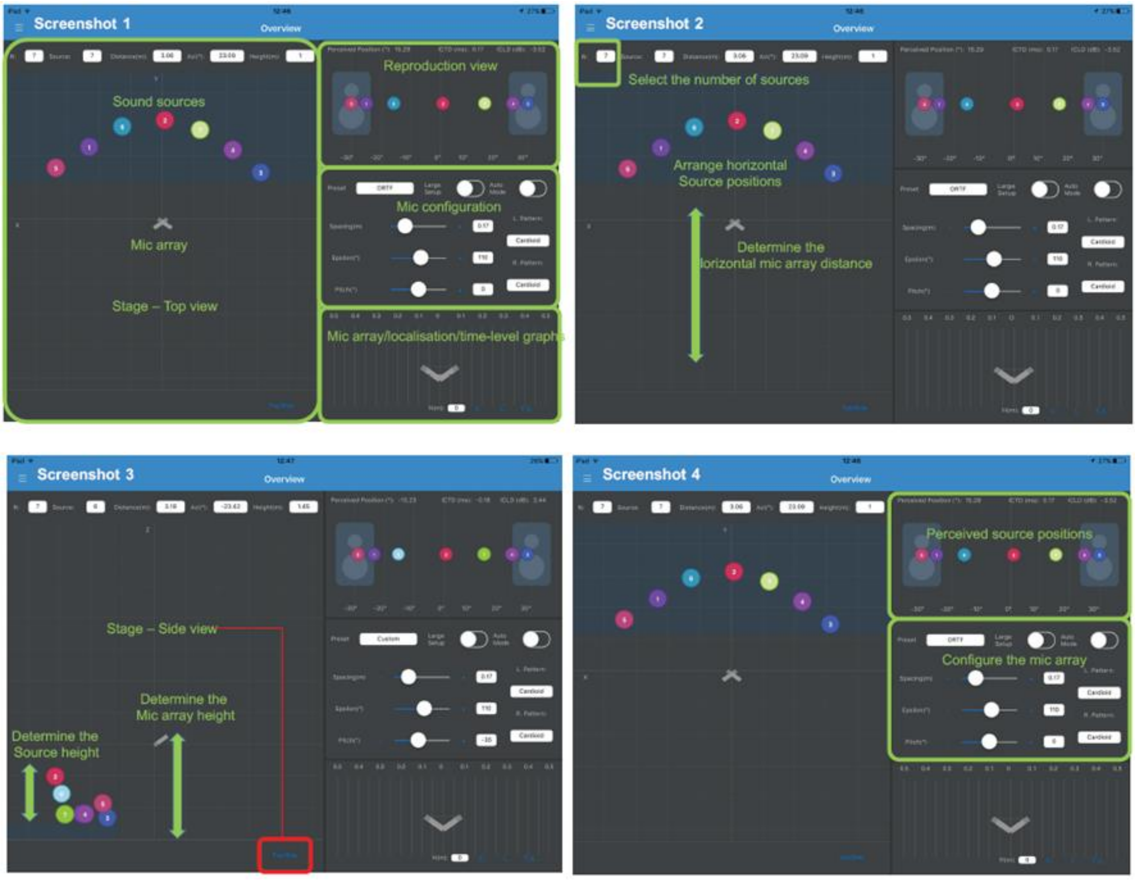


Figure 9.1.7: MARRS Microphone Array Recording and Reproduction Simulator (Lee *et al.*, 2017b)

In addition, care was taken that the chosen arrays would allow to capture the different temporal parts of an acoustical event (direct sound, early reflections, reverb) in such a balance which would allow to achieve the desired conditions for the Main- and F/R-Array comparison. The significance of the different sound attributes for a specific spatial impression can be seen in Figure 9.1.8:

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

Figure 9.1.8: Interrelation Between Sound Attribute and Spatial Impression (Theile & Wittek, 2011, p. 5)

9.1.2 The Importance of Early Reflections in Recording

As the *"spatial impression is the advantages [sic] of three-dimensional multichannel sound recording"* (Hamasaki & Van Baelen, 2015, p. 3), it makes sense to summarise the role of specific sound components on spatial perception and the practical consequences on the design of 3D microphone arrays.

Besides the strong dominance of direct sound in a sonic event based on its high influence on source imaging and timbre (see Figure 9.1.8), early reflections seem to be of particular interest in the context of 3D recordings:

> *"A single reflection coming in from a specific direction – say, the top-right corner of the rear part of the hall – should be reproduced as such; it must not be picked up by the 'wrong' mikes. This would be the case, for example, when using omnidirectional microphones in a room-microphone array [corresponding to the ambience array of the F/R-Array in the current investigation]…Perception is particularly stable when the reflections come in from the original directions of the upper half space… Therefore, the reproduction must be absolutely consistent with a real spatial environment."* (Theile & Wittek, 2011, p. 6)

Theile & Wittek got affirmed by Hamasaki & Van Baelen stating that

> *"the direction of each early reflection and late reverberation has a much more important role in creating spatial impression for natural sound recording. Therefore, it is necessary to situate the rear and upper microphones in order to catch the reflections coming from behind and above a listener."* (2015, p. 3)

Furthermore, Howie *et al.* observed that LEV seems to increase *"if there is adequate spatial balance in the direction of arriving reflections."* (2016, p .2)

As confirmed by Howie *et al.*, all of this leads to the conclusion that ideally, the microphone capsule direction should roughly mirror playback speaker direction. A possible exception can be the top front microphones which may need to be angled away from the sound source in order to minimise direct sound capture (see Chapter 9.5.5). This does not mean that the microphone placement is dictated by the layout and relative distances between reproduction loudspeakers. "*Rather, microphone placement and optimization should be based on capturing an ideal reverberant sound field."* (2016. p. 4) This, on the other hand, confirms the claim of Bowles stated in Chapter 2.1 that it is necessary *"to keep the height channel information specific to sound arriving from above"* (2015, p. 5) and also emphasises the importance of ceiling reflections (*ibid.*, p. 1).

Accoring to Theile & Wittek, giving special attention to the directions of early reflections during the recording process gives the human ear the chance to recognise and understand the reflection pattern of a sound event, which is crucial to imaging a spatial environment realistically (2012, p. 7). Figure 9.1.9 depicts a reflection pattern of a sound event in a schematic way:

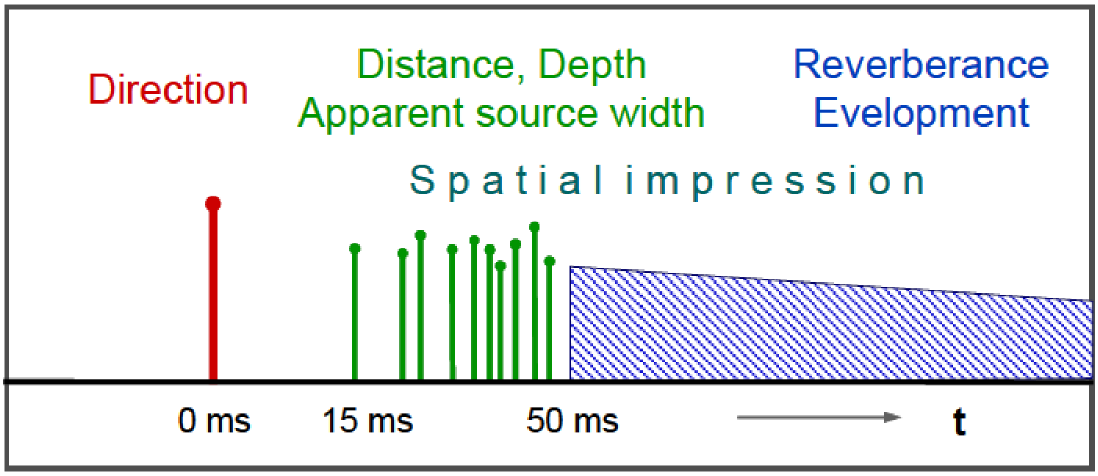


Figure 9.1.9: Influence of Attributes on Sound Impression over Time (*ibid.*, p. 5)
(Red; Direct sound, Green; Early reflections, Blue; Reverberation)

Key parameters of reflection patterns include the timing structure in relation to direct sound, levels and spectrum, and, as stated above, the horizontal and vertical incidence directions. As already visualised in Figure 9.1.8, indirect sound allows for reproducing the recording space, whereas the relation between direct and indirect sound determines the spatial attributes of a sound event (*ibid.*, p. 6).

Early reflections (occurring at a delay of 15 to 50 ms) play a key role in spatial hearing and are most important as they affect the perception of distance, depth, and spatial impression and thus deserve special attention when it comes to recording (*ibid.*). The importance of early reflections on spatial impression and the influence of the microphone array on their capture can be seen as a possible reason why there seems to be a common sense about the importance of capturing ambient sound from specific directions of the recording space, as could be seen above.

Apart from this, when studying the setups of the different studies it was observed that great care was taken to capture lateral (early) reflections. Originally their importance for the perception of spatial impression has been shown by Barron who found *“that the degree of spatial impression is probably related to the ratio of lateral to non-lateral sound arriving within 80msec of direct sound.”* (1971, p. 1) The influence and importance of lateral sound energy and reflections for spatial impression in terms of LEV was later confirmed by Hanyu *et al.* (1999).

A practical implication of this can be seen in investigations such as *Exploratory microphone techniques for three-dimensional classical music recording* by Howie & King where both, the main and the height layer contained cardioid microphones oriented 90°away from the sound source, aiming straight to the side-walls as can be seen in Figure 9.1.0. When evaluating the recording, they found that the lateral channels of their main layer seemed to increase envelopment, even if it didn’t contain enough diffuse decorrelated information (2015, p. 4).

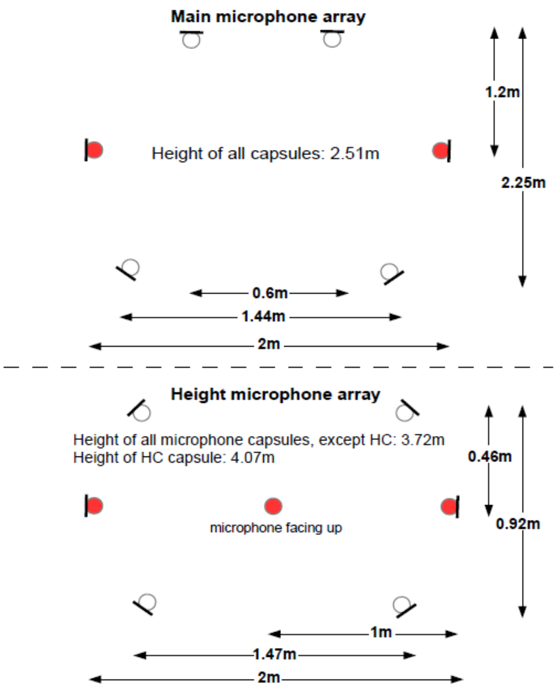


Figure 9.1.10: Microphone Arrays Top View (*ibid.*, p. 3)

However, as could be seen in Figure 9.1.8, it is not only the early reflections (including the lateral early reflections) contributing to the spatial impression, but also the reverberation. Reverb becomes important when dealing with the attribute of envelopment, as early reflections would not contribute to its perception (see Figure 9.1.8). As can be seen in Figure 9.2.2, different arrays will produce different reflection patterns when listening to the recording of the same acoustic event in the sweet spot of the loudspeaker setup. Therefore, it is important to understand what the possibilities and limitations of different array systems are when it comes to spatial impression (early reflections) and envelopment (reverberation).

9.1.3 Useful Tools: Acoustic Pressure Equalisers / Diffraction Attachments

> *"The diffraction attachment emulates the directional characteristics of the famous Neumann M50 microphone, which features a small diaphragm placed in an acrylic sphere, and exhibits a cardioid polar pattern in the upper mid and high frequencies, while remaining omnidirectional in the lower range. One example of these attachments is the Schoeps KA40 for the MK 2H microphone."* (King *et al.*, 2016, p. 2)

APEs (see Figure 9.1.11) increase the microphone directivity in the *"presence range"* between 1 and 4 kHz (Schoeps.de, 2018b) and thus offer an improved channel separation and localisation accuracy. These are some of the main reasons Morten Lindberg would use them in his 2L-Cube (Lindberg, 2015, p. 15).

In the context of a Main-Array, the diffraction attachments for the backward facing surround omnis allow them to be placed closer to the source due to their increased directional behaviour in the HF content, and thus allow to capture *"a more dynamic and therefore interesting ambient program.* Many engineers have adopted this method, although several other techniques have been developed and used with equal success*."* (King *et al.*, 2016, p. 2)

Figure 9.1.11: Schoeps MK2H Omnidirectional Microphone with a KA40 Diffraction Attachment (*ibid.*)

9.1.4 Order of Procedure when Choosing the Setup

At this point the question arises in what order each of the parameters of an array should be analysed in relation to the characteristics of the sound source and its acoustic environment. Williams proposes the following order (2010, p. 15):

1) Studying the frequency response curve of a microphone especially with regards to the bass response in relation to the directivity of the microphones (referring to the increasing LF roll-off of directional microphones)

2) Position of the microphone system, optimised for the best possible D/R ratio and balance between the individual sound sources (the arrays proposed by Williams belong to the category of the Main-Arrays, explained in Chapter 2.2)

3) Stereophonic sound image: The SRA is chosen according the sector occupied by the sound source

4) Angular distortion: not applicable, as the source is a single instrument placed in the centre

5) Mono/Stereo computability: not applicable, as for stereo a separate AB array is used, and the likelihood of the recorded material needed in mono is negligible

9.2 Justification of the Chosen Arrays

9.2.1 Main-Array

<u>Main Layer</u>

The wide A/B was taken as a first basis to start with and uses 9 omnidirectional microphones, one microphone for each channel, placed according to the Auro-3D speaker configuration, as shown in Figure 9.2.1. There are several reasons speaking in favour of this configuration in the current context:

Firstly, omnidirectional microphones have been preferred consistently as a main system for classical music recording (King *et al.*, 2016; Bowles, 2015, p. 5; Hamasaki & Van Baelen, 2015, p. 4). The reason for that is mainly related to the tone colour, as an omni-directional microphone has much more richness in the low frequency range compared with a cardioid microphone (*ibid.*).

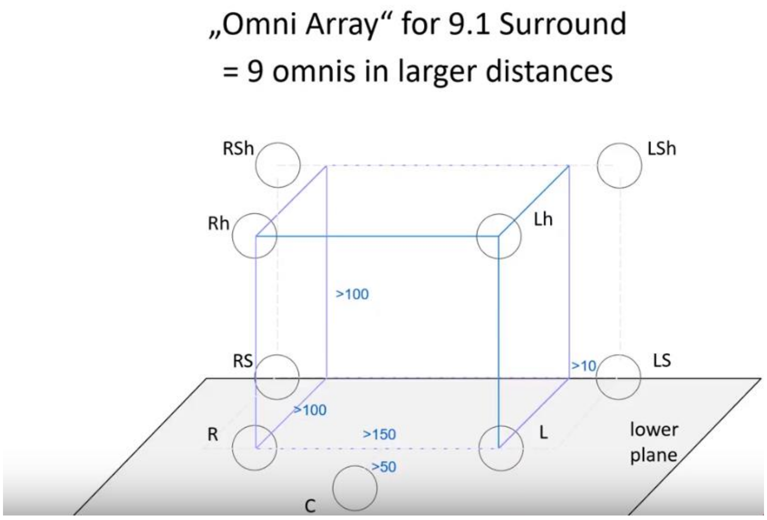


Figure 9.2.1: (Wittek, 2018)

Secondly, the increased amount of reverb of this setup is supposed to enhance the perception of envelopment (see Figure 9.1.8), although at cost of localisation accuracy and the impression of depth (Theile & Wittek, 2011, p. 10). The reason can be seen in Figure 9.2.2, depicting the reflection patterns of two different arrays in the sweet spot of an Auro-3D speaker array (see Chapter 2.1). The OCT setup was designed to optimise the channel separation. Its configuration consists exclusively of directional polar patterns and thus takes the directionality of early reflections into account (see Chapter 9.1.2), resulting in a clear pattern with separated groups of early reflections which is necessary for the perception of depth, and spatial impression (see Figure 9.1.8 and Chapter 9.1.2). The large AB, on the other hand, does not provide a clear pattern of early reflections in its capture as can be seen in Figure 9.1.8. In turn, the reverberant contribution starts soon (compared to the OCT-array, as depicted).

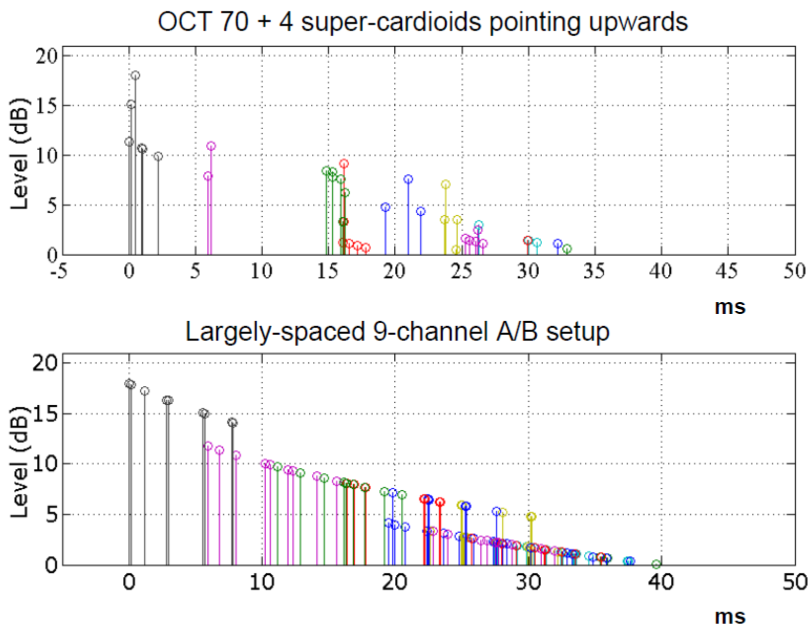


Figure 9.2.2: Reflection Patterns of AB and OCT Microphone Arrays (Theile & Wittek, 2011, p. 10).

Theile & Wittek describe this phenomenon the following way:

> *"[In the AB setup] obviously, there are hardly any utilizable discrete reflections, and reverb builds up very quickly. Even the direct signal has a wide and reverberant character; however, this may actually be desired: Recording in long-reverb spaces where the diffuse-field (the envelopment) dominates the listening experience – for example, in a church – results in a great surround sound; presence and imaging stability can still be enhanced using spot microphones. Achieving a degree of imaging, depth, and distance perception corresponding to the recording room will definitely not be achieved."* (*ibid.*, p. 11)

This was confirmed by Riaz *et al.*, stating that

> *"the combination of spot microphones and individual position A arrays [including an omni rig] works well to aid localisation of individual sound sources and capture more of the room's ambience inducing a greater sense of the recording space."* (2017, p. 5)

Also regarding the factor of DF-decorrelation for achieving the perception of envelopment (see Chapter 9.1.1), this array has some advantages:

> "*Regarding recording/reproduction of diffuse sound such as reverberation, the largely spaced A/B ensures maximum de-correlation [sic] even at the low frequencies and thus supports natural immersive perception of the surrounding ambience."* (Theile & Wittek, 2012, p. 15)

This is worth mentioning as the Main-Array will have to provide sufficient decorrelated diffuse sound, as there is no separate ambience array providing diffuse sound only. As the recording took place in a church where the reverb dominates the listening experience, the addition of spot microphones in a three-dimensional omni setup to improve localisation was found to work well (Riaz *et al.*, 2017, p. 5), and a largely spaced AB setup provides a maximum LF decorrelation, this setup was considered worth considering as a first basis. Furthermore, widely spaced AB techniques were also genuinely preferred in an informal microphone array comparison at the ICSA 2011 (Zielinsky, 2018).

This 9-channel omni array, on the other hand, can be considered as the basis for Morten Lindberg's 2L Cube, which can be seen in Figure 9.2.3 and

"is really a direct consequence of the speaker configuration in the AURO 3D playback system. Time of arrival, SPL and on-axis HF texture is directly preserved in this 5.1.4 or 7.1.4 microphone configuration. Proportions are cubical and the dimensions could vary from 120 cm for a large orchestral array down to 40 cm in an intimate chamber musical context. I always use omnis in the Main-Array. But depending on the room, the music and the instruments I alternate between the DPA 4003 and the 4041 with the larger membrane, the latter providing a more focused on-axis texture." (Lindberg, 2015, p. 15)

Figure 9.2.3: 2L-Cube by Morten Lindberg (Lindberg, 2018a)

As could be seen in Figure 9.2.3, Lindberg takes advantage of APEs, whose advantages are described in Chapter 9.1.3.

The common basis of these two approaches, being omni based with the additional option of using APEs (Lindberg) to achieve increased HF directionality for localisation and an improved signal separation as demanded by Theile & Wittek (see Chapter 9.1.1) led to further considerations in the direction of a Bowles-Array which will be introduced in the next section.

The Bowles-Array (see Figure 9.2.4), was taken in consideration as it would allow to combine all the different requirements discussed so far in one array. Firstly, it consists of for four omnidirectional microphones and depending in the size of the source and the and/or the desired imaging, a directional microphone can be used for the centre channel. This is important to mention as in case of a Main-Array there are arguments against a centre microphone, as it further complicates ICLD and ICLT relationships between the front channels with possibly negative consequences on dynamics and timing (Faulkner, 2019b) and the danger of narrowing the frontal image:

> *“It is relatively difficult to use the center microphone to obtain adequate sound image localization in the frontal sound field compared with an ordinary two-channel stereo microphone. It is necessary to adjust the level of the center microphone carefully to find the point of compromise between stereo image width and sound image localization.”* (Hamasaki & Van Baelen, 2015, p. 3)

Furthermore, it is much easier this way to obtain the desired channel separation in the front channels as the distance between the mikes is increased. Also, it adds another interesting factor to the comparison, as the FR approach entails a centre channel. Another difference to the wide AB and 2L-cube is the increased spacing of the surround microphones compared to the front pair. Similarly, this increases the DF-decorrelation, being another important requirement for the ambience part of stereophonic arrays to achieve an increased impression of spaciousness (see Chapter 9.1.1).

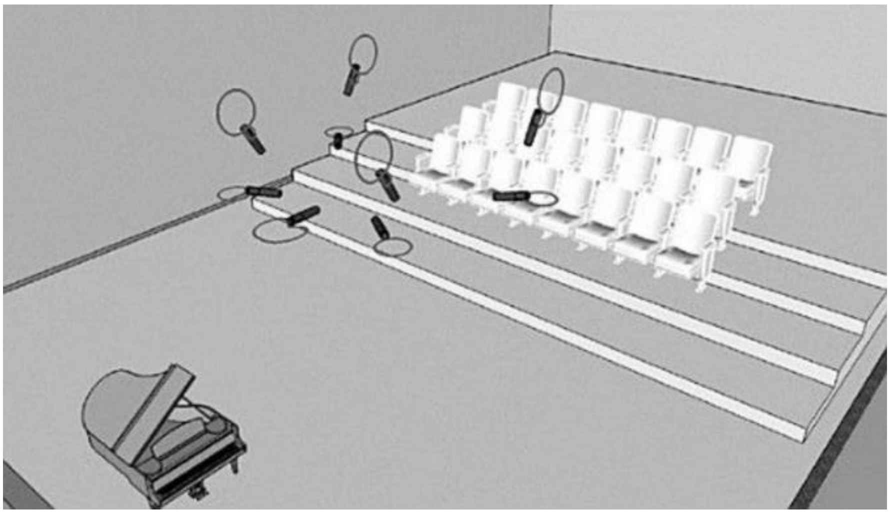


Figure 9.2.4: Bowles-Array (Roginska, & Geluso, 2018, p. 239)

As omnidirectional mics become more increasingly directional past 1 kHz, the microphones of the main layer are angled downwards to 30-45° (Bowles, 2015, p. 5). The height channels are directional, either supercardioid (Roginska & Geluso, 2018, p. 238) or hypercardioid (*ibid.*, p. 6) to overcome the possible problems when using omnis as height microphones.

> *"The height layer was designed to capture sound reflections coming from the ceiling and higher areas of sidewalls. Therefore, the [height layer of the] array points 30° upward from the horizontal plane, rather than pointing the microphone's axes directly above… This way the front two height microphones pick up a higher concentration of front ceiling and high wall reflections. Similarly, the two rear height microphones pick up more of the rear ceiling and high wall reflections. Like other microphone arrays that capture height information, the distance between the main and height will vary."* (Sungyoung Kim as in Roginska & Geluso, 2018, p. 238)

Furthermore, the height microphones are angled outward to provide a better separation between left and right height channels, resulting in a wider image in the height layer, which helps to define the height layer as distinct from the main layer (Bowles, 2015, p. 6).

All these aspects seemed to take into account many of the requirements mentioned in Chapters 2.3 and 9.1.1.: The possibility of vertical coincident spacing, the increased channel separation and DF-decorrelation for the ambient sound through the use of directional mikes and the increased rear spacing, increased channel separation at the front (stereo instead of LCR), the necessary separation between the main and the height layer through the use of directional height channel microphones (compared to an omni only setup), the directional capture of lateral early ceiling reflections (also due to the directional pattern), and last but not least the use of omni microphones in the main layer for a preferred frequency response.

A compromise was made, however, when choosing cardioid microphones for the height layer: The high level of rear sound rejection of the cardioids allows for placement and angles in favour of ambient sound capture, and their directionality allows for an increased decorrelation between the ambient signals. Furthermore, cardioids are less prone to LF loss than the originally used hypercardioid or supercardioid microphones, whereas the supercardioid pattern is used already in the Hamasaki-Cube (see next Chapter), and allows a certain degree of comparison. At the same time Howie *et al.* claimed that a certain degree of LF roll-off is desirable as this would increase the pitch height effect (2016, p. 4). Bowles, on the other hand stated the problem of *“excessive low-frequency presence in the height layer, making those sound sources poorly localized”* (2015, p. 5) and mentioned the issue of comb-filtering, similarly to Lee & Gribben (2014).

Further reasons against using an omnidirectional pattern for the height layer like in the wide AB or 2L-Cube are that firstly it is undenied that direct sound in the height channels has to be suppressed sufficiently (see Chapter 9.5.3). Also, while it was claimed, that the omnidirectional height microphones provided a *“rich, even room sound”* (King *et al.*, 2016, p. 2), the same authors noted that they

> “*contained too much direct sound and not enough decorrelated or diffuse sonic information. This was especially noticeable in the 2 front height channels, which had a tendency to 'pull' certain elements of the image up toward the higher ring of loudspeakers as the signal was raised in level.”* (Howie & King, 2015, p. 3)

In any case, omnidirectional microphones would never reach the desirable minimum of 9.5dB ICLD to achieve the localisation threshold in a coincident placement, as two coincident omnis display no ICLD. Therefore, comb-filtering and image shifts would be expected (see Chapters 9.5.3 and 9.5.4).

The compromise of an omni microphone with diffraction attachments was considered but no official listening test has been conducted so far to find out whether there is any preference for the “omni plus attachment” or cardioid pattern for the height layer microphones. Informal comparisons suggest that the “omni plus attachment” solution sounds more natural and even in timbre with a sufficient amount decorrelation from the direct sound, whereas the cardioid pattern seemed to provide more control over the vertical image (King *et al.*, 2016, p. 4).

As the height channels need to be balanced fairly high in the mix to have enough impact (see Chapter 2.1), an omni-directional mic was not considered suitable for the height layer, especially in combination with the vertical coincident placement as the pattern is still omnidirectional up to the HF range. The option of an "omni plus attachment" seemed still way too much of a risk for uncontrolled direct sound capture, which would limit the possibilities of comparisons and combining the two approaches during the mix to a big extent. Also, based on the explanations in Chapter 9.5.3, rejection of direct sound in the height channels gives more freedom in mixing, regardless of following a main or F/R-Array approach. As the Main-Array has no possibility to change the D/R ratio in the main layer (unlike the applied F/R-Array approach described in the next Chapter), it was considered important that at least the height channels would provide enough flexibility during the mixing stage. For all these reasons, the choice fell on the cardioid pattern for the height layer in the Main-Array.

9.2.2 F/R-Array

The foundation for the FR was mainly the study *An Informal Comparison Between Surround-Sound Microphone Techniques* by Kassier, Lee, Brookes and Rumsey, comparing different combinations of front and back arrays for their preference (2005). Although this study investigated the preference of surround arrays and not 3D arrays, it was considered as a useful basis for further decisions, as 3D arrays are based upon surround arrays. The arrays in question were the Fukada-Tree, OCT, INA-3 and a near coincident technique for the front, and the IRT-Cross, Hamasaki-Square, a dummy head and a spaced two-channel cardioid technique for the back. It turned out that there was a clear overall preference for the Fukada-Tree for the front, and the Hamasaki Square for the back, as can be seen in Figure 9.2.5:

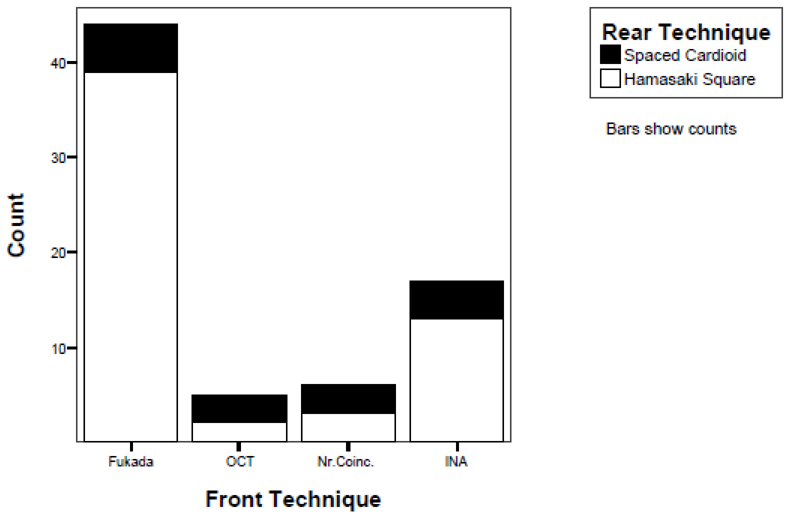


Figure 9.2.5: Overall Preference of all Subjects and all Programme Items (*ibid.*, p. 11)

The setup of the Fukada-Tree and Hamasaki Square in the mentioned study can be seen in Figures 9.2.6 and 9.2.7. They have been spaced 7m apart from each other (*ibid.*, p. 7):

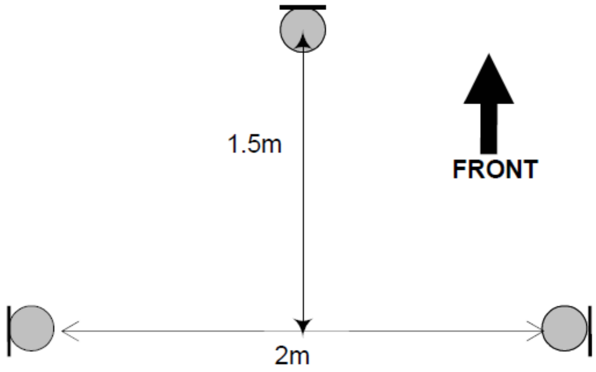


Figure 9.2.6: Fukada-Tree in Kassier *et al.* (2005, p. 3)

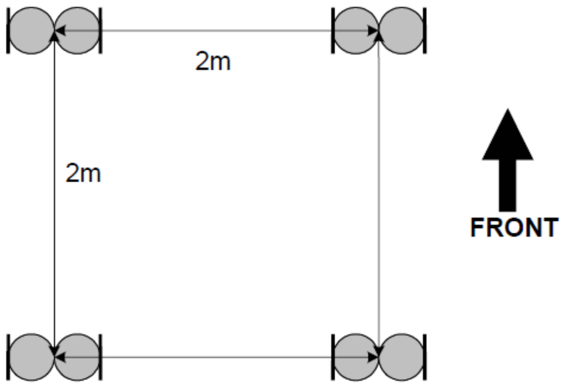


Figure 9.2.7: Hamasaki Square in Kassier *et al.* (2005, p. 6)

As can be seen in Figure 9.2.5, the dummy head and the IRT-Cross are missing. The authors' initial informal listening tests showed that only a surprisingly small difference in perception could be noticed between the Spaced Cardioid technique and the dummy head, as well as between the IRT-Cross and the Hamasaki Square. It was therefore decided to reduce the number of rear techniques involved in the test to one two-channel technique (Spaced Cardioid) and one four-channel technique (Hamasaki Square). Anyway, the IRT-Cross may not have been the most appropriate solution in this context: Holman states that there is the limitation that *"some direct sound will reach especially the front facing microphones and pollute its use as a pickup of principally reverberation"* (Holman, 2008, p. 93), which, in that case would defeat the purpose of its use as an ambience array. This was confirmed by Theile, who suggests that *"this array [Hamasaki-Square] is a better option for achieving good spatial impression compared to the IRT-Cross."* (Kassier *et al.*, 2005, p. 6)

In summary, there was a general consensus amongst the subjects of the study that the Fukada-Tree provided a natural distance to the recorded sources, an appropriate source width, a balanced spatial context, timbral balance and an increased low frequency extension of the tested front arrays. The Hamasaki Square, on the other hand, seemed preferred, as it blended the ambience from the rear with that from the front and was more coherent, produced wider sources, was more enveloping and exhibited less phase issues. All these attributes have been defined according to Rumsey's scene-based paradigm (Kassier *et al.*, 2005, p. 13).

Considering these statements and the clear overall preference of the Fukada-Tree (see Figure 9.2.5) it was decided to further investigate the technique of the Fukada-Tree. Further considerations related to the Fukada-Tree have been made taking into account that the array at the back will be the Hamasaki-Cube configuration. The Hamasaki-Cube was chosen on the basis of the strong preference over all subjects and items of the Hamasaki Square in the mentioned study (see Figure 9.2.5). It is an extended version of the Hamasaki Square, with four additional supercardioid height microphones, placed in a square on top of the Hamasaki Square, thus creating a cube (Hamasaki & Van Baelen, 2015, p. 4).

The Fukada-Tree is based on a Decca Tree, using cardioids instead of omnis *"to reduce the amount of reverberant sound pickup by the front mikes."* (Rumsey, 2013, p. 197) In the context with the Hamasaki-Cube, this seemed to be of major importance, as the Hamasaki-Cube does not provide any localisation cues, since it is especially designed to reject direct sound (as explained later in this Chapter), and direct sound, as could be seen in Chapter 9.1.1 plays a major role also regarding other parameters in an auditory event (see Figure 9.1.8). However, care has to be taken when using Main-Arrays consisting only of cardioid microphones: Firstly, their pickup exhibits a lack of LF content compared to omnis and the need to compensate for this was stated by Hamasaki & Van Baelen:

> *"An omni-directional microphone has much more richness in the low frequency range compared with a cardioid microphone. When directional microphones are used for recording multichannel sound, it is necessary to have a supplementary means of making up for the loss of energy at low frequencies."* (Hamasaki & Van Baelen, 2015, p. 8)

Secondly, it was found by Luthar and Maltezos, that

> *"using only the KM140's [cardoid] as the LCR microphones caused an auditory disconnect between the height channels and the horizontal sound stage, which was not found when the LCR were either all omnidirectional or with a cardioid in the center. The use of only cardioids had a tendency to flatten the frontal image and the sound was perceived as more aggressive and lacking any sense of 'air'… The positive effect of the height channels was minimized, however, when the LCR configuration was all KM 140s."* (2015, pp. 3 & 4)

It must be noted, that the setup of Luthar & Maltezos did only consist of the Main-Array approach (three microphones at the front, two omnis at the back, and four omnis for the height layer). Therefore, it did not include a separate ambience array at the back which could have been used to even out certain balance issues in the perception of ambience through the routing of ambient sound into the front channels which have been the ones being *"aggressive and lacking air."*

Nevertheless, combining these statements this means that when using the Fukada-Tree as a front array, entailing three cardioid microphones, a way has to be found to

a) Compensate for the lack of LF content
b) Compensate for the possible lack of “air” and reducing “aggressiveness” to avoid a perceived separation of the main and height layers

Luckily, further investigations revealed that both, the Fukada-Tree and the Hamasaki method provided means to reduce these issues: Firstly, *“the widely spaced outer pair [of the Fukada-Tree] should produce a large interchannel time difference, providing a good sense of 'spaciousness' and 'openness'…”* Kassier *et al.*, 2005, p. 3)

Before proceeding further it should be mentioned that there is an extended version of the Fukada-Tree being a surround array with separate treatment of front and rear. As can be seen, the Fukada surround configuration would normally entail two backward facing cardioids for the surround channels. However, according to Holman it can also be completed with an IRT-Cross or a Hamasaki-Square instead (Holman, 208, p. 94). Either way the Hamasaki-Square achieved much better scores than the spaced cardioid approach in Kassier *et al.* (2005), which was one of the main reasons it was chosen to replace the cardioid pair.

Secondly, the Fukada surround array provides further options through the use of additional omni microphones, as can be seen in Figures 9.2.8 and 9.2.9:

> *“Omni outriggers are sometimes added [to the Fukada-Tree] as shown, typically panned between L-LS and R-RS, in an attempt to increase the breadth of orchestral pickup and to integrate front and rear elements.”* (Rumsey, 2013, p. 197)

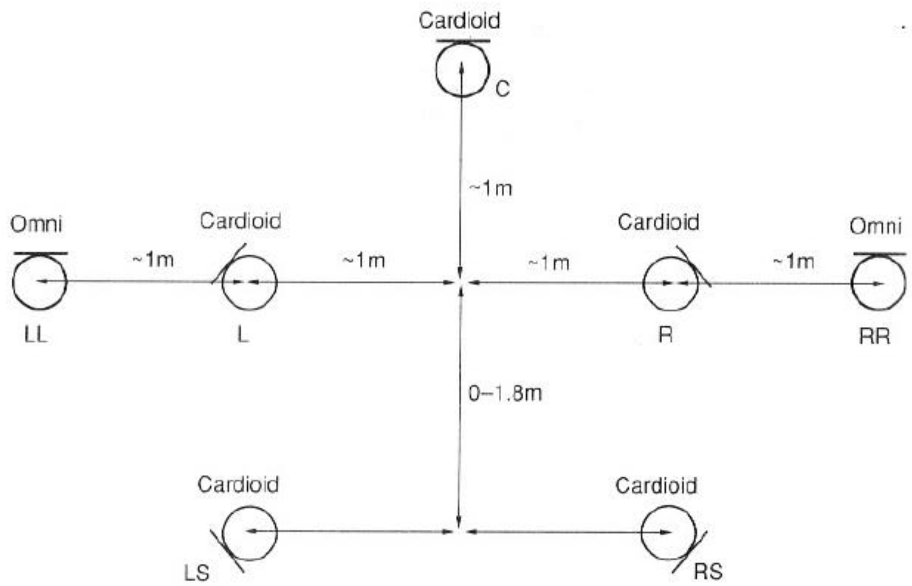


Figure 9.2.8: Fukada-Tree According to Rumsey (2013, p. 196)

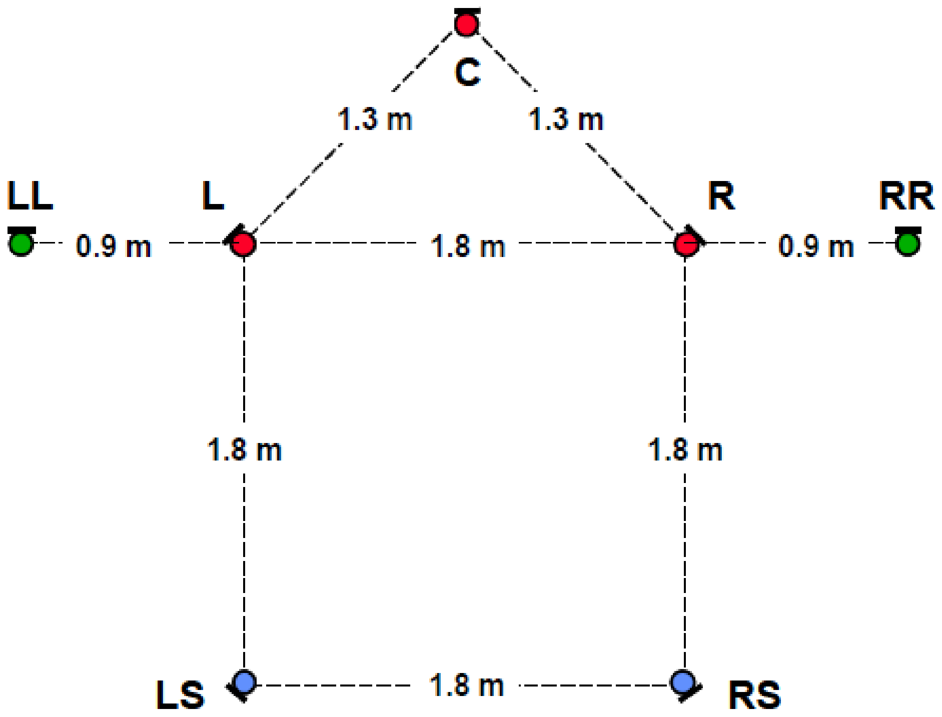


Figure 9.2.9: Fukada-Tree According to Theile (2001, p. 18)

Thirdly, when having a closer look to the Hamasaki approach in Figure 9.2.10, a similar extension can be noticed: Instead of the square only, Hamasaki proposes as well a front array, consisting of three cardiods (and a baffle), and as well two additional omnis, similar to Fukada. In this case though the omnis do not serve the purpose to be panned between the front and surround channels. They are mainly added to compensate for the lack of LF content (as stated previously Hamasaki himself declared the need to do so). *"These omnis are low-pass filtered at 250Hz and mixed with the left and right front signals to improve the LF sound quality."* (Rumsey, 2013, p. 197) Similarly, Theile proposes additional low-pass filtered omnis in his OCT array, although at 100Hz and routed to the front left and right, whereas the centre channel would be high-pass filtered above the same frequency (ibid, p. 199).

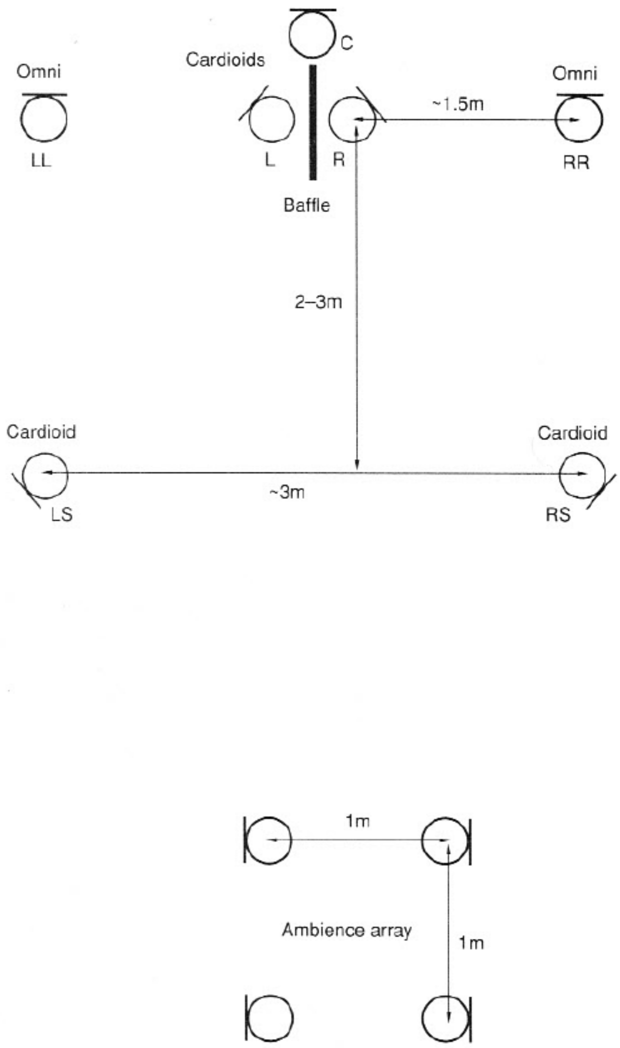


Figure 9.2.10: Hamasaki Surround Sound Recording Array (Rumsey, 2013, p. 198)

Last but not least, the Hamasaki approach has the advantage of entailing another ambience array in addition to the back spaced cardioids, which offers further possibilities in mixing to work against *"the lack of air"* and to even out the ambient sound between the front-back and main-height areas. This is necessary as in the Hamasaki approach the low-pass filtered omnis are used for providing LF content in the front channels and thus don't contribute much to the perception of "air" as this association related to frequency content above 10 kHz (Izhaki, 2018, p. 219). Therefore,

> *"the pair of microphones furthest towards the front [of the Hamasaki-Square] are [sic] routed to channels L and R or panned between L-Ls or R-RS. The degree of L-LS or R-RS panning is dependent on the amount of desired spatial information in the front loudspeakers, and also seems to rely on the headroom of spatial image in the front array that is used in combination."* (Kassier *et al.*, 2005, p. 6)

Nevertheless, care must be taken when routing these signals into the L/R channels as high levels of ambience sound will reduce the localisation accuracy (Riaz *et al.*, 2017, p. 5) which would defeat the purpose of having cardioid microphones in the front array.

Regarding the need to convey an adequate impression of the room the importance of capturing early reflections from specifically above (ceiling reflections) and the sides (lateral reflections) was stated in Chapter 9.1.2. In the Hamasaki-Cube this is perfectly realised as the four bi-directional microphones are pointing straight to the side walls and the four supercardioids straight up to the ceiling. Additional lateral early reflections will be captured by the Fukada-Tree through the off-axis angle of 90° aiming towards the side walls. This allows at the same time for a more focused centre image within the LCR arrangement. Apart from capturing the preferred directions of early reflections, the Hamasaki-Cube allows for a maximal rejection of direct sound, which on the other hand greatly improves the flexibility during the mixing stage.

This optimised capture of ambient sound becomes even more important considering that in the case of the Hamasaki-Square the channels will be routed not only to the height and surround channels, but to all channels except the front centre. Holman affirms this when saying that *"all directions are helpful in the production of the feeling of envelopment, so reverberation returns and multichannel ambiences should apply to all of the channels, with uncarrelated sources."* (2008, p. 188) The requirement of uncorrelated channels for ambient sound capture has been discussed in Chapter 9.1.1 and is considered fulfilled in this setup.

9.3 Justification of the Microphone Choice

9.3.1 Small or Large Diaphragm Microphones?

For the justification and fairness of a comparison, it was considered important to use microphones of one sort for the whole experiment, either small or large diaphragm, to allow for some sonic consistency. Although the better signal to noise ratio of large diaphragms would be especially advantageous for the application for a distant pickup of ambient sound where the average sound level is low (Holman, 2008, p. 74), the overall characteristics of small diaphragms was in their favour for the intended setup. This especially includes their better off-axis frequency response in terms of less colouration (Bartlett, 2017; Holman, 2008, p. 94;) and the consistency of their pickup patterns (consistent directionality across the frequency spectrum) which LDCs lack, and their overall flatter frequency response (Bartlett, 2017). The factor of the off-axis colouration plays a big roll in this setup as the majority of the microphones are placed in a considerable off-axis position and the resulting colouration could impact the consistency of the comparison between the arrays. In that respect, the main reason for placing the microphones off-axis in this setup is the reduction of direct sound. The effectiveness of this approach is strongly linked to the pickup patterns, meaning that the consistency of the directionality of the chosen polar patterns is essential. Besides that, the problems of a mix of small and large diaphragm mikes in the context of a comparison were raised by Kassier *et al.* (2005, p. 13). All these considerations lead to the decision to aim for a small diaphragm only setup.

9.3.2 Microphone Brands and Models

In a classical or acoustic source context it can be said that

> *"the most four most commonly used brands in commercial recording are DPA, Neumann, Schoeps, and Sennheiser. These companies have been in the business of designing microphones for use in main systems for a very long time."* (King, 2017, p. 41)

Apart from that, the microphones chosen for the setup of this experiment can all be considered as industry standards in the context of acoustical recordings, as they consistently reappeared in the AES publications and books listed in the reference list and bibliography.

9.4 Justification of the Microphone Placement

This chapter aims to justify any microphone spacings or angles, which are not already dictated by the specific array types themselves, which are explained in Chapter 9.2.

9.4.1 Main-Array

<u>Vertical Spacing (coincident)</u>

A detailed explanation for the use of a vertical coincident placement for the mikes for the height channels can be found in Chapter 9.5.4. However, summarised the reason can be found in the comb-filtering that occurs when direct sound is present in the height layer of a vertically spaced array, whereas no spectral magnitude reduction is caused by the addition of the coincident layer to the main layer. As the proximity of the height microphones to the source would lead with great chance to the presence of direct sound in the height channels this placement follows Lee & Gribben's suggestion of a vertical coincident placement since *"vertical interchannel crosstalk is inevitably present."* (Lee & Gribben, 2014, p. 882)

<u>Array Placement</u>

The following statement summarises the difficulties of the optimal Main-Array configuration and its placement (for the position of the current array please see floorplan, Chapter 9.6.4):

> *"The fixed positions and polar patterns of the front and rear microphones would result in an inevitable compromise between the representation of optimised directional images and spatial or room impression. For example, the front triplet should be optimised not only with respect to the recording angle of direct sound from the front but also with respect to the balance of direct and indirect sound intensity in conjunction with the rear microphones. In addition, the position and directivity of the rear microphone array should not be decided exclusively for the characteristics of the ambient sound, but also for the suppression of the direct sound due to the relatively short distance between the front and the rear microphones."* (Kassier *et al.*, 2005, p. 2)

Besides that, the array should be placed within the reverberant radius (critical distance) to capture enough direct sound (Griesinger, 1999, p. 28). As preliminary measurements of the church have been taken, an approximate volume was calculated on 5260 cubic metres. RT60 was measured with an audio analyser and found to be 1.41 seconds (the proximity of the audio analyser to the sound source was 2 metres for that value, thus simulating a possible placement of a Main-Array).

Therefore, when assuming γ to be around 0.5 (semi-directional), the critical distance of the room was calculated to be 2.46 metres (= 0.057 $\sqrt{(\gamma V)/RT60}$ = 0.057 $\sqrt{(0.5 \times 5260 m^3)/1.41 s}$). Needless to say, this calculation only served as a reference to make an informed decision regarding the placement of the Main-Array. The outcome of this estimation shows that the placement of the Main-Array as depicted in the floorplan is fully within the critical distance.

<u>Microphone Spacings</u>

> *"The basis of most of these [five-channel main microphone] arrays is pair-wise time-intensity trading… usually treating adjacent microphones as pairs covering a particular section of the recording angle around the array and possibly hoping that the signals from the other microphones will be either low enough in level or long enough delayed not to affect the image in the sector concerned too much."* (Rumsey, 2013, p. 191)

As stated in Chapter 2.2.1, the Main-Array can be considered as one compact system, which often aims to provide a 360° capture. Michael Williams developed the concept of critical linking. Critically linked arrays show no overlap of SRAs (Williams & Le Dû, 2010). Following this procedure, the first step was to determine the optimal microphone spacing between the front microphones, based on the source size, the desired imaging and the acoustic space, as suggested by Bowles (2015, p. 5). A valid starting point seemed to be a L/R distance of 67cm as *"that gives fairly decent in-phase coherent information from the centre of the image and gives a wide enough ambience that it sounds believable on its own."* (Faulkner n.d.; Simmons and Faulkner, 2016)

In a next step the MMAD software (Williams, 2009) was used to define the distance between the front and rear microphones, the two rear microphones, and all the angles. The following solution could be found, which according to Williams should ensure critical linking without applying any offsets (Williams & Le Dû, 2010):

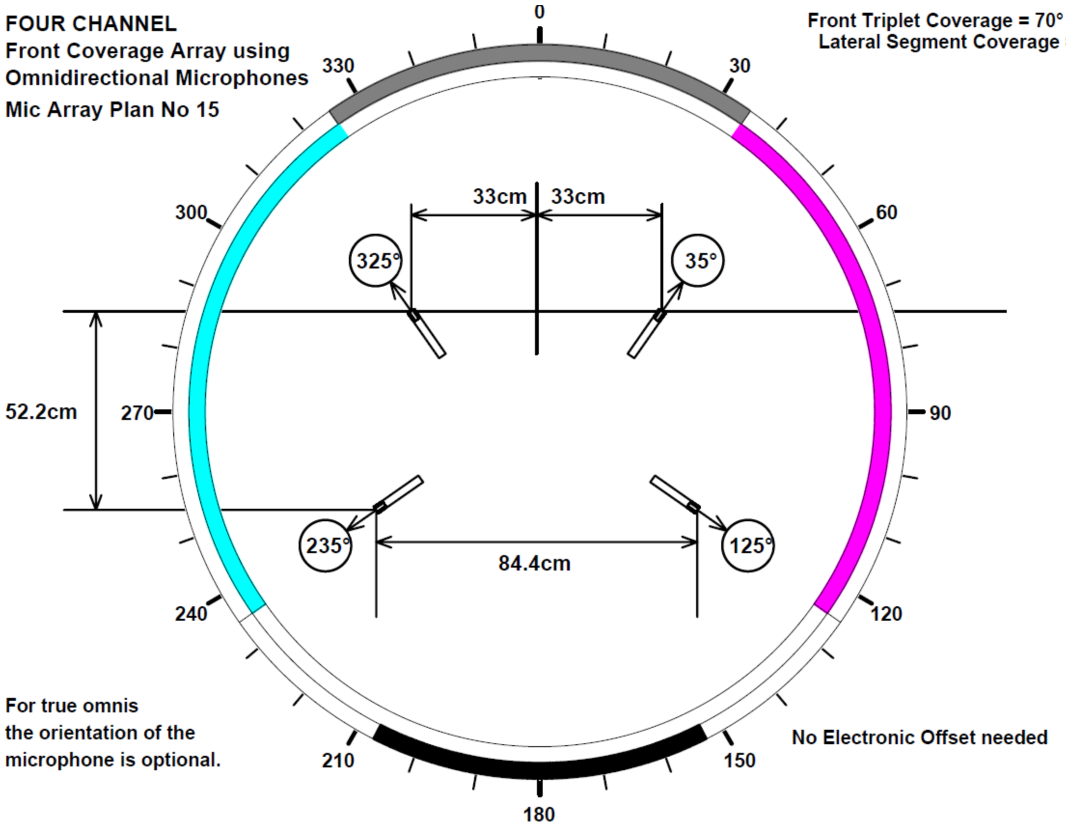


Figure 9.4.1: Critical Linking according to Williams (2009)

Angle between the Main and Height Layer Microphones

The importance of reaching an ICLD of 9.5dB (localisation threshold in a coincident setup) between the main and height layer microphones is explained in detail in Chapter 9.5.5. The necessary angle to achieve this with an omni main layer and a cardioid height layer in a vertical coincident setup was supposed to be around 120° (Lee, 2019). However, when doing a preliminary test session to ensure a sufficient ICLD between the main and the height layer (see Appendix Chapter 9.6.2), it was found that 110° would also be sufficient. The angle of 110° was found to allow for angles which could cope better with Bowles' demands regarding the axis of the microphones (see Chapter 9.2.1).

9.4.2 F/R-Array

Fukada-Tree

According to Holman the setup of this configuration has specific distances and angles (2008, p. 94), although different authors declare slightly different distances. The following section provides a short overview on the basic considerations when deciding the microphone placement for the Fukada-Tree. Holman declares a distance between the two outer cardoids of six feet, and a distance of five feet to the frontal centre mic from the centre of this line, which is similar to the suggestion of Theile (see Figure 9.2.9). Kassier *et al.* applied similar spacings as can be seen in Figure 9.2.6, leading to a SRA of 108° (2005, p. 3). Rumsey, on the other hand, suggests an equilateral triangle of around 1 meter to each side (see Figure 9.2.8).

In addition to that, *“the configuration of the tree can vary depending on the hall´s acoustic characteristics, while the microphone intervals may be changed conforming to the orchestra´s [source] size and formation.”* (Theile, 2001, p. 18)

Concerns about the Fukada-Tree have been raised based about the fairly long distances between each microphone. According to Kassier *et al.*,

> *“there is a potential problem in localisation of sound sources, as there is a strong precedent effect triggered between L&C, or C&R... Therefore, it is difficult for the Fukada Tree to achieve a balanced distribution of the phantom sources although there are three solid localisation areas…”* (2005, p. 3)

However, since in the current recording the source is not an ensemble, but a solo source placed centre, this concern was thought to be of minor importance. Apart from that, *“comments tended to indicate that subjects believed localisation was good in the Fukada Tree”* (*ibid.*, p. 13), including ensemble items.

Referring to Howie *et al.*, one might also consider placing the Front Main-Array (within the FR-Array) somewhat closer to the source than typical for a stereo only recording to minimise the ambient sound capture (2016, p. 4). As can be seen in the floorplans (Chapter 9.6.4), this was put into practice.

<u>Main-F/R-Array Distance</u>
Lee & Gribben suggest that the Hamasaki-Cube should be situated *“beyond the critical distance of a large recording venue, where the D/R ratio is below 1, in order to capture diffused sound mainly.”* (2014, p. 882) As described in Chapter 9.4.1, the critical distance of the space was estimated on 2.46 metres. Therefore, as can be seen in the floorplan (Chapter 9.6.4), the Hamasaki-Cube can be considered to be placed past the critical distance in the current setup. Furthermore, the 3:1 rule can be applied, as there are two different microphone systems in action. According to King, in such a combination of two different systems, the more distant system should be placed at least three times the distance to the source than the closer system to minimise comb-filtering (2016, p. 75). In addition to that, it has to be taken into account that

> *“the further the rear array is from the recorded sources, the more early reflections, the higher the reverberant-to-direct ratio and the higher the density of reflections. However, at least 10dB suppression of the direct sound is required in the rear channels versus the front channels [in techniques with front rear separation].”* (Kassier *et al.*, 2005, p. 2)

At the same time a spacing between microphones above 10m should be avoided according to Rumsey, as the precedence effect is breaking down for distances exceeding that value, which will result in audible echoes (2013, p. 174). As before, all these considerations have been taken into account in the current setup.

Hamasaki-Cube

The following description by Holman outlines the placement and the aspects of the Hamasaki-Square, which forms the main layer of the Hamasaki-Cube:

> *"The Hamasaki-Square array is another setup useful in particular for hall ambience. In it, four bidirectional mikes are placed in a square of 6-10 ft on a side, with their nulls facing the main sound source so that the direct sound is minimized, and their positive polarity lobes facing outwards. The array is located far away and high up in the hall to minimize direct sound [or, in other words to obtain the maximum ratio of reverberant to direct sound]. The front two are routed to L and R and the back two are routed to LS and RS. Side-wall reflections and the side component of reverberation are picked up well, while back wall echoes are minimized."* (Holman, 2008, p. 93)

Originally, the suggested distance by Hamasaki for his square was 1m between each microphone but based upon calculations and subjective listening tests that suggestion was adjusted to 2-3m (Kassier *et al.*, p. 5). Hamasaki claimed that the distance to provide enough low frequency decorrelation above 100Hz between two omni mikes in the reverberant field has to be at least 2m (*ibid.*) while Griesinger (1999, p. 44) and Rumsey (2013, p. 197) have suggested a distance greater than reverb radius of the recording venue. Furthermore, Hamasaki conducted a subjective listening test to compare the spatial impression between each pair with a distance of 1m, 2m and 3m. It was found that most listeners preferred the spacing of 3m to 2m and 2m to 1m (*ibid.*). Based on this and the findings of Rumsey & Lewis proposing a distance between rear microphones between 3 and 4m (2002), a spacing of 3m was considered to be a valuable starting point.

In summary, it can thus be said that the spacing of 3m, as applied, considers all these aspects. It even considers the demands of Griesinger and Rumsey, as the microphones are placed further from each other than the critical distance of the room (2.46m)

Opposed to the Main-Array approach the vertical spacing in the Hamasaki-Square has not been found of any major concern regarding comb-filtering as "*the issue of comb-filtering is not serious for diffused ambience signals"* (Lee *et al.*, 2014, p. 7), as explained in detail in Chapter 9.5.4. Based on the possibilities provided by the microphone stands, the height layer of the Hamasaki-Cube was set to 5m, and the main layer to 3m.

Regarding the SRA of the Hamasaki-Cube it can be said that its primary function is "*to capture as much indirect sound as possible"* (Hamasaki & Van Baelen, 2015, p. 4). This configuration is thus not intended to convey any kind of localisation cues or (phantom) source images. Neither is it the aim of the F/R-Array approach to create phantom images over 360° as stated in Chapter 2.2.1. Therefore, the parameter of the SRA was not taken into account for the Hamasaki-Cube.

9.5 Psychoacoustics

9.5.1 Horizontal and Vertical Aural Perception / ICLD, ICTD, ICCC

Firstly, the main reason for a difference between horizontal and vertical aural perception lies in the fact that the two ears are spaced apart and thus generate interaural cues.

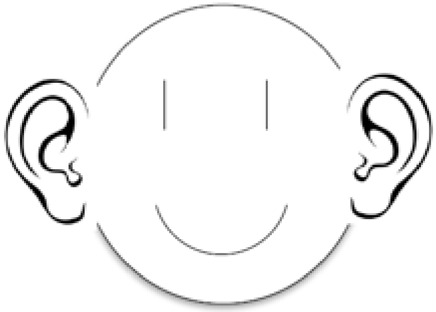


Figure 9.5.1: Two Horizontally Spaced Ears (Lee, 2018a, p. 7)

When listening to a horizontal stereo reproduction setup, the interchannel cues (Inter Channel Level Differences, Inter Channel Time Differences or Inter Channel Cross-correlation Coefficient) of the system are translated into interaural cues, as shown in Figure 9.5.2:

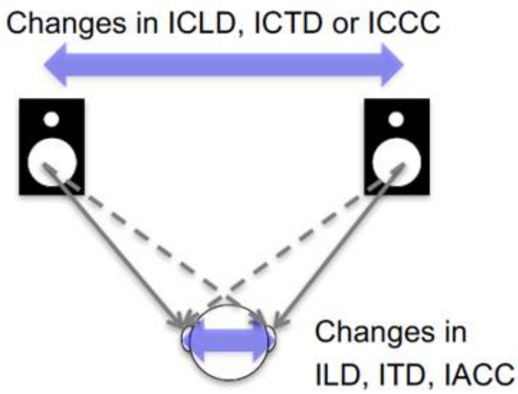


Figure 9.5.2: Horizontal Aural Perception (Lee, 2018a, p. 8)

However, when listening to a vertical stereo setup (correspondent to perceiving the median plane), no interaural changes will result. Thus, vertical localisation solely relies on spectral cues, as shown in Figure 9.5.3:

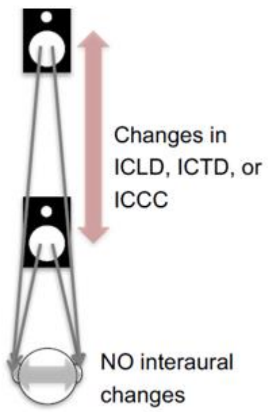


Figure 9.5.3: Horizontal Aural Perception (Lee, 2018a, p. 9)

Combining the two scenarios, leading to a similar setup as the Auro-3D setup, the vertical localisation will be generated by spectral cues, interaural cues, and the phantom elevation effect (not explained within this work).

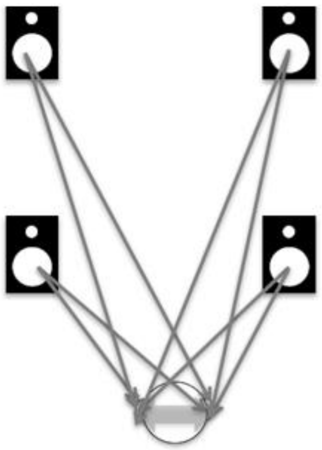


Figure 9.5.4: Horizontal and Vertical Aural Perception (Lee, 2018a, p. 11)

9.5.2 Practical Implications of Psychoacoustic Findings on Microphone Array Design

As each, main and F/R-Arrays can be applied in many different forms, whereas each form has its own purposes, this Chapter contextualises the chosen setup with recent research findings and thus demonstrates why exactly these two configurations and their hybrid version is interesting to compare. For a detailed justification of the setup, floor plans and photos please see the Appendix Chapters 9.2, 9.3, 9.4 and 9.6.

9.5.3 Vertical Interchannel Crosstalk (ICCT) / Vertical Phantom Image Shift

One of the currently most researched principles in the perception of the median plane are the effects of the vertical interchannel crosstalk (ICCT), where *"a (delayed) direct sound [is] captured by a height microphone that aims to capture ambience."* (Lee, 2018b, p. 15)

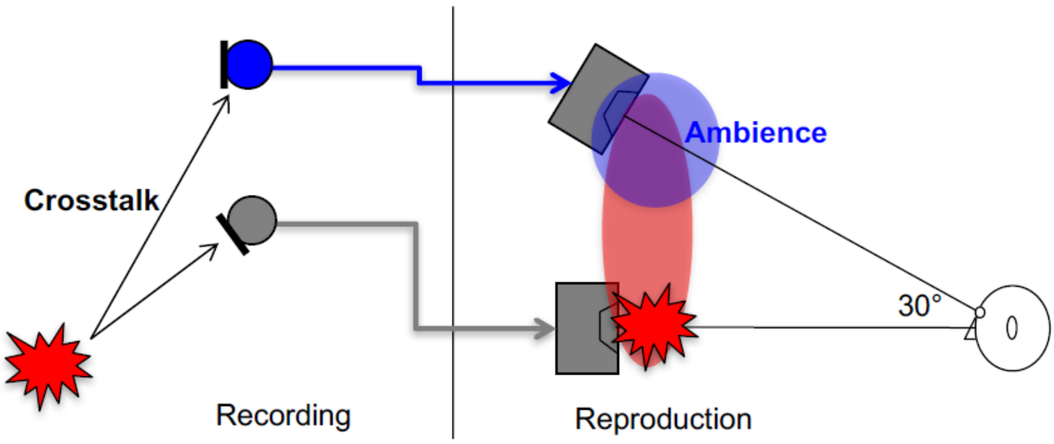


Figure 9.5.5: Vertical Interchannel Crosstalk (Lee, 2018b, p. 15)

This has practical implications on the design of 3D microphone arrays:

> *"When recording for such formats, it is necessary to pay close attention to the amount of direct sound present in the height layer signal. The reason for this is as follows. Should there be excessive direct sound in the height layer then, at the reproduction stage, sound sources may be perceived as vertically-oriented phantom images at intermediate positions between the main and height loudspeaker layers. Additional spatial and timbral effects may also be perceived, depending on the time and level relationships between the direct sounds in the respective layers. Collectively, these properties comprise an interference effect referred to as 'vertical interchannel crosstalk'."* (Wallis & Lee, 2017, p. 1)

An example of vertical interchannel crosstalk can be seen in Figure 9.5.5. Furthermore, vertical ICCT has practical implications on the microphone array design:

> *"The ability to locate sound images fully at one loudspeaker would be particularly important in designing vertical multichannel microphone techniques… This gives rise to a question as to how the distance and angle between microphones should be configured."* (Lee, 2011, p. 2)

Figure 9.5.6 provides an idea of different concepts how direct sound can be suppressed in the height channel microphone. The -6dB figure refers to the localisation threshold, which will be explained later and was adjusted in further studies to -9.5dB for vertical placements.

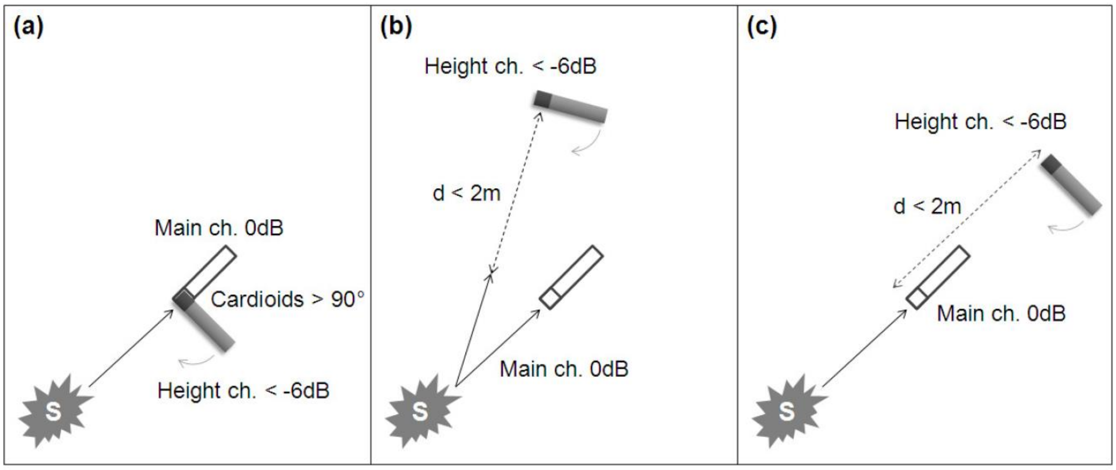


Figure 9.5.6: Height Channel Microphone Configurations (Lee, 2011, p. 9)

Regarding the research questions this would indicate that Main-Arrays are more prone to vertical ICCT and all its effects than F/R-Arrays, as Main-Arrays are placed closer to the sound source where the chance of direct sound in the height channels is increased.

9.5.4 Vertical Microphone (Layer) Spacing and Comb-Filtering

Furthermore, one of the major differences of the two chosen configurations is the vertical spacing between the microphones of the main and the height layer.
Based on the effect of vertical ICCT and the non-significant effect of vertical interchannel cross-correlation (ICC, explained later in Chapter 9.5.6), the vertical spacing has practical implications, as can be seen in the following part. In that regard, the arrays of the present study operate opposed as one constellation has vertical spacing whereas the other is vertically coincident and thus, they are based on different psychoacoustic principles.

In the current investigation, the parameter of vertical spacing is used as a means for creating different conditions with regards to ICLD, ICTD and the D/R ratio between the different arrays. The ICLD has influence on the level of ICCT, and therefore on imaging and spectral magnitude changes, the ICTD will influence the amount of comb-filtering and is also related to the ICC, and the D/R ratio is an important means to achieve the desired spatial impression. Apart from that, it has been shown that a vertically coincident microphone layer spacing was generally preferred over vertically spaced microphone positions for the attributes spaciousness and preference (Lee & Gribben, 2014), as can be seen in Figures 9.5.7 and 9.5.8:

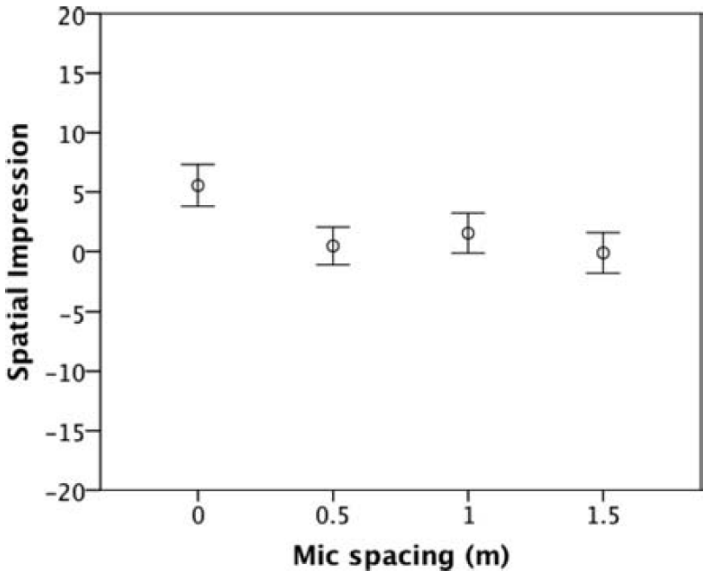


Figure 9.5.7: Microphone Spacing vs Spatial Impression for all Tested Sources (Lee & Gribben, 2014, p. 874)

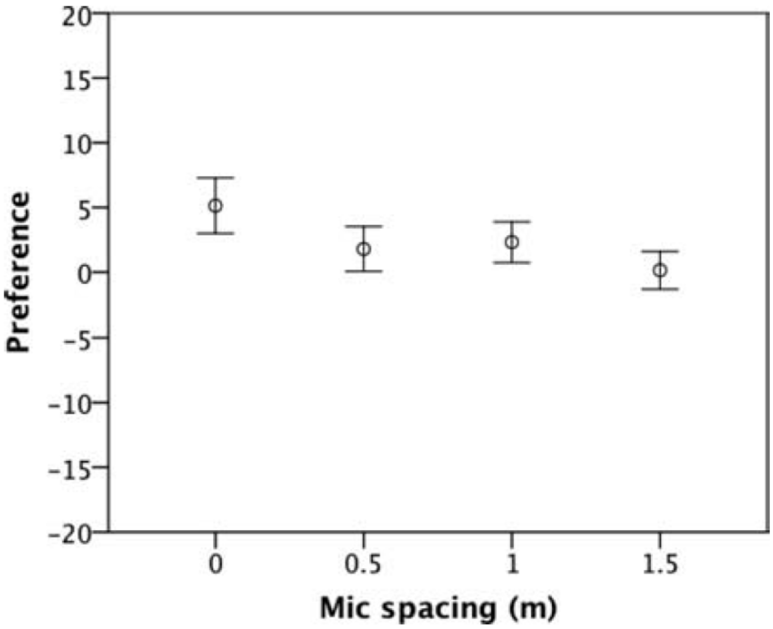


Figure 9.5.8: Microphone Spacing vs Preference for all Tested Sources (Lee & Gribben, 2014, p. 875)

The explanation of Lee & Gribben for this was that

> *"the main and crosstalk signals of the coincident layer were summed constructively at the listener's ear without comb-filtering. As shown in Fig. 12 [Figure 9.5.9 in the current paper], there was no spectral magnitude reduction caused by the addition of the coincident layer to the main layer across the whole frequency [sic]. However, the spaced layers, which had a time delay between the main and crosstalk signals, caused somewhat destructive magnitude changes to the main layer at a number of frequency regions."* (2014, p. 881 & 882)

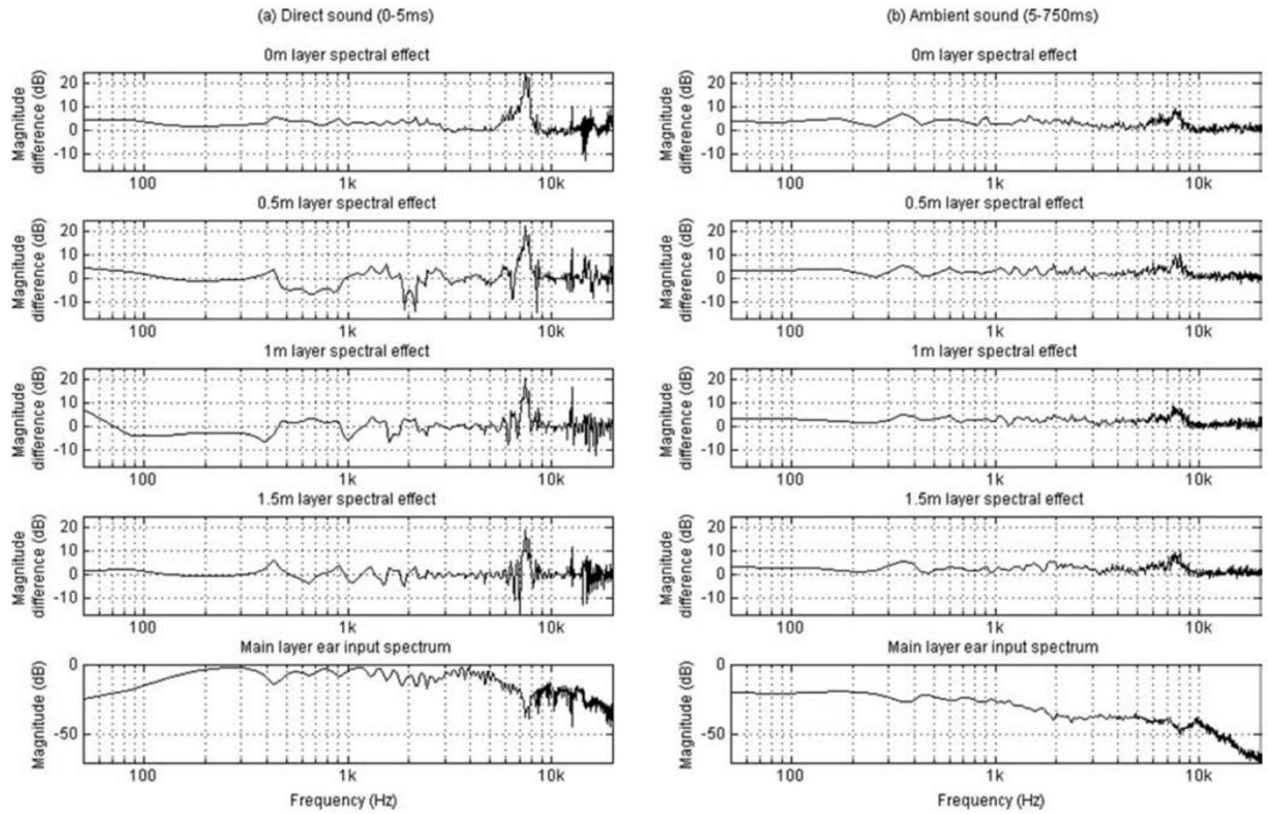


Fig. 12. Spectral magnitude differences of the left ear impulse response of the main and height layers to that of the main layer only; (a) direct sounds (0ms…5ms) (b) ambient sounds (5ms…750ms); 1024-point average FFT; the bottom panels show the magnitude frequency response of the ear input signal for the main layer.

Figure 9.5.9: Spectral Magnitude Differences Resulting Through Vertical Microphone Spacing (Lee & Gribben, 2014, p. 879)

In short, comb-filtering in this situation comes from a lack of vertical ICLD of direct sound in the presence of vertical ICTD. Therefore,

> *“it might be suggested that in practical recording situations where vertical interchannel crosstalk is inevitably present due to the desired angle and polar pattern of height microphone [sic]…, a vertically coincident 3D main microphone array could be beneficial compared to a vertically spaced array since coincident signals cause no comb-filtering at the ear.”* (Lee & Gribben, 2014, p. 882)

Nevertheless, the use of an F/R-Array works based on different psychoacoustic principles and therefore has its own right to be included in the comparison, especially when considering that

> *“the issue of comb-filtering is not serious for diffused ambience signals as can be seen … [referring to the same diagram as Figure 9.5.9]. Applying a vertical spacing between microphones in a diffused field would not be of critical concern in terms of the comb-filtering of direct sound, although it could still affect the tonal colour of the reproduced ambient sound.”* (Lee *et al.*, 2014, p. 7)

This means that both techniques will have tonal consequences, but it is difficult to say what their nature is and how they differ, which demonstrates the need of a widely-defined comparison of these two techniques.

9.5.5 Main-Array vs F/R-Array / Localisation Threshold / Masking Threshold / Vertical Image Spread (VIS)

The ambience array of the current experiment is a Hamasaki-Cube, as proposed by Lee & Gribben and placed beyond the critical distance of a large recording venue, where the D/R ratio is below 1 in order to capture diffused sound mainly (2014, p. 882). On the one hand, the lack of ICCT prevents comb-filtering to a big extent, on the other hand it reduces the possibility of VIS (Vertical Image Spread) compared to the Main-Array. At the same time, the possibility to adjust the D/R ratio is much bigger than with the Main-Array. With the Main-Array, as it most likely contains direct sound in the height channels, the range of possible D/R ratio adjustment will be limited to the localisation threshold. The localisation threshold according to Wallis & Lee is defined as

> *"the minimum amount of attenuation of direct sound necessary in the height layer for the main channel signal to be localised at the position of the main layer. It is important to note that the localisation threshold is not a complete masking of the direct sound in the height layer [as the masking threshold is not yet reached]. Instead, although the perceived location of the main channel signal would be unaffected, the aforementioned spatial and timbral effects of vertical interchannel crosstalk would remain somewhat audible."* (2017, p. 2)

As the researchers could prove differences in the resultant spectrum when a sound source was presented only through the main layer of loudspeakers compared to the main and the height layer with the localisation threshold applied, the following conclusion was reached:

> *"Direct sounds can be present in the height layer provided they are attenuated with respect to those in the main layer by either 9.5 dB (in the case of 0 ms ICTD) or 7 dB (in the case of 1–10 ms ICTD) without the perceived location of the main channel signal being affected [being the localisation thresholds for these ICTDs as in Figure 9.5.10]. Such a technique [applying the localisation threshold] could have potentially pleasing effects such as an increase in perceived VIS. However, it is currently not clear how the timbre of the main channel signal would be affected by such a technique [as differences in the resultant spectrum could be found] and, further, if the end result would be pleasing. … Such a study would make it possible to determine whether the localisation threshold should be applied or, conversely, if the direct sound in the height layer should be either masked or absent entirely. This would provide further insights on both image rendering and microphone techniques in the context of 3D audio production."* (Wallis & Lee, 2017, p. 17)

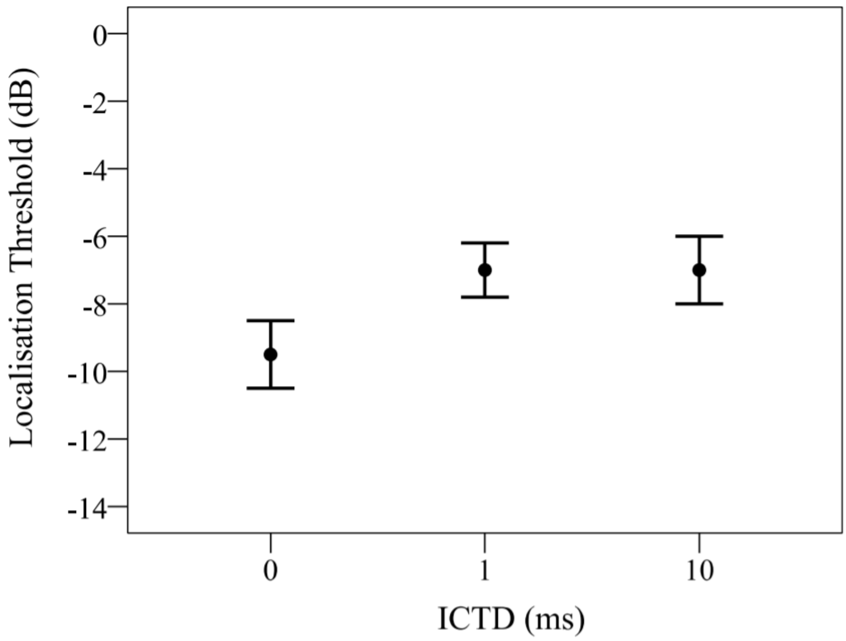


Figure 9.5.10: Localisation Thresholds for Different ICTDs (Wallis & Lee, 2017, p. 9)

This stands in contrast with the ambient array, having only ambient signals in the height layer, as demanded from Hamasaki & Van Baelen who claim that "*the rear and upper microphone's principle role is to catch indirect sound such as early reflections and late reverberation… It is necessary to capture as much indirect sound as possible, rather than the direct sound.*" (Hamasaki & Van Baelen, 2015, p. 4)

The current experiment, provides the possibility to do a widely-defined comparison of a crosstalk signal at the localisation threshold with a signal where direct sound is *"absent entirely"* (taken from the height layer of the Hamasaki-Cube) and would allow some indications on what the timbral tendencies of these two approaches are.

9.5.6 Vertical Interchannel Correlation (ICC)

However, unlike in the horizontal plane, where the ICC influences the apparent source width (ASW) to a big extent (Gribben & Lee, 2018, p. 537), it was demonstrated that *"the effect of vertical interchannel decorrelation [the equivalent to a low ICC] on vertical image spread (VIS) tends to be slight."* (*ibid.*, p. 552) *"This suggests that vertical decorrelation of signals in a practical scenario may be largely ineffective, i.e., where a direct comparison between conditions is unavailable."* (*ibid.*, p. 537) Bowles, however, claims that *"the lack of time-of-arrival differences between the main and height layers reduced the perceived differences between these two layers."* (2015, p. 5) Furthermore, it was shown that the vertical microphone layer spacing had little effect on the perception of environment-related spatial impression (Lee & Gribben, 2014, p. 883) and that the vertical decorrelation of low frequencies is unnecessary (Gribben & Lee, 2018, p. 553).

All these findings are important to note as the Hamasaki-Cube, with its large spacings is based on the principle of interchannel decorrelation to increase the perception of spaciousness, whereas the coincident Main-Array will go along with a notably higher ICC. However, although having a higher ICC the Main-Array is subject to ICCT, which is supposed to lead to a VIS and thus possibly could rival the big spacing of the Hamasaki-Cube.

9.5.7 Pitch Height Effect

The pitch height effect explains the phenomenon that a higher frequency tends to be localised at a higher position, rather than the presentation method (Lee *et al.*, 2014, p. 2). The pitch height effect is visualised in Figure 9.5.11, depicting the perceived elevation of octave and broadband stimuli for lower and upper loudspeaker presentations (top left and top right) as well as for 0ms (bottom left) and 10ms (bottom right) stereophonic conditions:

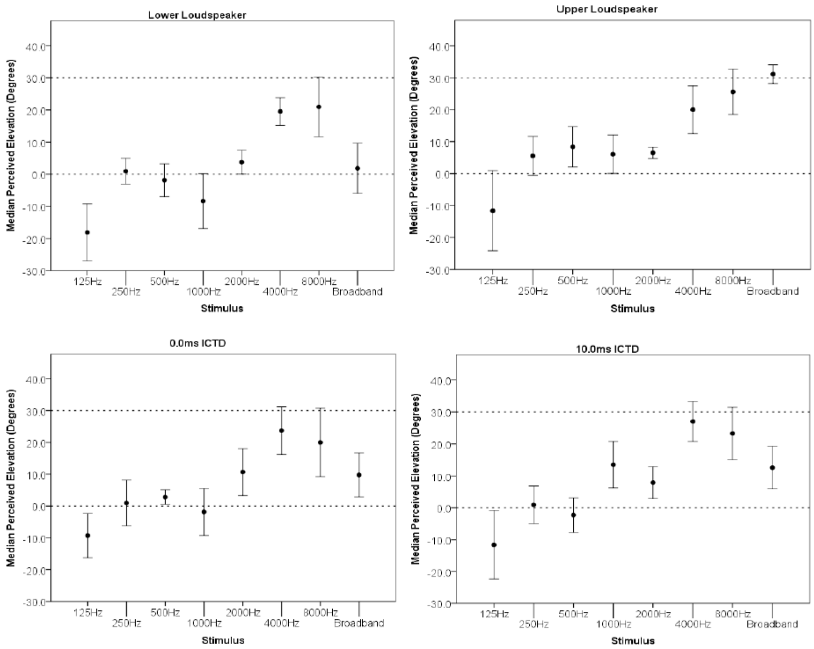


Figure 9.5.11: Pitch Height Effect (Lee *et al.*, 2014, p. 3)

9.6 Recording Methodology

A detailed justification of the chosen arrays including microphone choice and placement, detailed indications, measurements and calculations can be found in the Appendix.

9.6.1 Choice of Sources

As stated by Rumsey, it was considered important to record a variety of sources each differing with regards to frequency content and its acoustic envelope. The choice of cello, violin, djembe, handpan and guitar is similar to the choice of other studies in the field of psychoacoustics including height channels, such as Lee & Gribben (2013), Wallis & Lee (2016a), Wallis & Lee (2017), or Lee (2011), amongst others. Further, the repertoire played by the guitarist made it possible to include material containing extended techniques, as in Simurra & Queiroz (2017), to have a broader range of source material. The choice of solo sources was made to ease the perception of possible differences between the different source types.

9.6.2 Preliminary Recording Session

Prior to the recording session, a test session was organised in Bankstock Studios to ensure that the chosen angle between the main and height layer of the Main-Array would achieve at least an ICLD of 9.5dB (localisation threshold) when playing back white noise through a Marshall Code 25 amplifier, as its frequency content represents all sources mentioned above and its sustained nature would allow to visually read the obtained ICLD between the two layers. A detailed explanation of the importance of achieving the localisation threshold can be found in Chapter 9.5.5. Although the acoustic properties of the recording space are different, the acoustically controlled environment would at least give indications whether such an ICLD could be achieved. The result of the test session (Figures 9.6.2 and 9.6.3) suggested that an ICLD of 9.5dB would be realistic, as depicted in Figure 9.6.1.

Figure 9.6.1: ICLDs in Test Session


Figures 9.6.2 and 9.6.3: Test Setup Main-Array

9.6.3 Recording Procedure

The recording took place at All Saints East Finchley, London, a church being well-known amongst engineers for its acoustics (Faulkner, 2019a). For the recording, all microphone signals were routed to three RME Octa Mic II preamps, an Antelope Orion 32 interface, and recorded into Pro Tools in PCM wave format at 96kHz/24bit resolution.

To allow for consistency throughout the comparison of the different techniques and especially for experimenting with a hybrid approach the goal was to organise modular microphones having all the same preamps but different capsules to obtain the desired polar patterns. As Schoeps microphones can be considered as an industry standard and have been applied in many of the quoted studies, such as in Howie *et al.* (2016 & 2017), amongst others, the choice fell on the Schoeps Colette series with the CMC 6 preamp. Apart from that, it seemed to be the only modular system available from the rental companies in question enabling to cover all polar patterns needed for the comparison and at the same time providing the option of adding diffraction attachments. For further justifications regarding the microphone choice the reader is referred to the Appendix.

As having used the same mic model for all channels, matching the microphone sensitivity was considered achieved apart from slight sensitivity differences caused by the different capsules for obtaining the different polar patterns, which are an experimental constant.

During the recording the goal was to match the input gain on all channels to minimise the differences of a possible colouration of the signal through the preamp amongst channels and to maintain the natural level relationships between channels. However, when strictly adhering to this the S/R ratio of the Hamasaki-Cube would exceed acceptable limits. Therefore, the gain applied to the channels of the Hamasaki-Cube was matched within the low and height layer but was higher than the gain of other channels.

Due to the sudden appearance of interference, few channels did contain radio signals above a certain preamp gain level. Considering the rather tight time constraints, the only efficient way to handle this problem was to reduce the gain of the affected channels. This difference in gain, was later compensated during stimuli creation.

9.6.4 Floor Plans

Figures 9.6.4 and 9.6.5 show the setup of the different arrays. Please note that the AB stereo pair appearing in the input list used to derive a separate stereo recording is not shown in the floor plans, as it was not part of the current investigation.

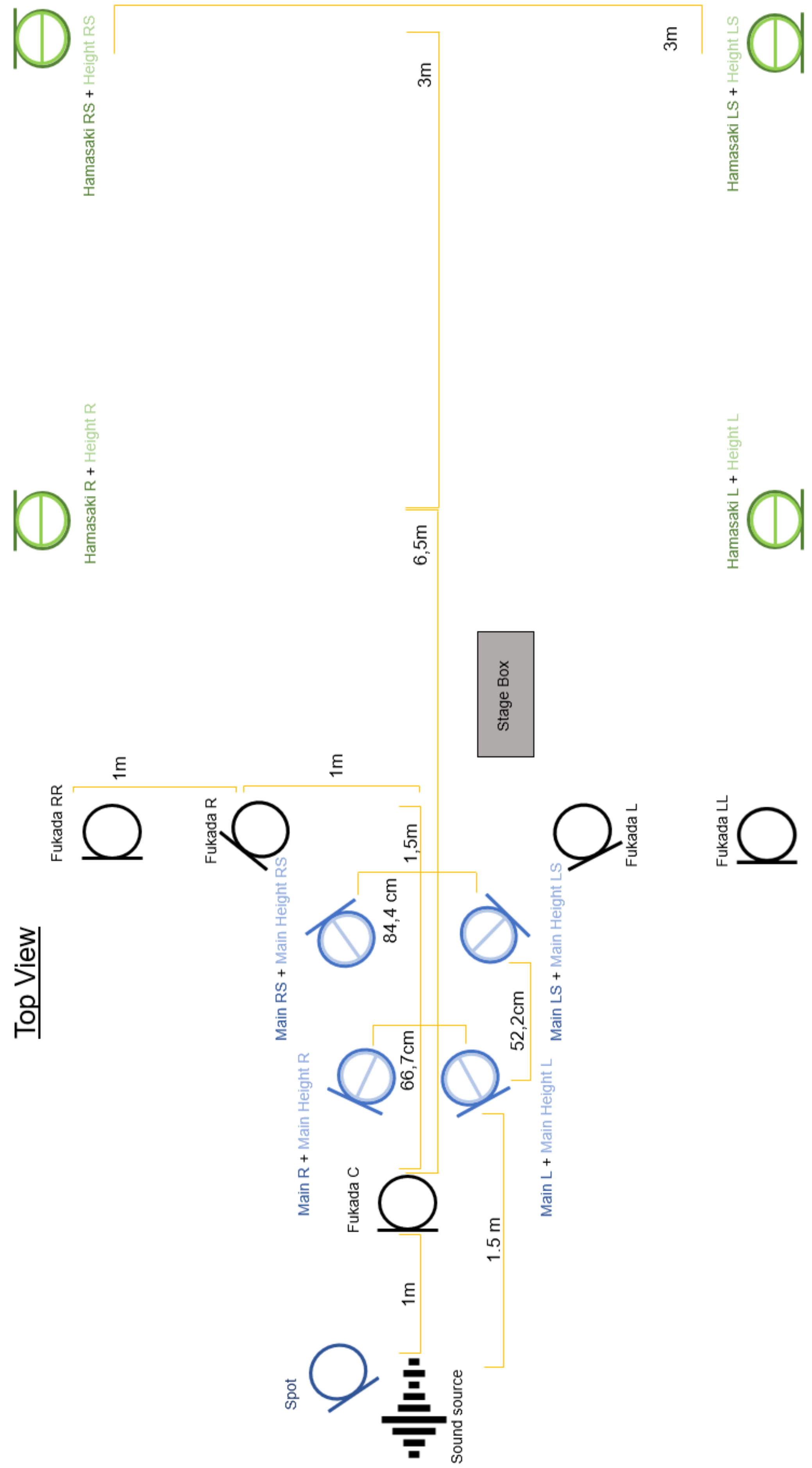


Figure 9.6.4: Floor Plan Top View

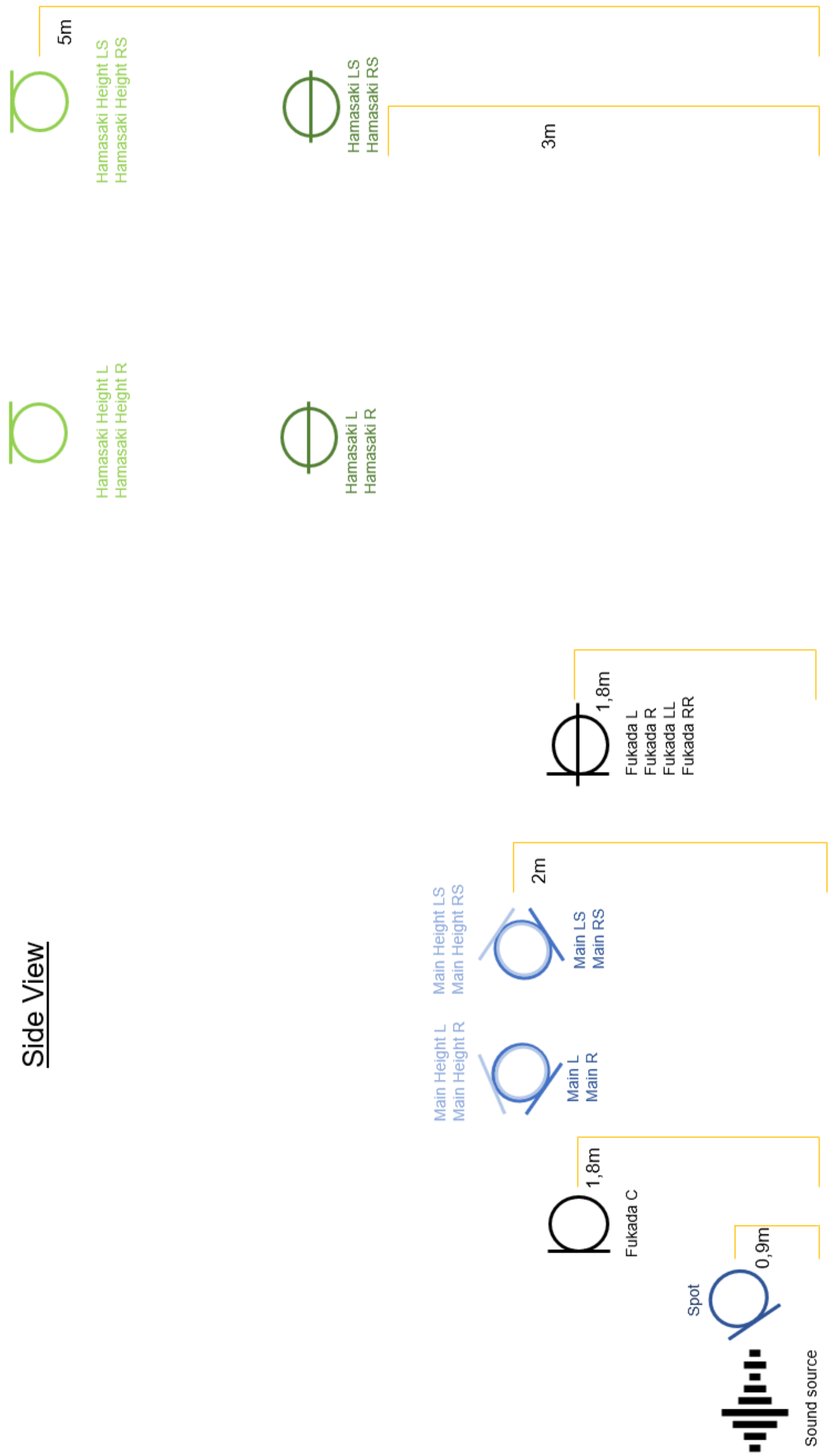


Figure 9.6.5: Floor Plan Side View

9.6.5 Input List

The colours used in the input list (Table 2) are correspondent to the floor plans (Figures 9.6.4 & 9.6.5) to ease identification.

Input RME	Position	Capsule on Schoeps CMC 6	Polar Pattern	Axis	Height
1	Main L	MK2H	Omni	30° outwards, 45° to source	2m
2	Main R	MK2H	Omni	30° outwards, 45° to source	2m
3	Main LS	MK2H with diffr. attachment	Omni	30° outwards, 45° to floor	2m
4	Main RS	MK2H with diffr. attachment	Omni	30° outwards, 45° to floor	2m
5	Main Height L	MK4	Cardioid	110° from main upwards	2m
6	Main Height R	MK4	Cardioid	110° from main upwards	2m
7	Main Height LS	MK4	Cardioid	110° from main upwards	2m
8	Main Height RS	MK4	Cardioid	110° from main upwards	2m
9	Fukada L	MK4	Cardioid	45° outwards	1,8m
10	Fukada C	MK4	Cardioid	0°	1,8m
11	Fukada R	MK4	Cardioid	45° outwards	1,8m
12	Fukada LL	MK2	Omni	0°	1,8m
13	Fukada RR	MK2	Omni	0°	1,8m
14	Hamasaki L	MK8	Figure-8	Positive lobe 90° outwards	3m
15	Hamasaki R	MK8	Figure-8	Positive lobe 90° outwards	3m
16	Hamasaki LS	MK8	Figure-8	Positive lobe 90° outwards	3m
17	Hamasaki RS	MK8	Figure-8	Positive lobe 90° outwards	3m
18	Hamasaki Height L	MK41	Supercardioid	0° to ceiling	5m
19	Hamasaki Height R	MK41	Supercardioid	0° to ceiling	5m
20	Hamasaki Height LS	MK41	Supercardioid	0° to ceiling	5m
21	Hamasaki Height RS	MK41	Supercardioid	0° to ceiling	5m
22	Spot	MK4	Cardioid		0,9m
23	AB L	MK2H	Omni	30° outwards, 45° to source	2m
24	AB R	MK2H	Omni	30° outwards, 45° to source	2m

Table 9.6.6: Input List

9.6.6 Photos of the Recording Session

Figure 9.6.7: Recording Setup Overview from Top Centre Close

Figure 9.6.8: Recording Setup Overview from Top Left

Figure 9.6.9: Recording Setup Overview from Top Right

Figure 9.6.10: Recording Setup Overview from Top Centre Far

Figure 9.6.11: Recording Setup Side View

Figure 9.6.12: Recording Setup Front View

Figure 9.6.13: Recording Setup Back View


Figure 9.6.14: Position Marking of the Main- and Front Array


Figure 9.6.15: Position Marking of the Hamasaki-Cube

9.7 Objective Measures

Objective measurements have been taken to fall back on when trying to explain the obtained listening test results. In order to gain further insight into the acoustics of the recording space in case needed during the evaluation of the results, RT60 and IRs have been derived from the recording space. In addition, dummy head recordings have been conducted to gather an objective reference for comparison to the timbral qualities perceived in the auditory evaluation.

9.7.1 Dummy Head Recording of the Reproduced Mixes

The placement of the dummy head was the same as when matching the stimuli levels (listening position and ear height, see Figure 3.4.9, Chapter 3.4). The input gain of the RME Octa Mic Preamp was matched and tested using pink noise playback from the centre speaker. The same procedure was applied by Lee & Gribben (2014, p. 877) and Howe *et al.* (2018, p. 5) for obtaining an objective reference when analysing the results of the subjective listening test.

9.7.2 Impulse Responses

Impulse responses of bursting balloons located at the position of the sound sources were taken at the position of the two arrays with a Behringer ECM 8000 measurement microphone, as can be seen in Figures 9.7.1 and 9.7.2:

Figure 9.7.1: Impulse Response Recording of Balloons on the Position of the Front Arrays

Figure 9.7.2: Impulse Response Recording of Balloons on the Position of the Back Arrays

9.7.3 RT60 Measurement

RT60 was measured with a Phonic PAA3 audio analyser under the same conditions as the IRs. The RT60 at the front position was measured 1.41seconds and the RT60 at the back 1.5 seconds.