Transitioning Studio Practice from Stereo to 3D: Single Instrument Capture with a Focus on the Vertical Image

Bryan Martin

Department of Music Research

Schulich School of Music

McGill University, Montreal

Submitted August 2019

A thesis submitted to McGill University in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

Abstract

This dissertation investigates new methodologies for the creation of three-dimensional audio images of individual musical instruments in recorded music. The recordings are made for playback in environments equipped with surround loudspeaker arrays that include loudspeakers located above and below the listener. These techniques are aimed at creating natural sounding representations of individual instruments in multi-channel recording and mixing of classical music, pop and jazz, but can also be used for any instance where a three-dimensional capture of audio is desired. To accomplish this, a suitable adjustability of three-dimensional spatial presence of music sources is needed. This capacity could be used to bring out individual artistic interpretations of naturalness of sound based on preferences of engineers and sound designers. The development of these methods evolved from prior research in stereo and multi-channel music recording techniques, spatial hearing and perception, psychoacoustics and from the evaluation of experiments in music mixing and the recording of individual musical instruments. Initial experiments investigated methods for expanding the size of sonic objects projected onto two and three dimensions of horizontal and vertical loudspeaker planes. The goal was to develop the best practices for the distribution of the audio spectrum in hemispherical multi-loudspeaker playback systems. Following that effort, techniques were developed for the creation of immersive mixes of popular music from monophonic multi-track source material in a multi-channel playback environment featuring several layers of loudspeakers. One of the most daunting challenges in creating believable three-dimensional mixes is expanding mono/point-source audio sources into a three-dimensional audio image. To achieve this aim, discrete microphone arrays were developed to capture the image of a sound source into a coherent vertical representation of the original. Ambisonic and HOA tools were not considered as they were not common in the development of

studio recording practices, do not relate to the transition from stereo, and are mostly unfamiliar to recording studio personnel.

Listening tests were conducted to validate the effectiveness of three-dimensional capture to determine the importance of spatial attributes in the vertical image of sound, and to qualify the spatial characteristics and realism of three-dimensional reproduction. A graphical method was employed to represent the perceived spatial attributes of the microphone arrays designed to create three-dimensional audio images when played back in a multi-channel listening environment featuring several layers of loudspeakers.

Résumé

Cette thèse examine de nouvelles méthodologies pour la création d'images audio en trois dimensions d'enregistrements de divers instruments de musique. Les images audio ainsi créées sont destinées à la lecture dans des environnements équipés de réseaux de haut-parleurs surround, situés au-dessus et au-dessous de l'auditeur. Ces techniques visent à créer des représentations sonores naturelles des divers instruments lors de l'enregistrement et du mixage à canaux multiples de musique classique, de pop et de jazz. Elles peuvent également être utilisées lorsqu'une capture audio en trois dimensions est souhaitée. Pour ce faire, il est nécessaire d'ajuster de manière appropriée dans l'espace tridimensionnel les diverses sources de musique. Ces techniques d'enregistrement pourraient être utilisées pour faire ressortir des interprétations artistiques individuelles en fonction des préférences des ingénieurs et des concepteurs sonores. Le développement de ces méthodes a été élaboré à partir de recherches antérieures sur les techniques d'enregistrement musical stéréo et multicanaux, l'audition et la perception spatiales, la psycho-acoustique, ainsi que de l'évaluation d'expériences de mixage musical et de l'enregistrement de divers instruments de musique. Les premières expériences ont porté sur les méthodes permettant d'augmenter la taille des objets soniques projetés sur deux ou trois dimensions des plans horizontaux et verticaux de haut-parleurs. L'objectif était de développer les meilleures pratiques pour la distribution du spectre audio dans un environnement de lecture hémisphérique. Suite à cet effort, des techniques ont été développées pour créer des mixages immersifs de musique pop, réalisés à partir de sources monophoniques multicanauxs, dans un environnement de lecture à canaux multiples comportant plusieurs plans de haut-parleurs. L'un des défis les plus difficiles à relever pour créer des mixages tridimensionnels crédibles consiste à transformer des sources audio monophoniques en images audio tridimensionnelles. Pour atteindre cet objectif, des matrices de

microphones ont été configurées pour capturer l'image d'une source sonore dans une représentation verticale, cohérente avec la source originale. Des tests d'écoute ont été menés pour valider l'efficacité de la capture en trois dimensions afin de déterminer l'importance des attributs spatiaux dans l'image verticale du son, ainsi que pour qualifier les caractéristiques spatiales et le réalisme de la reproduction en trois dimensions. Une méthode graphique a été utilisée pour représenter les attributs spatiaux perçus à partir des configurations de microphones, conçues pour créer des images audio tridimensionnelles lors de la lecture dans un environnement d’écoute à canaux multiples comportant plusieurs plans de haut-parleurs.

Acknowledgements

None of this would have happened without the support and love of my wife Johanne Champagne, and extensive comic book exploits with my son Etienne.

I would like to convey my sincere appreciation to my co-supervisors, Prof. Wieslaw Woszczyk and Prof. Richard King for their guidance, support and indulgence throughout this process.

A huge heaping of gratitude and good times to my partner in 3D adventures Will Howie, and the continuum of support and therapy from the Sound Recording PhD family: Diamond Dave Benson (the mother), Jack Kelly, Denis Martin, Diego Quiroz Orozco, Dr. Brett Leonard, Dr. Doyuen Ko, and Dr. Sungyoung Kim (the brothers).

To an unfinished conversation with Eleanor Stubley who believed (and who still owes me that drink).

A huge thanks for technical support and musings on cheap guitar trading to Ieronim Catanescu, and to CIRMMT technical support and allowing me the use of absurd caches of audio equipment: Yves Méthot and Julien Boissinot.

And lastly the boys in the band, who were here before and will be here forever after: Goran Petrovic (on going deep and finding the root) and Michael Greenfield (Monday, Thursday café, shop therapy, and the source of ‘tone’), there are no words.

This research would not have been possible without extraordinary financial support from the Fonds de recherche Société et culture Québec (FQRSC), as well as generous support from McGill University, the Schulich School of Music, and the Centre for Interdisciplinary Research in Music Media and Technology (CIRMMT).

Preface

This dissertation is presented as a manuscript thesis, wherein the bulk of the document is taken from previously published papers. For Chapters 3 through 6, the text appears generally as originally published, apart from changes to formatting, figures, tables, and reference numbers. Some text and references have been updated or eliminated to reflect the grouping of these manuscripts into a single document. Also, some discussion sections have been expanded to include information omitted from the original published versions owing to document length limitations. Written consent to reproduce previously published material in these manuscripts has been granted by my various co-authors, as well as the Managing Editor of the Journal of the Audio Engineering Society. All figures within this manuscript fall under the sole property and ownership of the author, Bryan Martin, unless otherwise noted.

Original Scholarship and Distinct Contributions to Knowledge

The study in **Chapter 3.2** is an investigation into mixing techniques for the creation of three-dimensional audio images of individual instruments in popular music reproduced in the 22.2 Multi-channel Sound System [1]. Results showed that technical requirements to achieve three-dimensional believability and immersion are far greater than that required for conventional stereo presentation, and the use of conventional mono and stereo tools necessitates a large time investment and tedious workflow that must be performed on each musical element of conventional mono and stereo source material.

Chapter 3.3 is an investigation into the playback levels of height-channel information that are considered to be effective in music mixing. The findings were that there is a minimum level of height information below which subjects could not differentiate added height content; that the subjects could perceive three distinct content levels during testing; and that the level of the immersive content needs to be substantially louder than that from loudspeakers in the horizontal plane to be perceived. Lastly, a preference question suggests that subjects preferred a more immersive environment than the more-subtle levels of immersion when given the choice.

Chapter 3.4 is an investigation into techniques and strategies for the creation of compelling high-fidelity mixes of popular music in the 22.2 Multi-channel Sound System. Results from the evaluation of *3-Dimensionality* and *Immersion* suggest that it is possible to create believable three-dimensional immersion using conventional stereo spatial tools in a 22.2 playback environment from monophonic multi-track source material. The workflow, however, is inefficient and time-intensive, and may not be commercially viable. Results of an expanded mixing 'sweet spot' in immersive mixing environments were inconclusive and may be program- and mix-method dependent. One of the primary obstacles that hindered the examination of the effectiveness of the

created virtual acoustic is the high number of audio channels required for creating a believable 'virtual reality'. An observation that surprised the author after decades of commercial music production experience was the ineffectiveness of many proven techniques employed in commercial two-channel stereo delivery.

The study in **Chapter 4** compares a selection of microphone arrays for music recording to create vertical and three-dimensional images. The three-dimensional images are generated by routing direct signals from the microphones to discrete loudspeaker channels. Most published investigations addressing three-dimensional microphone arrays have focused on the capture of ensembles in live situations. These techniques prioritize the direct sound in the middle loudspeaker layer, ambience in the height layer and employ no bottom layer loudspeakers. This study focuses on single instrument capture in recording studio situations. Results showed that subjects discerned the different arrays from each other, and that a listener can easily identify each array. The majority of the subjects rated the array images having an extended verticality over the mono source, and that the 3D microphone arrays provide a greater depth of image than the mono capture. Although the subjects had not been asked to examine the arrays for three-dimensional criteria prior to testing, their combined observations point to good performance by the arrays in capturing vertical imaging and enhanced dimensionality when played through loudspeaker systems that include height channels.

The investigation detailed in **Chapter 5** utilizes a simple graphical method in an effort to represent the perceived spatial attributes of the microphone arrays and a mono/point-source signal. Direct sound from the individual microphones within each array is assigned to the specific loudspeakers of the 22.2 playback system, which most closely mirrors their position during capture. The subjects' representations support that these arrays clearly capture much more

information than a single microphone. In examining the images provided by these arrays in the context of immersive/3D mixing and post-production, a case can be made that they will contribute to a more efficient and improved workflow.

Chapter 6 employs a 3D panning tool to manipulate audio images recorded with advanced close-capture microphone arrays for three-dimensional imaging. The 3D microphone arrays used in this study were: Coincident-XYZ, M/S-XYZ and Non-coincident-XYZ/five-point (See 4.3.1 for detailed description). Instruments of the orchestral string, woodwind, and brass sections were recorded. The objective of the test was to determine the point of three-dimensional expansion onset, preferred imaging, and image breakdown point. Subjects were presented with a continuous dial to manipulate the three-dimensional spread of the arrays, allowing them to expand or contract the microphone signals from 0˚ to 90˚ azimuth/elevation. The results showed that the M/S-XYZ array is the perceptually "biggest" of the capture systems under test and displayed the fasted sense of *expansion onset*. The coincident and non-coincident arrays are much less agreed upon by subjects in terms of *preference* in particular, and also in *expansion onset*.

Contributions of Authors

For all previously published work presented in this thesis (Chapters 3-7) I was the principal author, and was responsible for all background research, development of the research questions, design and implementation of new recording techniques, design and administration of listening tests, and interpretation of the results. Listed below are the contributions of my various co-authors.

Chapter 3.2

Martin, Bryan, "Mixing Popular Music in Three Dimensions," presented at the Innovation in Music 2015, Cambridge, UK, 2015.

Chapter 3.3

Martin, Bryan; King, Richard; Leonard, Brett; Benson, David; Howie, Will, "Immersive Content in Three Dimensional Recording Techniques for Single Instruments in Popular Music," presented at the 138 AES Convention, Warsaw, Poland, 2015.

Richard King and Will Howie aided in framing the research question and interpreting the results. Brett Leonard created the Max/MSP patch used to administer the listening test. Dave Benson and Brett Leonard conducted the statistical analysis of the listener data and aided in interpreting the results.

Chapter 3.4

Martin, Bryan; King, Richard, "Three Dimensional Spatial Techniques in 22.2 Multi-channel Surround Sound for Popular Music Mixing," presented at the 139 AES Convention, New York, NY, USA, 2015.

Richard King aided in framing the research question and interpreting the results.

Chapter 4

Martin, Bryan; King, Richard; Woszczyk, Wieslaw, "Microphone Arrays for Three-Dimensional Capture of Acoustic Instruments," presented at the 2016 AES International Conference on Sound Field Control, Guilford, UK, 2016.

Richard King and Wieslaw Woszczyk aided in framing the research question. Richard King aided in interpreting the results.

Chapter 5

Martin, Bryan; King, Richard; Woszczyk, Wieslaw, "Subjective Graphical Representation of Microphone Arrays for Vertical Imaging and Three-Dimensional Capture of Acoustic Instruments, Part I," presented at the AES 141 Convention, Los Angeles, CA, 2016.

Richard King and Wieslaw Woszczyk aided in framing the research question. Richard King and Denis Martin aided in interpreting the results.

Martin, Bryan; King, Richard; Woszczyk, Wieslaw, "Subjective Graphical Representation of Microphone Arrays for Vertical Imaging and Three-Dimensional Capture of Acoustic Instruments, Part I," presented at the AES 141 Convention, Los Angeles, CA, 2016.

Richard King and Wieslaw Woszczyk aided in framing the research question. Richard King and Denis Martin aided in interpreting the results.

Denis Martin conducted the statistical analysis of the listener data and aided in interpreting the results.

Chapter 6

Jack Kelly created the Max/MSP patch used to administer the listening test. Brett Leonard conducted the statistical analysis of the listener data. Brett Leonard and Jack Kelly and aided in interpreting the results.

Contents

ABSTRACT ..... **I**

RÉSUMÉ ..... **III**

ACKNOWLEDGEMENTS ..... **V**

PREFACE ..... **VI**

ORIGINAL SCHOLARSHIP AND DISTINCT CONTRIBUTIONS TO KNOWLEDGE ..... **VII**

CONTRIBUTIONS OF AUTHORS ..... **X**

CONTENTS ..... **XII**

LIST OF TABLES ..... **XX**

LIST OF FIGURES ..... **XXI**

1 INTRODUCTION **1**

1.1 MOTIVATION 1

1.2 RESEARCH GOALS 3

1.3 STRUCTURE OF THESIS 4

2 BACKGROUND **7**

2.1 BACKGROUND OF RECORDED MEDIA AND REPRODUCTION 7

2.2 MULTI-CHANNEL AUDIO 8

2.2.1 Stereophonic Sound *8*

2.2.2 Quadraphonic Sound *9*

2.2.3 Brian Eno: "An Ambient Speaker System" *11*

2.2.4 Channel-based Surround Sound *12*

2.2.5 5.1 Surround Sound *14*

2.2.6 7.1 Surround Sound *16*

2.2.7 Multi-channel Formats that include Height Channels *16*

2.3 SPATIAL HEARING 24

2.3.1 Localization and Localization Error *25*

2.3.2 Perception in the Horizontal Plane *26*

2.3.3 Envelope Cues *28*

2.3.4 Median Plane Localization and Monaural Cues *30*

2.4 STEREO RECORDING 33

2.4.1 Useful Angle of Acceptance *34*

2.4.2 Coincident Microphone Techniques *36*

2.4.3 Near-Coincident Microphone Techniques *40*

2.4.4 Spaced Microphone Techniques *42*

2.5 3D AUDIO RECORDING TECHNIQUES 42

2.5.1 William Howie's Orchestral Recording Technique Optimized for the 22.2 Multi-channel Sound System. 43

2.5.2 Michael William's Arrays for Rendering 3D Sound Fields *44*

2.5.3 Günther Theile's OCT-9 Array *45*

2.5.4 Morten Lindberg's 2L-Cube Array *46*

2.5.5 Gregor Zielinsky's Twins Cube and Twins Square Arrays *47*

2.5.6 Helmut Wittek and Günter Theile's ORTF-3D *48*

2.5.7 David Bowles' Microphone Array *49*

2.5.8 Paul Geluso's Z-Microphone Array *50*

2.5.9 NHK Portable Spherical Microphone Array for Super Hi-Vision 22.2 Multi-channel Audio *51*

2.5.10 Richard King's Omni-directional Height Array with Diffraction Attachments *52*

2.5.11 Wieslaw Woszczyk's and Paul Geluso's 3D Sound Field Array *52*

2.6 SUBJECTIVE ASSESSMENT OF SPATIAL SOUND QUALITY 53

2.6.1 Spatial Audio Assessment *54*

3 EXPERIMENTS IN IMMERSIVE MIXING 57

3.1 SCHULICH SCHOOL OF MUSIC STUDIO 22 57

3.1.1 Music Studio 22 Properties 58

3.1.2 Room Shape and Dimensions 60

3.1.3 Background Noise 61

3.1.4 Loudspeaker Geometry and Configuration 61

3.1.5 Loudspeaker System Properties 62

3.1.6 The Measured Operational In-Room Loudspeaker Response [155] 64

3.2 MIXING POPULAR MUSIC IN THREE DIMENSIONS: EXPANSION OF THE KICK DRUM SOURCE IMAGE 65

3.2.1 Abstract 65

3.2.2 Introduction 65

3.2.3 Goal of Three-Dimensional Mix Investigation 66

3.2.4 Test Environment 66

3.2.5 Experimental Design 66

3.2.6 Techniques Used to Expand the Kick Drum Image 68

3.2.7 Conclusions 74

3.2.8 Future Work 75

3.3 IMMERSIVE CONTENT IN THREE DIMENSIONAL RECORDING TECHNIQUES FOR SINGLE INSTRUMENTS IN POPULAR MUSIC .. 76

3.3.1 Abstract 76

3.3.2 Introduction 76

3.3.3 Test Design 77

3.3.4 Testing Software & Methodology 82

3.3.5 Subjects 84

3.3.6 Results and Analysis 84

3.3.7 Conclusions 87

3.3.8 Possibilities for Future Work 88

3.4 THREE DIMENSIONAL SPATIAL TECHNIQUES IN 22.2 MULTI-CHANNEL SURROUND SOUND FOR POPULAR MUSIC MIXING. 89

3.4.1 Abstract 89

3.4.2 Experimental Design 89

3.4.3 Test Environment 90

3.4.4 Construction of the Virtual Acoustic 91

3.4.5 Placement of Discrete Delays 92

3.4.6 Placement of Discrete Reverberations 93

3.4.7 Multi-Channel Impulse Response Layer 96

3.4.8 Mix Evaluation 97

3.4.9 Test Subjects 97

3.4.10 Results and Analysis 98

3.4.11 Conclusions 99

3.4.12 Future Work 101

4 MICROPHONE ARRAYS FOR VERTICAL IMAGING AND THREE-DIMENSIONAL CAPTURE OF ACOUSTIC INSTRUMENTS 103

4.1 ABSTRACT 103

4.2 INTRODUCTION 104

4.3 TEST DESIGN AND METHODS 105

4.3.1 Microphone Arrays 105

4.3.2 Recording Studio 111

4.3.3 Control Room Monitor Environment 111

4.4 ARRAY LEVEL MATCHING 111

4.4.1 Intra-Array Level Matching 111

4.4.2 Inter-Array Level Matching 112

4.5 TEST SUBJECTS 112

4.6 TEST MATERIAL 114

4.7 Test Procedure 115
4.7.1 Playback of Microphone Arrays for Testing 115
4.7.2 Test Familiarization and Training 115
4.8 Results 116
4.8.1 Results of Training Tests 116
4.8.2 Results of Main Test 117
4.8.3 Questionnaire for Qualitative Observations 119
4.8.4 Subject Rating of the Training and Main Test 119
4.8.5 Subjects Qualitative Observation of Arrays 120
4.9 Discussion of Results 123
4.9.1 Training Test 123
4.9.2 Main Test 123
4.9.3 Perceived Verticality of Audio Image Captured by Arrays Compared to Mono 124
4.9.4 Perceived Depth of Audio Image Captured by Arrays Compared to Mono 124
4.9.5 Perceived Immersion of Audio Image Captured by Arrays Compared to Mono 124
4.9.6 Perceived Size of Audio Image Captured by Arrays Compared to Mono 124
4.10 Conclusion 125
4.11 Future Work 125
5 SUBJECTIVE GRAPHICAL REPRESENTATION OF MICROPHONE ARRAYS FOR VERTICAL IMAGING AND THREE-DIMENSIONAL CAPTURE OF ACOUSTIC INSTRUMENTS 126
5.1 Abstract 126
5.2 Introduction 127
5.3 Test Design and Methods 128
5.4 Array Level Matching 128
5.5 Test Subjects 128
5.6 Test Material 128

5.7 Test Procedure 129
5.7.1 Playback of Microphone Arrays for Testing 129
5.7.2 Test Familiarization and Training 129
5.8 Results 131
5.8.1 Graphical Results of Test 131
5.8.2 Assessment of Test by Subjects 132
5.9 Discussion and Analysis of Results 133
5.9.1 Discussion of Maximum Extents 133
5.9.2 Discussion of Graphical Representations 134
5.9.3 Statistical Analysis 135
5.10 Conclusion 141
5.11 Future Work 142
5.12 Drawings of Audio Images 143
5.12.1 Mono Drawings 143
5.12.2 Coincident Array Drawings 145
5.12.3 Non-Coincident Array Drawings 147
5.12.4 MS-XYZ Drawings 149
5.12.5 Accumulated Array Drawings 151

6 SUBJECTIVE ASSESSMENT OF THE VERSATILITY OF THREE-DIMENSIONAL NEAR-FIELD MICROPHONE ARRAYS FOR VERTICAL AND THREE-DIMENSIONAL IMAGING 153

6.1 Abstract 153
6.2 Introduction 154
6.2.1 Immersive Audio Recording and Production 154
6.2.2 Stereo vs 3D Reproduction 155
6.2.3 A More Complete Sonic Picture 155
6.3 Test Design and Methods 156

6.3.1 Testing Environment 156

6.3.2 Microphone Arrays 156

6.3.3 Array Level Matching 156

6.3.4 Test Subjects 156

6.3.5 Test Material 157

6.3.6 Test Procedure 157

6.3.7 Training and Familiarization 162

6.3.8 Testing 162

6.3.9 Post Testing Questionnaire 164

6.4 RESULTS 164

6.4.1 Nuisance variables 164

6.4.2 The Effect of Array 166

6.4.3 Image Range 168

6.4.4 The Effect of Experience 173

6.5 DISCUSSION 173

6.6 CONCLUSIONS 174

6.7 FUTURE WORK 175

7 CONCLUSIONS 176

7.1 SPECIFIC CONCLUSIONS BY CHAPTER 176

7.1.1 Chapter 3.2 176

7.1.2 Chapter 3.3 177

7.1.3 Chapter 3.4 178

7.1.4 Chapter 4 179

7.1.5 Chapter 5 180

7.1.6 Chapter 6 181

8 FUTURE WORK 183

8.1 Research Listening Environments 184

8.1.1 A Proposed 27.2 Configuration for Maximum Immersion 184

8.2 Investigation Into the Benefits of Bottom Layer Loudspeakers 186

8.3 Continued Research in Vertical Audio Imaging 186

8.4 More Uniform Test Methods 186

9 GLOSSARY OF TERMS 188

10 BIBLIOGRAPHY 193

List of Tables

TABLE 1: STUDIO 22 LOUDSPEAKER LAYOUT AND NAMING CONVENTIONS. 60

TABLE 2: STUDIO 22 DIMENSIONS 60

TABLE 3: TRACK LISTING FOR 1980S MULTI-TRACK USED FOR EXPERIMENT. 67

TABLE 4: MICROPHONES USED IN RECORDING. 78

TABLE 5: PARAMETERS OF DISCRETE DELAYS. 94

TABLE 6: PARAMETERS OF STEREO REVERBERATIONS. 94

TABLE 7: MICROPHONES USED IN RECORDING. 106

TABLE 8: TRAINING TEST TRIALS. 117

TABLE 9: MAIN TEST TRIALS. 118

TABLE 10: DRAWING TRIALS 131

TABLE 11: MICROPHONE ARRAYS AND CORRESPONDING PANNING OBJECT NUMBERS. (SEE FIGURES 118-121). 158

TABLE 12: LOUDSPEAKER CONFIGURATION USED FOR TESTING 159

TABLE 13: MEAN AND VARIANCE VALUES FOR PREFERENCE BY ARRAY, SUMMED ACROSS ALL SUBJECTS. 167

TABLE 14: ARRAY EXPANSION ONSET: AZIMUTH AND ELEVATION. 169

TABLE 15: ARRAY IMAGE BREAKDOWN: AZIMUTH AND ELEVATION. 169

List of Figures

FIGURE 1: RIAA STANDARD RECORDING AND REPRODUCING CHARACTERISTIC. 9

FIGURE 2: QUADRAPHONIC LOUDSPEAKER ARRANGEMENT. 10

FIGURE 3: BRIAN ENO'S AMBIENT SPEAKER SYSTEM. 11

FIGURE 4: FOUR-CHANNEL SURROUND SOUND FOR MAGNETIC FILM.[56], (USED BY PERMISSION).14

FIGURE 5: 70MM SIX-CHANNEL SURROUND SOUND FOR MAGNETIC FILM.[56], (USED BY PERMISSION). 14

FIGURE 6: ITU-R BS.775-1 5.1 REFERENCE LOUDSPEAKER ARRANGEMENT. 15

FIGURE 7: ITU-R BS.775-2 7.1 REFERENCE LOUDSPEAKER ARRANGEMENT. 16

FIGURE 8: 5.0 AND 7.0 WITH 2 HEIGHT CHANNELS 21

FIGURE 9: 10.2 TYPE A SOUND SYSTEM WITH 2 HEIGHT CHANNELS. 21

FIGURE 10: 10.2 TYPE B SOUND SYSTEM WITH 3 HEIGHT CHANNELS. 22

FIGURE 11: 9.1 AURO-3D SOUND SYSTEM WITH 4 HEIGHT CHANNELS. 22

FIGURE 12: 11.1 AURO-3D SOUND SYSTEM WITH 6 HEIGHT CHANNELS. 23

FIGURE 13: 11.1 (7.1+4) SOUND SYSTEM WITH 4 HEIGHT CHANNELS. 23

FIGURE 14: 22.2 SOUND SYSTEM WITH 9 HEIGHT CHANNELS. 24

FIGURE 15: LOCALIZATION ERROR. 26

FIGURE 16: INTERAURAL TIME DELAY (ITD). 27

FIGURE 17: MAXIMUM INTERAURAL TIME DELAY (ITD). 28

FIGURE 18: TRANSIENT ENVELOPE CUE DIAGRAM FROM [42]. 29

FIGURE 19: CONE OF CONFUSION. 30

FIGURE 20: LOCALIZATION BLUR IN THE MEDIAN PLANE FOR CONTINUOUS SPEECH BY A FAMILIAR PERSON [73, 97] 32

FIGURE 21: STEREO PLAYBACK IMAGING OF SOUND SOURCES LOCATED WITHIN THE ANGLE OF ACCEPTANCE. 35
FIGURE 22: STEREO PLAYBACK IMAGING OF SOUND SOURCES LOCATED OUTSIDE THE ANGLE OF ACCEPTANCE. 35
FIGURE 23: XY COINCIDENT MICROPHONE PAIR. 37
FIGURE 24: BLUMLEIN ARRAY. 38
FIGURE 25: MS STEREO BASIC MICROPHONE CONFIGURATION. 39
FIGURE 26: MS MATRIX CIRCUITS. 39
FIGURE 27: MS MATRIX CIRCUIT AT RECORDING CONSOLE. 40
FIGURE 28: COMPARISON OF NEAR-COINCIDENT MICROPHONE ARRAYS. 41
FIGURE 29: WILLIAM HOWIE'S RECORDING TECHNIQUE OPTIMIZED FOR 22.2 MULTI-CHANNEL SOUND [111], (USED BY PERMISSION). 43
FIGURE 30: WILLIAMS' 12-CHANNEL M.A.G.I.C. ARRAY [66]. (USED BY PERMISSION). 44
FIGURE 31: GÜNTHER THEILE OCT-9 ARRAY [66].(USED BY PERMISSION). 45
FIGURE 32: MORTEN LINDBERG'S 2L-CUBE [66]. (USED BY PERMISSION). 47
FIGURE 33: GREGOR ZIELINSKY'S TWIN CUBE ARRAY [66]. (USED BY PERMISSION). 48
FIGURE 34: ORTF-3D. (USED BY PERMISSION). 49
FIGURE 35: BOWLES MICROPHONE ARRAY [66]. (USED BY PERMISSION). 50
FIGURE 36: PAUL GELUSO'S SPACE MZ ARRAY INCORPORATING FOUR VERTICALLY ORIENTED BIDIRECTIONAL MICROPHONES PAIRED WITH HORIZONTALLY ORIENTED CARDIOID MICROPHONES [66]. (USED BY PERMISSION). 51
FIGURE 37: PROPOSED NHK SPHERICAL MICROPHONE [122]. (USED BY PERMISSION). 51
FIGURE 38: OMNIDIRECTIONAL MICROPHONE WITH DIFFRACTION ATTACHMENT [123].(USED BY

permission). 52
Figure 39: Tetrahedral Microphone Capsule. 53
Figure 40: Studio 22. 58
Figure 41: Studio 22 RT30. 60
Figure 42: Studio 22 loudspeaker layout. 62
Figure 43: ME 25 free field frequency response [156]. 63
Figure 44: The averaged operational room response of all 22 loudspeakers. It represents the averaged spectral balance of sound radiation when all loudspeakers are operating [155]. 64
Figure 45: Original monophonic source image. 71
Figure 46: Low center loudspeaker added to image. 71
Figure 47: Rear middle center loudspeaker added to image. 72
Figure 48: Subwoofer loudspeakers expand image. 72
Figure 49: Narrow left and right middle loudspeakers adds focus to image. 73
Figure 50: Middle left and right rear surround speaker increase immersive content of image. 73
Figure 51: Expanded image is placed in hemispherical acoustic. 74
Figure 52: Eight Channel Surround Loudspeaker Ring. (Height channels copy first ring of loudspeakers 1.5 meters above). 77
Figure 53: Recording studio microphone positions corresponding to control room loudspeaker positions. 79
Figure 54: Recording studio microphone positions. 80
Figure 55: Spectral plot of recording studio (10 second sine sweep). 80

FIGURE 56: REVERBERATION TIME (RT30) OF RECORDING STUDIO. 81
FIGURE 57: FREQUENCY RESPONSE OF A FULL-RANGE LOUDSPEAKER IN THE RECORDING STUDIO. 81
FIGURE 58: 1959 HARMONY MONTEREY ARCHTOP GUITAR. 82
FIGURE 59: GRAPHICAL USER INTERFACE (GUI). 83
FIGURE 60: IMMERSION RATINGS GROUPED BY HEIGHT CHANNEL LEVELS 85
FIGURE 61: PREFERENCE CHOICES FOR ALL SUBJECTS: IMMERSIVE STIMULI WERE PREFERRED IN 98 OUT OF 145 TRIALS. 86
FIGURE 62: CONSISTENCY SCORES: EXAMPLES OF THREE SUBJECTS EXHIBITING HIGH, MEDIUM AND LOW CONSISTENCY SCORES 86
FIGURE 63: STUDIO 22 LOUDSPEAKER CONFIGURATION. 91
FIGURE 64: POSITIONING OF DISCRETE DELAYS. 95
FIGURE 65: REVERBERATION POSITIONS. 96
FIGURE 66: SUBJECTS RATINGS OF 3-DIMENSIONALITY AND IMMERSION. 98
FIGURE 67: SWEET SPOT: ASYMMETRICAL LEFT-TO-RIGHT. 99
FIGURE 68: SWEET SPOT ESTIMATIONS, LEFT AND RIGHT 99
FIGURE 69: COINCIDENT ARRAY. 107
FIGURE 70: M/S-XYZ ARRAY 107
FIGURE 71: MICROPHONE LOCATIONS IN NON-COINCIDENT/FIVE-POINT CAPTURE ARRAY FOR CELLO RECORDING. 108
FIGURE 72: NON-COINCIDENT/FIVE-POINT CAPTURE ARRAY. 109
FIGURE 73: MEASUREMENT OF SCHOEPS MK4 CAPSULES IN COINCIDEN 4-MIC ARRAY 110
FIGURE 74: MEASUREMENT OF SCHOEPS MK4 CAPSULES IN COINCIDENT STEREO PAIR. 110
FIGURE 75: MEASUREMENT OF SCHOEPS MK4 SINGLE CAPSULE 111

FIGURE 76: AGE OF SUBJECTS. 113
FIGURE 77: HOW SUBJECTS IDENTIFY THEMSELVES. 113
FIGURE 78: SUBJECTS' YEARS OF MUSICAL TRAINING. 113
FIGURE 79: SUBJECTS' YEARS OF AUDIO EXPERIENCE. 114
FIGURE 80: MICROPHONE-TO-LOUDSPEAKER POSITION DURING TESTS. 115
FIGURE 81: RESULTS OF TRAINING TESTS. 117
FIGURE 82: RESULTS OF MAIN TEST 118
FIGURE 83: SUBJECT RATING OF TRAINING TEST. 119
FIGURE 84: SUBJECT RATING OF MAIN TRIAD/ABX TEST. 120
FIGURE 85: SUBJECTS RATING FOR AUDIO IMAGE VERTICALITY. 121
FIGURE 86: SUBJECTS RATING FOR DEPTH OF AUDIO IMAGE. 122
FIGURE 87: SUBJECTS RATING FOR IMMERSION OF ARRAY IMAGE COMPARED TO MONO IMAGE. 122
FIGURE 88: SUBJECTS RATING FOR SIZE OF AUDIO IMAGE. 122
FIGURE 89: HORIZONTAL/VERTICAL GRID. 130
FIGURE 90: DEPTH GRID. 130
FIGURE 91: AVERAGED MAXIMUM EXTENT COMPARISON BETWEEN MICROPHONE ARRAYS. 132
FIGURE 92: SUBJECTS RATING OF TEST DIFFICULTY. 133
FIGURE 93: HORIZONTAL LEFT. 137
FIGURE 94: HORIZONTAL RIGHT 138
FIGURE 95: VERTICAL HIGH. 139
FIGURE 96: VERTICAL LOW. 140
FIGURE 97: DEPTH. 141
FIGURE 98: CELLO MONO. 143

FIGURE 99: TENOR SAXOPHONE MONO. 143
FIGURE 100: TRUMPET MONO. 144
FIGURE 101: VIOLA MONO. 144
FIGURE 102: CELLO COINCIDENT ARRAY. 145
FIGURE 103: TENOR SAXOPHONE COINCIDENT ARRAY. 145
FIGURE 104: TRUMPET COINCIDENT ARRAY. 146
FIGURE 105: VIOLA COINCIDENT ARRAY. 146
FIGURE 106: CELLO NON-COINCIDENT ARRAY. 147
FIGURE 107: TENOR SAXOPHONE NON-COINCIDENT ARRAY. 147
FIGURE 108: TRUMPET NON-COINCIDENT ARRAY. 148
FIGURE 109: VIOLA NON-COINCIDENT ARRAY. 148
FIGURE 110: CELLO MS-XYZ. 149
FIGURE 111: TENOR SAXOPHONE MS-XYZ. 149
FIGURE 112: TRUMPET MS-XYZ. 150
FIGURE 113: VIOLA MS-XYZ. 150
FIGURE 114: ALL MONO DRAWINGS. 151
FIGURE 115: ALL COINCIDENT ARRAY DRAWINGS. 151
FIGURE 116: ALL NON-COINCIDENT ARRAY DRAWINGS. 152
FIGURE 117: ALL MS-XYZ DRAWINGS. 152
FIGURE 118: PANNING AXIS OF DISCRETE LOUDSPEAKERS. 159
FIGURE 119: PANNING AXIS OF DISCRETE LOUDSPEAKERS DISPLAYING THE MINIMUM 0^{o} MONOPHONIC IMAGE 160
FIGURE 120: PANNING AXIS OF DISCRETE LOUDSPEAKERS DISPLAYING THE ORIGINAL 30^{o} IMAGE

EXPANSION. 160

FIGURE 121: PANNING AXIS OF DISCRETE LOUDSPEAKERS DISPLAYING THE MAXIMUM 90º IMAGE EXPANSION. 161

FIGURE 122: 3D DETAIL OF LISTENING CONFIGURATION. 161

FIGURE 123: TEST GUI. 163

FIGURE 124: TEST GUI WITH SELECTION. 163

FIGURE 125: PLOT OF VARIANCE OF PREFERENCE, SUMMED ACROSS ALL SUBJECTS, VERSUS TRIAL. THE LIGHT LINES ARE PLOTS OF THE RAW DATA, WHILE THE HEAVIER LINES ARE TRENDS SHOWING THE OVERALL DIRECTION OF VARIANCE USING A LINEAR BEST-FIT METHOD. FAMILIARITY WITH THE ARRAY SEEMS TO MANIFEST ITSELF OVER TIME AS VARIANCE DECREASES SIGNIFICANTLY FOR BOTH THE COINCIDENT AND NON-COINCIDENT ARRAYS. 166

FIGURE 126: HISTOGRAM OF EXPANSION ONSET, SUMMED ACROSS ALL SUBJECTS, VERSUS ARRAY. THE M/S-XYZ SHOWS BOTH THE EARLIEST EXPANSION ONSET AND THE STRONGEST AGREEMENT BETWEEN SUBJECTS WHILE THE COINCIDENT AND NON-COINCIDENT ARRAYS ALSO EXHIBIT A GREAT DEAL OF AGREEMENT, THE PEAK FREQUENCY OF RESPONSE FOR EACH IS APPROXIMATELY HALF OF THAT SEEN WITH THE M/S-XYZ ARRAY. 167

FIGURE 127: PREFERENCE VALUES FOR ALL ARRAYS, SUMMED ACROSS ALL SUBJECTS. 167

FIGURE 128: HISTOGRAM OF IMAGE BREAKDOWN (IN ±° AZIMUTH) BY ARRAY, SUMMED ACROSS ALL SUBJECTS. WHILE ALL ARRAY OPTIONS EXHIBIT THE EXPECTED TREND TOWARDS BREAKDOWN AT THE EXTREME EXPANSION OF THE IMAGE, THE ±15°, ±45°, AND ±75° PEAKS IN THE M/S-XYZ ARRAY ARE SOMEWHAT ODD. 168

FIGURE 129: IMAGE ONSET. 170

FIGURE 130: IMAGE BREAKDOWN 170

FIGURE 131: M/S XYZ ONSET TO BREAKDOWN RANGE.....171

FIGURE 132: NON-COINCIDENT ONSET TO BREAKDOWN RANGE. 172

FIGURE 133: COINCIDENT ONSET TO BREAKDOWN RANGE.....172

FIGURE 134: PROPOSED 27.2 MAXIMUM IMMERSIVE CONFIGURATION. 185

1 INTRODUCTION

1.1 Motivation

This research was initiated by the complete failure on the part of the researcher—a 30+ year professional in popular music production—to achieve a satisfactory result when tasked with providing an immersive presentation of popular music in an immersive multi-channel format for the 35th anniversary of the Schulich School of Music's Sound Recording Program at McGill University. This weeks-long failure brought into glaring light the ineffectiveness of the current offering of commercial production software and its digital audio workstations *(DAW*) platforms, but more pointedly the lack of understanding of the requirements for the creation of true three-dimensional audio images and environments.

The presentation of audio in three-dimensions is currently an area of intense interest in professional audio. It is being examined for delivery in cinema [2, 3], home theatre [3, 4], automobiles [5-7], headphones [8-10], through up-mixing [11, 12], via codecs [11, 13, 14] and as plugin applications for production in DAW [15-17]. The bulk of the investigations into recording techniques have largely been focused on classical music recording [18-20], broadcast [21, 22], film sound design [2, 3]and live event capture [23, 24] with little address of conventionally recorded popular music.

With the exception of the niche market composed of binaural releases [25], music recording and reproduction has been largely focused on monophonic and stereophonic releases with a limited number of 5.1 *surround sound* offerings. Stereo and 5.1 surround sound systems reproduce program at ear-level in a single horizontal plane. These listening paradigms do not deliver a fully immersive experience, nor do they approach the realism in reproduction of the playback formats that include height channels. Multi-channel sound systems that include height channels have been shown to improve immersion, envelopment, depth and presence in music reproduction by presenting a more realistic experience to the listener [26-28].

The introduction of multi-channel audio (beyond stereo and excluding Quad) largely took hold in the public consciousness when Dolby Labs suggested the 5.1 loudspeaker array as a format for cinema sound delivery in 1976 [29]. Early pioneers of height dimension in recording of music included *Periphony* (with-height sound reproduction, Michael Gerzon, 1973 [30]), 2+2+2 DVD releases [31], and experimental SACD releases of DMP recordings with a single rear-overhead channel (Telarc 2001 [32]). The industry has lately experienced the expansion of immersive formats that include several height channels. Some of the most notable formats are Dolby Atmos for cinema sound with up to 64 speaker channels [33], Auro 3D specifying up to 13.1 channels [34], and the 22.2 Multi-channel Sound System developed by the Japanese broadcaster NHK [1]. It should be noted that the above formats are for sound with picture and were not developed for the exclusive delivery of musical content.

Multi-channel commercial releases of popular music have been largely limited to 5.1 and 7.1 formats with no height channels present. Since 2015 there have been sporadic commercial releases of formats that include height channels. They have been largely from individual artists, but also labels such as *2L—the Nordic Sound*, *Unamas*, *Sono Luminus* and *Delphian Records*. The

sound-stage architecture and formats in these releases varies from 5.1 and 7.1 surround formats to 9.1 Auro and Dolby Atmos immersive formats.

3D audio is currently a 'hot topic' across the media landscape with the advent of virtual and augmented reality, a multitude of playback formats, and a wide variety of hardware capture options that range from *ambisonic* [35, 36] to 360º preconfigured microphone arrays [37]. However, little fundamental work has been undertaken on the best recording and mixing practices, or on validated capture techniques that reliably present believable three-dimensional audio images. This investigation is a small first step in these directions.

1.2 Research Goals

The research goals are:

1. Develop methodologies for creating three-dimensional audio images of individual musical instruments presented using commercially available software.
2. Develop methodologies for creating the simulation of three-dimensional immersive environments using commercially available software.
3. Determine the optimal playback levels of height-channel information that are considered to be effective in music mixing.
4. Develop microphone arrays for music recording of individual instruments to create adjustable vertical and three-dimensional images without the use of added processing.
5. Conduct subjective testing to examine the perceived three-dimensional extent of the audio images captured by each microphone array.

1.3 Structure of Thesis

This thesis is comprised of the following chapters:

Chapter 1: Introduction summarizes the goals and motivation for this research.

Chapter 2: Background provides a review of literature relating to the development and evaluation of spatial audio: spatial hearing, stereo and surround recording techniques, 3D audio recording techniques and the subjective assessment of spatial sound quality.

Chapter 3.2-3.3 investigate mixing techniques for the creation of three-dimensional audio images of individual instruments in popular music reproduced in the 22.2 Multi-channel Sound System. This research focuses on the technical requirements needed to achieve three-dimensional believability and immersion using conventional audio tools. This chapter also investigates the playback levels of height-channel information that are considered to be effective in music mixing.

Chapter 3.4: Three Dimensional Spatial Techniques in 22.2 Multi-channel Surround Sound for Popular Music Mixing investigates strategies for the creation of compelling high-fidelity mixes of popular music in the 22.2 Multi-channel Sound System. This investigation builds upon the findings in Chapter 3 and examines the design and implementation of early and late reflections, and reverberant fields. The techniques discussed include the expansion of spatial elements into three dimensions using conventional tools, and the implementation of multi-channel impulse responses for reverberant fields. A subjective listening test suggests that it is possible to create believable three-dimensional immersion using conventional stereo spatial tools in a 22.2 multi-channel playback environment from monophonic multi-track source material.

Chapter Error! Reference source not found.: **Development of Vertical Imaging Microphone T echniques**. It became apparent from the results and experiences garnered from the Chapter 3 that to create impactful and compelling content for immersive playback systems that it would be necessary to create sonic images of greater realism than could be generated though mix processing of mono- and stereophonic source material used in conventional multitrack recordings. The findings of Chapter 3 suggested that traditional close capture techniques did not provide enough information for the requirements of 3D believability. It was hypothesized that spatially enhanced images of greater realism and impact could be created during the recording process.

Chapter 4: Microphone Arrays for Vertical Imaging and Three-Dimensional Capture from Monophonic Program progresses from the creation of 3D audio images by artificial means at the mixing stage to research into the capture of a 3D audio image at the recording stage. The first study compares a selection of microphone arrays for music recording to create vertical and three-dimensional images without the use of added processing. The majority of the subjects rated the images captured by the microphone array as having an extended verticality over the mono source, and that the 3D microphone arrays provide a greater depth of image than the mono capture. Results also showed that subjects could discern the arrays from one another, and also that each array could be easily identified. A further study employed a simple graphical method in an effort to represent the perceived spatial attributes of the microphone arrays and a mono/point-source signal. The subjects' representations support that these arrays clearly capture much more information than a single microphone.

Chapter 5: Subjective Graphical Representation of Microphone Arrays for Vertical Imaging and Three-Dimensional Capture of Acoustic Instruments is an investigation that employs a simple graphical method in an effort to represent the perceived spatial attributes of three microphone arrays designed to create vertical and three-dimensional audio images. Three separate arrays were investigated in this study: Coincident, M/S-XYZ and Non-coincident/Five-point capture. Test subjects were asked to represent the spatial attributes of the perceived audio image on a horizontal/vertical grid and a graduated depth grid, via a pencil drawing. Results show that the arrays exhibit a greater extent in every dimension—vertical, horizontal and depth—compared to the monophonic image. The statistical trends show that the spatial characteristics of each array are consistent across each dimension.

Chapter 6: Subjective Assessment of the Versatility of Three-Dimensional Near-field Microphone Arrays for Vertical and Three-Dimensional Imaging employs a 3D panning tool to manipulate audio images recorded with advanced close-capture microphone arrays for three-dimensional imaging. The objective of this interactive test was to determine the point of three-dimensional expansion onset, preferred imaging, and image breakdown point.

Chapter 7 details the general conclusions of this research.

Chapter 8 discusses future work.

2 BACKGROUND

2.1 Background of Recorded Media and Reproduction

The role of the recording engineer from the inception of recorded media has been one of the capture and preservation of information—transferring the most accurate representation of the audio event to the listener. Throughout the 160+ year history of recorded sound there has been a constant progression of expanding bandwidth and increasing number of reproduction channels [1, 25, 38]. The recording medium has progressed from Leon Scott's lampblack-coated cylinder (1857) to the early Edison cylinder (1876, *phonograph cylinder*) and Emile Berliner's lateral tracking 'flat disk' (the *gramophone record*, and progenitor of the ubiquitous LP) [38] to the Hi-Res data files available today. The early media were monophonic, or single-channel playback, which had an upper cut-off frequency between 2-3 kHz [39]. The next great step in the improvement of recorded media was two-channel or stereophonic. One of the pioneers in this field was Alan Dower Blumlein who created a means for recording two channels on a single disc in 1933 [40, 41]. This bore a great increase in the transfer of information and, therefore, 'realism', owing to the fact that normal human audio perception is binaural—where two ears are employed by the brain to interpret the incoming sounds [39, 40, 42]. The stereo LP (Long Play) $33^{1/3}$ rpm record was introduced in 1958 [43] with the capability of reproducing 20 kHz. The analog disc (which over its 100-year commercial reign sold over 30 billion units) was supplanted in 1983 by the digital compact disc (CD) format invented by Sony/Philips [39].

The first introduction of a consumer multi-channel audio format (beyond the standard two-channel stereo) was Quadraphonic (four-channel) which appeared as the Quadraphonic 8-Track Tape (1970) and the Quadradisc in 1972. The music industry's failed experiment with Quadraphonic sound was largely over by 1976 [44].

It was in the context of sound-for-picture that multi-channel audio gained an early foothold. For the music consumer, this began in 1976 when Dolby Labs suggested the 5.1 loudspeaker array as a format for cinema [29]. Since that time, there has been an expansion of immersive formats that include height channels. Some of the most notable formats are Dolby Atmos for cinema sound [2], Auro 3D specifying up to 13.1 channels [34], and the 22.2 Multi-channel Sound System developed by the Japanese broadcaster NHK [1].

2.2 Multi-channel Audio

2.2.1 Stereophonic Sound

Multi-channel audio for commercial distribution and consumption began with the declaration of the international standard for stereo records which was decided at the RIAA (Recording Industry Association of America) meeting in Indianapolis in December of 1957 [45]. There were two competing systems for the standard: the V-L system developed by Decca in London, U.K., and the Westrex 45-45 system in Hollywood, California. The Westrex 45-45 was chosen.

"*The Westrex Stereo Disk Recorder . . . records two stereophonic channels in a single groove with a single stylus. The axes of the two recordings are at 90 degrees to one another, each being at 45 degrees with the horizon plane of the record"* [46].

The RIAA *Standard Recording and Reproducing Characteristic*—commonly known as the RIAA equalization curves—are shown in Figure 1[47]. The yield from the combined curve is designed to be a linear frequency response from 20 Hz to 20 kHz.

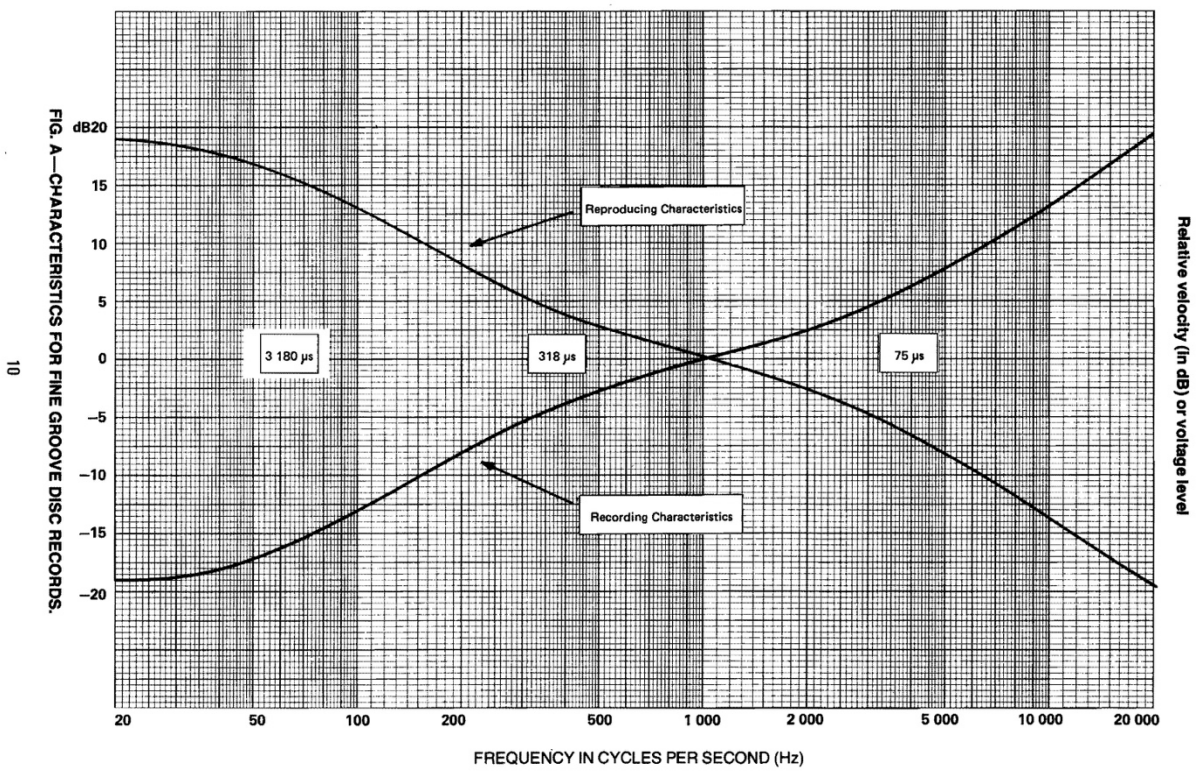


Figure 1: RIAA Standard Recording and Reproducing Characteristic.

2.2.2 Quadraphonic Sound

The inception of what has come to be termed “Surround Sound” began in early 1968 in a conversation between Thomas Mowrey of Vox Records, and Robert Berkovitz of Acoustic Research (based in Cambridge, Massachusetts). The concept discussed was a rectilinear, four-microphone array where one conventional stereo pair was pointed directly at the performing ensemble, and a second stereo pair was pointed away from the ensemble to pick up the indirect or ambient sound. This array would supply 360º field of capture to be played back on four similarly positioned loudspeakers, and thus ‘surround’ the listener (Figure 2).

The technology for delivering the four channels of audio via a conventional stereo LP (or other 2-channel medium) was invented by bassoonist Peter Scheiber. The company to commercialize this technology, Audiodata, was formed by Mowrey and Scheiber in 1969 [48].

Scheiber's invention became known as "matrix quadrophony". The primary shortcoming of the circuit was that, while there was complete electrical separation between diagonally opposing channels, the separation between adjacent channels was only 3 decibels. This 3dB separation (or lack of) was perceived as a diminishing of the sound field. Subsequent circuits (such as the *QS* developed by Sansui, and *SQ* developed by Columbia Records and Sony) improved on this shortcoming, but arrived too late to rescue the format commercially, and never completely overcame the obstacle of a diminished stereo sound field.

Dubbed the *". . . stereo of the future"* by High Fidelity magazine in 1969 [49] and *"the New Surround Sound"* by Electronic World in 1970 [50], "matrixed quad" had a short and ultimately unsuccessful life, which was largely over by 1976.

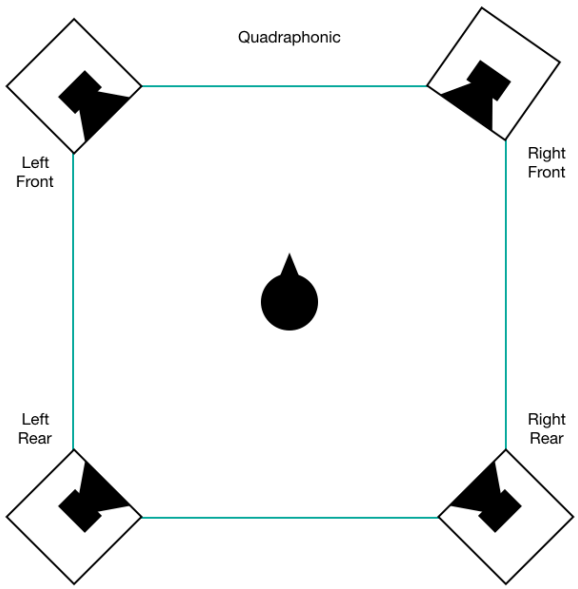


Figure 2: Quadraphonic Loudspeaker arrangement.

2.2.3 Brian Eno: “An Ambient Speaker System”

After the demise of quadraphonic sound, no commercial surround formats for music were promoted until 1999 with the launch of the Super Audio CD (SACD) by Sony and Philips, and in 2000 with the introduction of DVD-Audio from the consortium of Matsushita, Toshiba and Warner [51]. But apparently there was one non-commercialized and un-promoted option.

In 1982 Brian Eno released the fourth album in his ambient music series—Ambient 4: On Land [52]. The album art included the loudspeaker and wiring diagram for an immersive audio configuration that Eno had been using for “many years” (Figure 3).

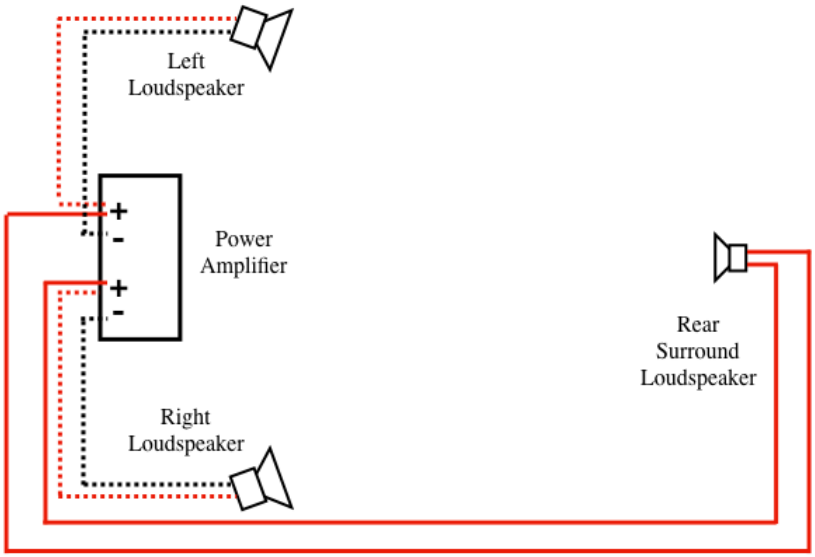


Figure 3: Brian Eno’s Ambient Speaker System.

Eno’s system is a simple three-loudspeaker setup. The only additional hardware required, beyond the standard stereo audio equipment, is the third loudspeaker. The surround loudspeaker is positioned behind the listener at the apex of a triangle formed by the three loudspeakers. The terminals of the surround loudspeaker are connected to the left and right positive terminals of the power amplifier.

Although Eno writes he *"arrived at this system by accident, and I don't really know why it works"*, his explanation in the accompanying album text seems to belie this modesty.

> *"What seems to happen is that the third speaker reproduces any sound that is not common to both sides of the stereo - i.e., everything that is not located centrally in the stereo image - and I assume that this is because the common information is put out of phase with itself and cancels out. More technically, the lower the impedance of the added speaker, the louder it will sound.*
> *If it is found to be too loud (although this rarely seems to happen), you can either insert a potentiometer (6-12 ohms, at least 10 watts) into the circuit, or move the speaker further away."*

And he further explains:

> *"The usage of this speaker in the three-way system is such that it will not be required to handle very low frequencies: therefore, a small or "mini" speaker will be adequate."*

The advantages of Eno's system are that of compatibility with any stereo recording, is of low-cost, and easy implementation in most domestic environments, furthermore, it is also free of any rights and licenses . . . which is likely the reason for it being obscured to the dusty archives of pop music history.

2.2.4 Channel-based Surround Sound

Channel-based surround (or immersive) audio formats have a 1:1 relationship between the number of audio channels and playback loudspeakers. Early work on this type of reproduction was undertaken at Bell Labs in the 1930s. In these experiments a spaced array of omni-directional

microphones were each connected directly to a corresponding loudspeaker in a listening room [53]. Steinberg and Snow [54] also conducted multi-channel research in this era. They found that three-channel audio gave convincing results in large auditorium with wide-screen pictures. In this three-channel format, the center channel provided a stabilizing effect for the central dialog channel for off-axis listeners. This format gained acceptance dating from its use by the Walt Disney company in the 1939 film *Fantasia*. This multi-speaker system came to be known as "Fanta-sound" and is detailed by Garity & Hawkins in a 1941 article in the SMPTE (Society of Motion Picture and Television Engineers) Motion Imaging Journal [55].

In the early 1950s *Stereophonic Sound* was promoted in tandem with the new wide-screen visual formats. Cinema *'Stereophonic Sound'* differed from the two-channel stereo that would later become the standard for phonograph records. *Stereophonic Sound* for cinema began and continues to utilize a minimum of four audio channels.

The dominant formats were four channel CinemaScope on 35mm film (Figure 4), and six channel Todd-AO on 70mm film (Figure 5). The minimum features of multi-channel film sound consist of several playback channels located in the front, and at least one in the rear. The rear effect(s) channel was initially reserved for special effects, but later became used for ambience. This approach provided a more immersive environment for the film goers. The application of using the rear channel(s) for immersive content came to be known as *'surround sound'*, and the effects channel as the *'surround channel'* [56].

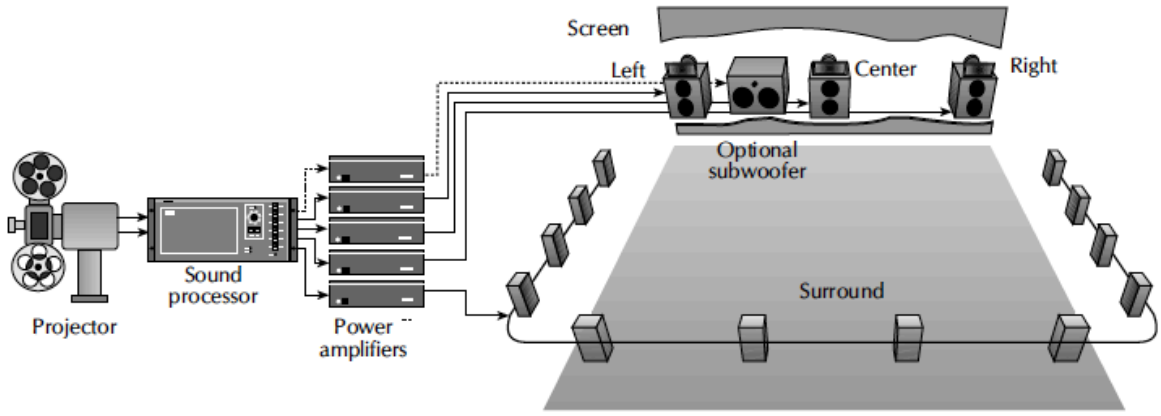


Figure 4: Four-channel surround Sound for magnetic film.[56], (Used by permission).

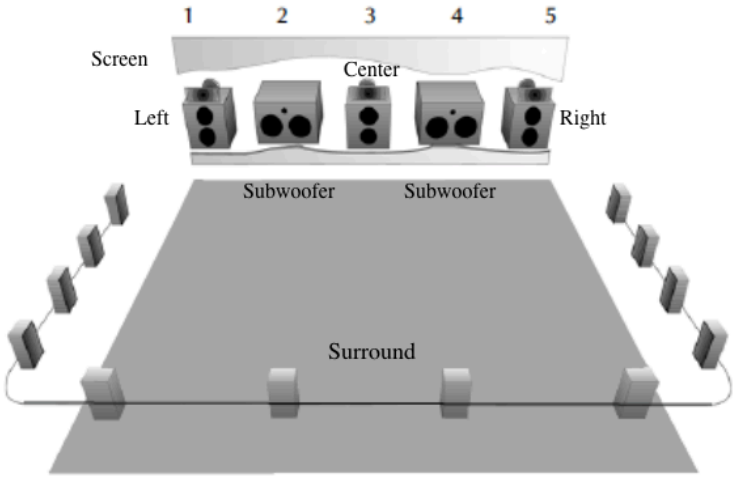


Figure 5: 70mm six-channel surround sound for magnetic film.[56], (Used by permission).

2.2.5 5.1 Surround Sound

The introduction of the compact disc (CD) in 1982 enticed the film industry to explore a digital format for sound in cinema. The industry agreed on a discrete channel format, which eventually became known as the 5.1 channel configuration. The 5.1 configuration contains five full range loudspeakers—left, center, right, left surround, right surround—with a subwoofer as the '.1' channel which carries the low-frequency effect (LFE) information.

The first 5.1 digital format appeared in 1990. This was the *Cinema Digital Sound* (CDS) format for 70mm print film. It was developed by Optical Radiation Corp. in conjunction with Kodak. 1992 saw the arrival of three competing formats for 35mm film: Dolby Digital, Digital Theater Sound (DTS), and Sony Dynamic Digital Sound (SDDS) [56]. The standard was formalized by the International Telecommunications Union ITU-R BS.775 in 1993, (Figure 6), [25, 57]. The IMAX format employed three synchronized CD players to reproduce digital sound over five horizontal channels (L,C,R,Ls,Rs) and one elevated channel at the centre top of the screen [58, 59].

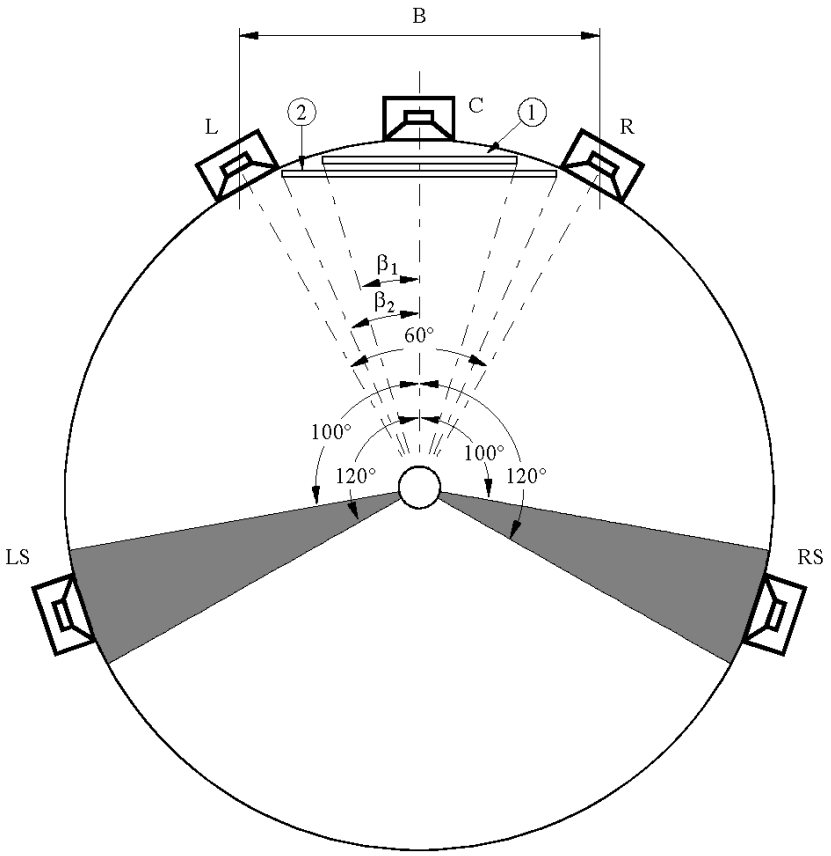


Figure 6: ITU-R BS.775-1 5.1 reference loudspeaker arrangement.

2.2.6 7.1 Surround Sound

The 7.1 surround sound format came about as a configuration that would provide a larger optimum listening area and increased envelopment. Two additional rear/side loudspeakers were added to the 5.1 surround specification (Figure 7). This specification is detailed in Recommendation ITU-R BS.775-2 [60].

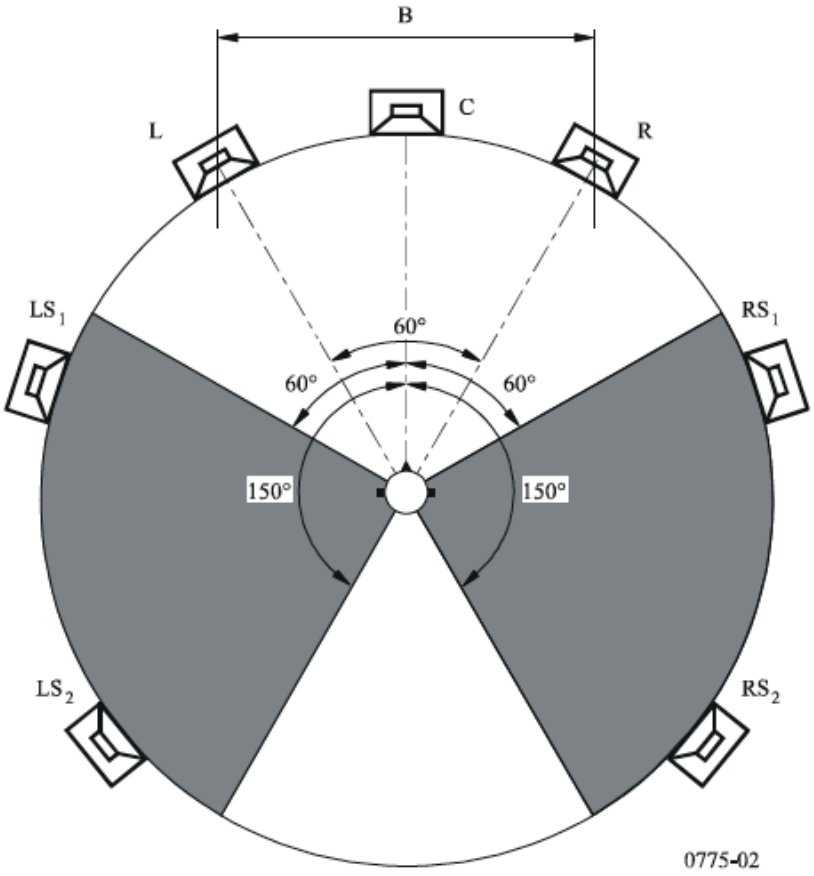


Figure 7: ITU-R BS.775-2 7.1 reference loudspeaker arrangement.

2.2.7 Multi-channel Formats that include Height Channels

Multi-channel formats that include height channels were developed to add a sense of three-dimensional spatial impression, to increase envelopment, and to provide a greater sense of listener engagement. This is achieved by augmenting the middle surround layer of loudspeakers (located

at ear level) with additional loudspeakers positioned above and below the middle listening plane. The Report ITU-R BS.2159-7 [61] states:

> *"Because each loudspeaker of the 5.1 channel sound system is set at the same height as the listener's ears, the sensation of spatial reality is fundamentally limited to the horizontal plane. For advanced multi-channel sound systems beyond the 5.1 channel sound system, the sensation of spatial impression should be enhanced around the listener, including in the upward/downward elevation sensation, reverberation and ambience."*

The ITU-R BS.2051–0 [62] specification provides a comprehensive summary of the loudspeaker configurations for multi-channel audio reproduction. The recommendation details the positional and directional configurations of loudspeakers using three layers: upper, middle, and bottom. The middle layer indicates the horizontal plane (ear level) while the upper and bottom layers indicate the height and ground level planes.

2.2.7.1 Middle Surround Sound Configurations with Two Height Channels

These configurations augment common 5.0 and 7.0 middle layer configurations with two additional height loudspeakers (Figure 8). Dolby [33] proposes a similar configurations. Kamekawa et al. [27] and Kim et al. [28] have found that height-channel content provides a natural-sounding perception of depth and improved envelopment.

2.2.7.2 10.2 Surround Sound System Type A

The Type A 10.2 (Figure 9) format was developed by THX creator Tomlinson Holman [63] and The Immersive Audio Laboratory (which is a part of the Integrated Media Systems Center at the University of Southern California). This sound system is based on the 5.1-channel layout in

Recommendation ITU-R BS.775 [57] and detailed in Report ITU-R BS.2159-7 [61]. This configuration is better known as the THX 10.2 system.

The configuration of the 10.2 Type A surround system consists of:

- Middle layer loudspeakers: Center 0º, Left -30º, Wide Left (LW) -60º, Right +30º, Wide Right (RW) +60º, Left Surround -110º and Right Surround +110º, Back Surround (BS) 180º.
- Upper layer loudspeakers: Left Height (LH) -45° and Right Height (RH) +45° elevated +45° above the median plane.
- Two LFE Subwoofers ('.2' channels) ideally located ±90º.

2.2.7.3 10.2 Surround Sound System Type B

The 10.2 channel sound system Type B (Figure 10) was developed by South Korea's National Radio Research Agency (RRA) jointly with Samsung Electronics and the Electronics Telecommunications Research Institute (ETRI) [64]. It defines the "Audio Signal Formats for Ultra High Definition (UHD) Digital TV" in the Republic of Korea, TTAK.KO-07.0098 in 2011 [61]. It became a standard for UHDTV broadcasting at the Asia-Pacific Broadcasting Union (ABU) in October 2013.

The Type B setup is based on the 5.1-channel standard. The configuration of the 10.2 Type B surround system consists of:

- Middle layer loudspeakers: Center 0º, Left -30º, Right +30º, Left Side (LS) -90º, Right Side (RS) +90º, Left Back (LB) -135º and Right Back (RB) +135º.
- Upper layer Loudspeakers: Left Height (LH) -45° and Right Height (RH) +45° Center Height Channel (CH) 180º elevated +45° above the median plane.
- Subwoofers: LFE1 left side on bottom layer, LFE2 right side on bottom layer.

2.2.7.4 Auro-3D Immersive Sound Systems

Auro Technologies NV was founded by Wilfried and Guy Van Baelen, based at Galaxy Studios, Mol, Belgium. The Auro-3D concept was developed in 2005. The Auro 3D loudspeaker setups range from 8.0 up to AuroMax 26.1 [3, 65]. All setups contain at least the four upper/height loudspeakers positioned directly above those in the middle ring to facilitate compatibility with existing commercial formats [66].

The most common setup is the Auro 9.1 configuration (Figure 11) based on the standard ITU-R BS.775 5.1 surround setup [67]. In the Auro 11.1 configuration (Figure 12) the upper layer mirrors the middle layer, but contains an additional loudspeaker located directly above the listener. This center-overhead loudspeaker has been named "the voice of god" (VOG). The VOG loudspeaker contributes to the production of a hemispherical sound field. Theile and Wittek [68] have reported that the Auro 9.1 configuration has improved depth, envelopment and spatial impression over conventional stereo reproduction.

2.2.7.5 11.1 Immersive Sound Systems

The 11.1 immersive sound system (Figure 13) is outlined in the ATSC 3.0 (Advanced Television Systems Committee) specification [69]. This configuration is based on the 7.1 surround sound format but includes four upper channels. This configuration is also the most common Dolby Atmos configurations for home theatre [33].

2.2.7.6 22.2 Multi-channel Sound System

The 22.2 multi-channel sound system (Figure 14) was developed by Nippon Hōsō Kyōkai (NHK) Science and Technology Research Laboratories (STRL) as part of their Super Hi-Vision

ultra-high resolution video system. This system scans 4000 video lines across a viewing angle of 100° [70].

The configuration of the 22.2 surround sound system consists of:

- Middle layer loudspeakers (10 channels): FC 0º, FLc -30º, FL -60º, FRc +30º, FR +60º, SiL -90º, SiR +90º, BL -135º, RL+135º, BC 180º.
- Upper layer loudspeaker (9 channels): TpFC 0º, TpFL -60º, TpFR +60º, TpSiL -90º, TpSiR +90º, TpBL -135º and TpRL +135º, TpBC 180º, TpC 0º overhead.
- Bottom Layer (3+2 channels): BtFC 0º, BtFL -60º, BtFR +60º, LFE1, LFE2.

The nine loudspeakers of the upper layer are employed to produce a more accurate localization of audio images in the elevated sound field. The VOG loudspeaker is also present to aid in the creation of a hemispherical sound field. The five requirements NHK set for the 22.2 system are summarized by Hamasaki [70]:

1. Integrity: the localization of an audio image anywhere on the UHD screen.
2. Periphony: the reproduction of sound coming from all directions surrounding the viewing position.
3. Presence: the reproduction of a natural, high-quality 3D acoustic space.
4. Compatibility: with existing multi-channel configurations.
5. Usability: the capability to support a live recording and live broadcasting.

In a comparative study between the 22.2 system, conventional stereo, and 5.1 surround [71], the 22.2 system was rated superior in the perceived attributes of front/rear and up/down discrimination, in movement, direction, and had greater reverberance and envelopment.

The 22.2 multi-channel sound format has been standardized by the International Telecommunications Union (ITU) [62] and Society of Motion Picture Engineers (SMPTE) [72].

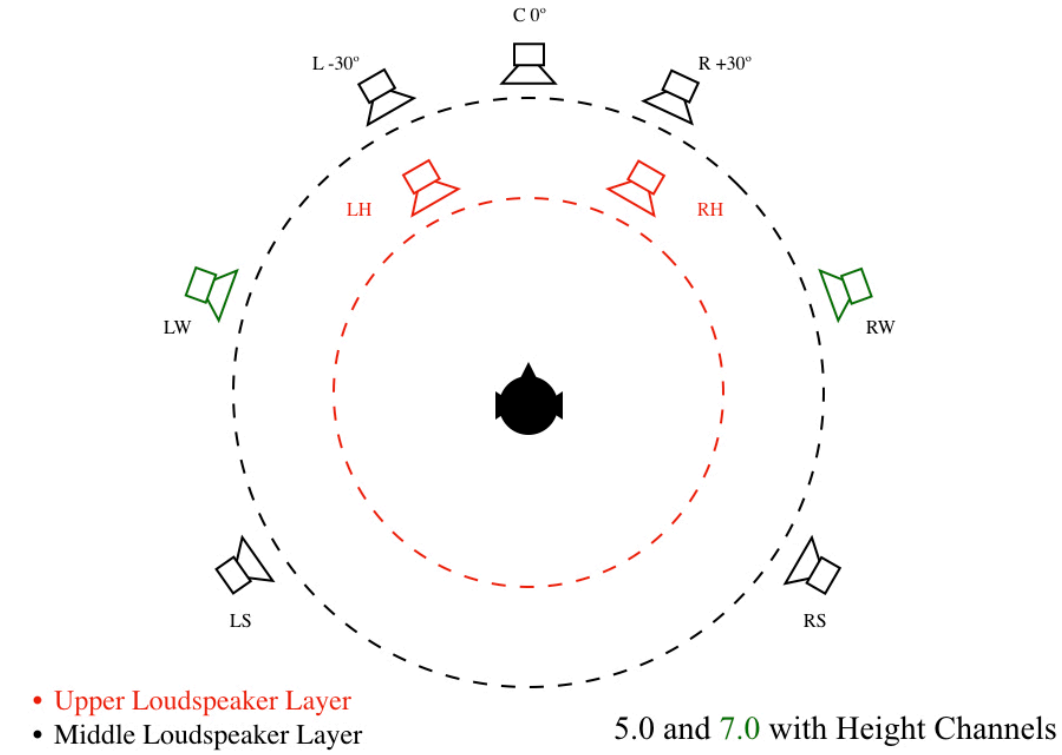


Figure 8: 5.0 and 7.0 with 2 Height Channels

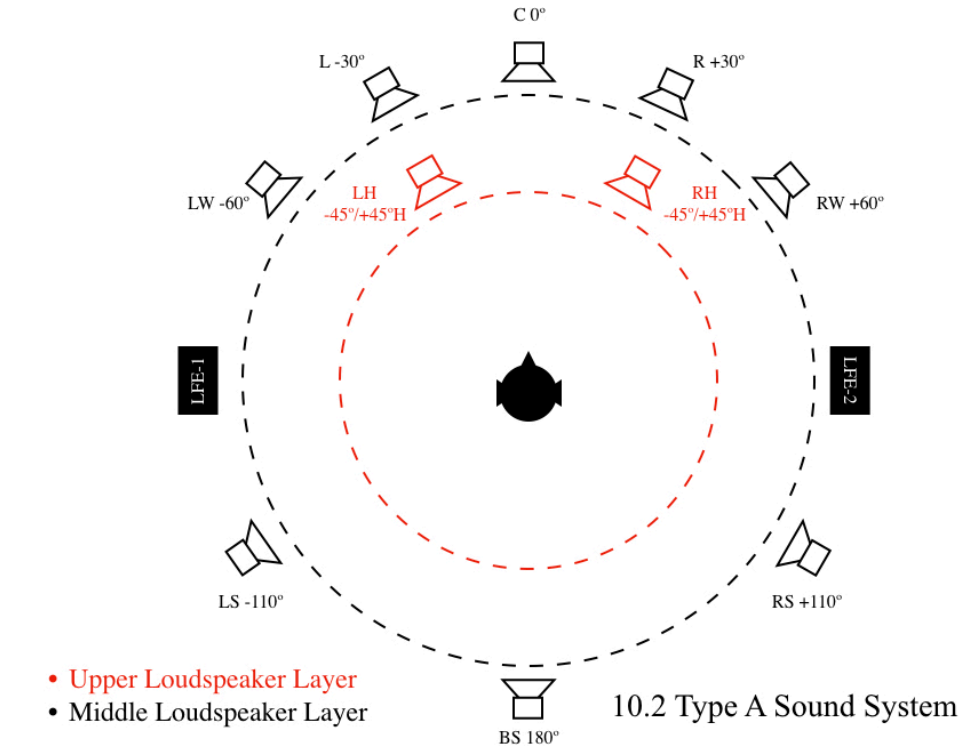


Figure 9: 10.2 Type A Sound System with 2 Height Channels.

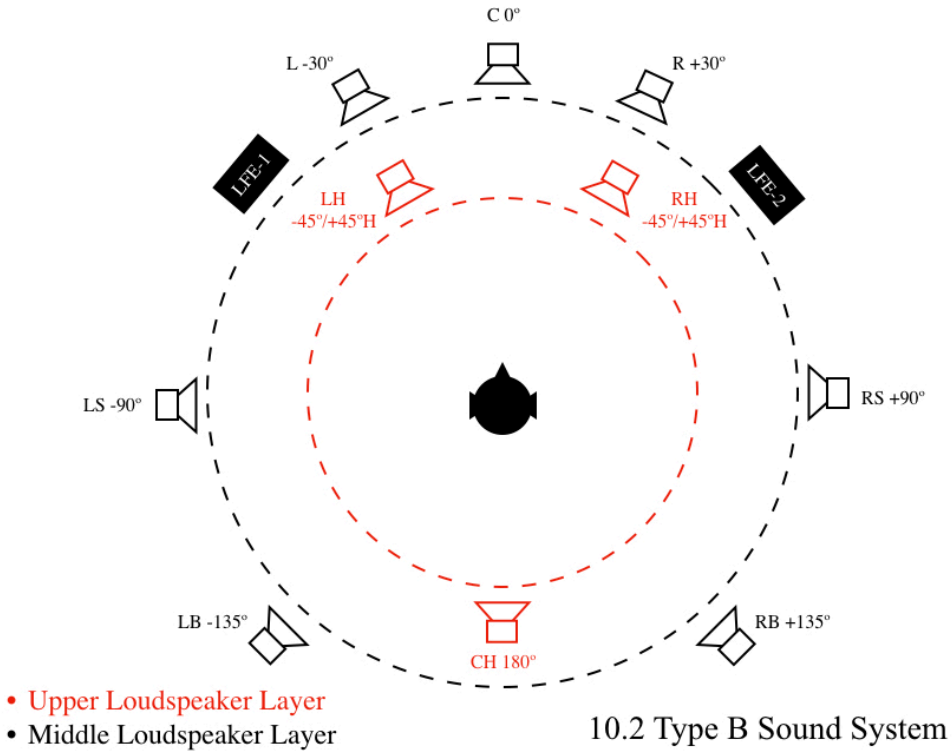


Figure 10: 10.2 Type B Sound System with 3 Height Channels.

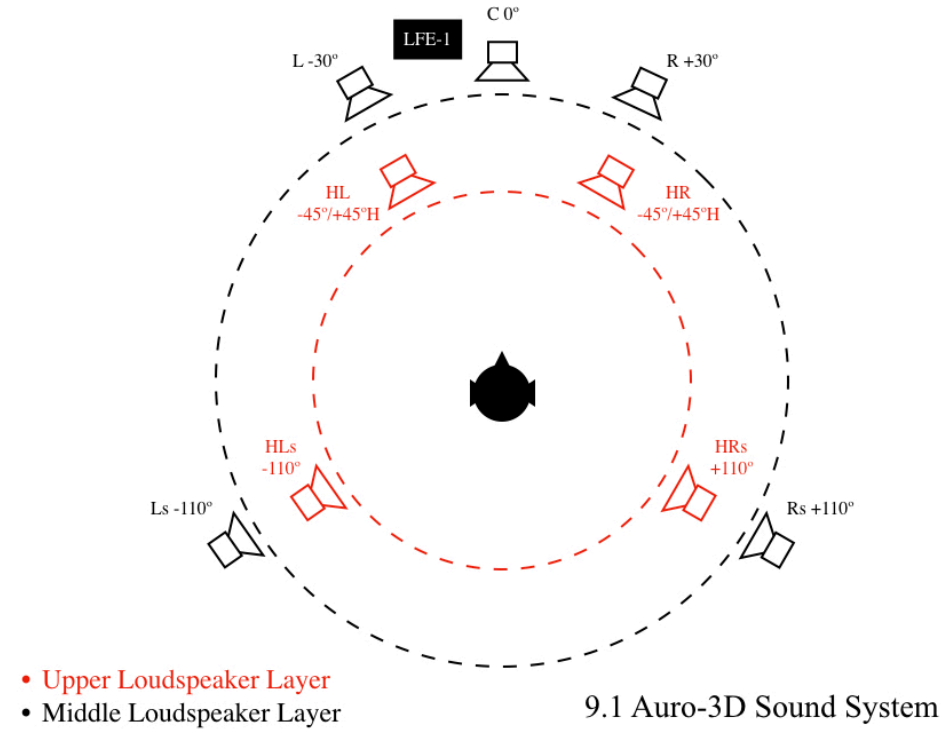


Figure 11: 9.1 Auro-3D Sound System with 4 Height Channels.

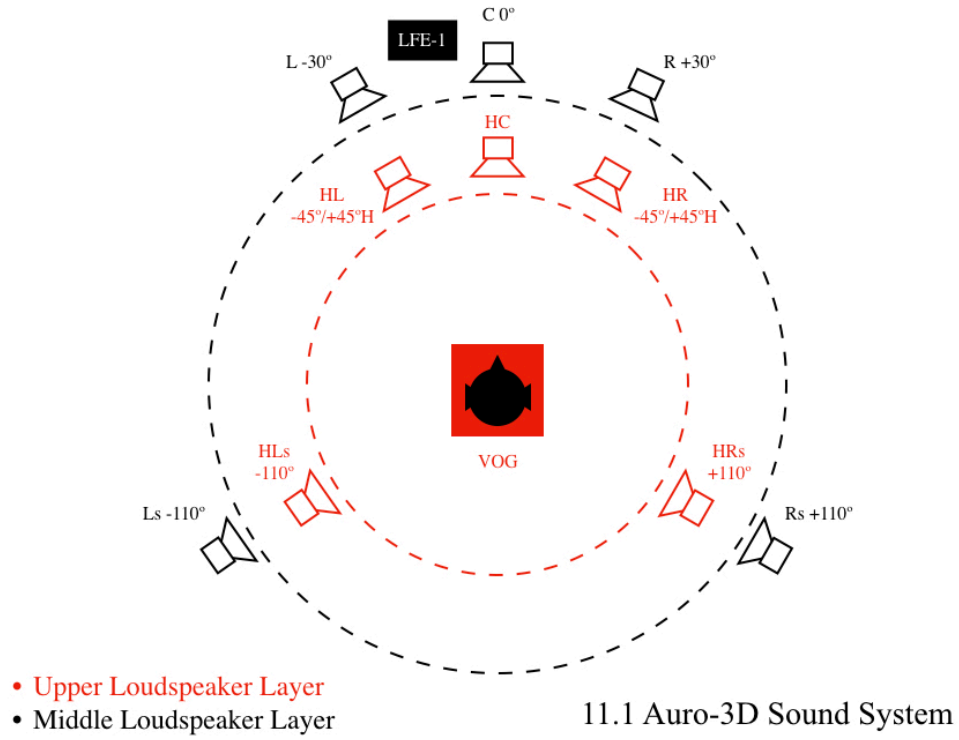


Figure 12: 11.1 Auro-3D Sound System with 6 Height Channels.

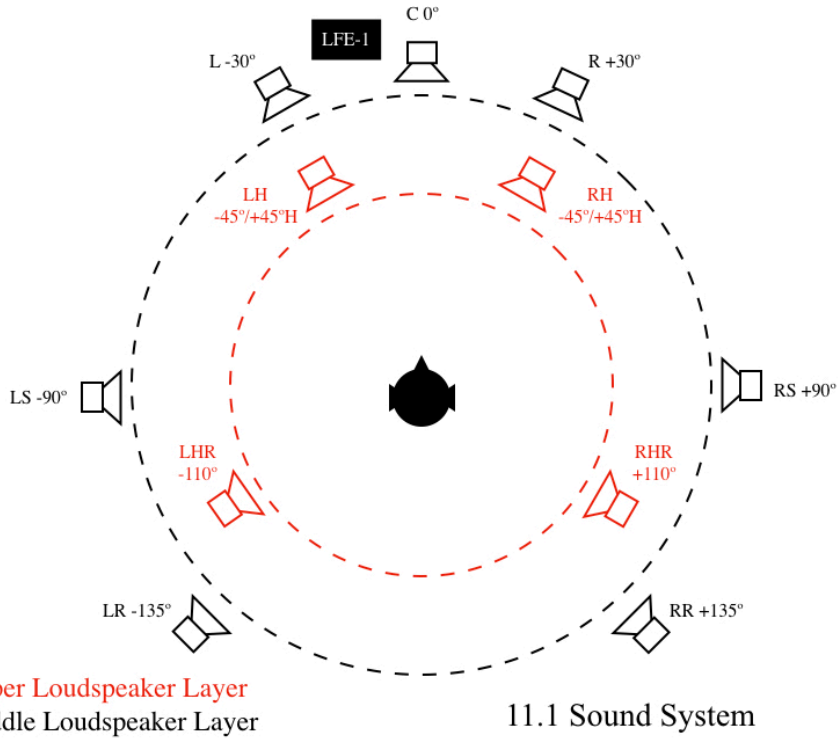


Figure 13: 11.1 (7.1+4) Sound System with 4 Height Channels.

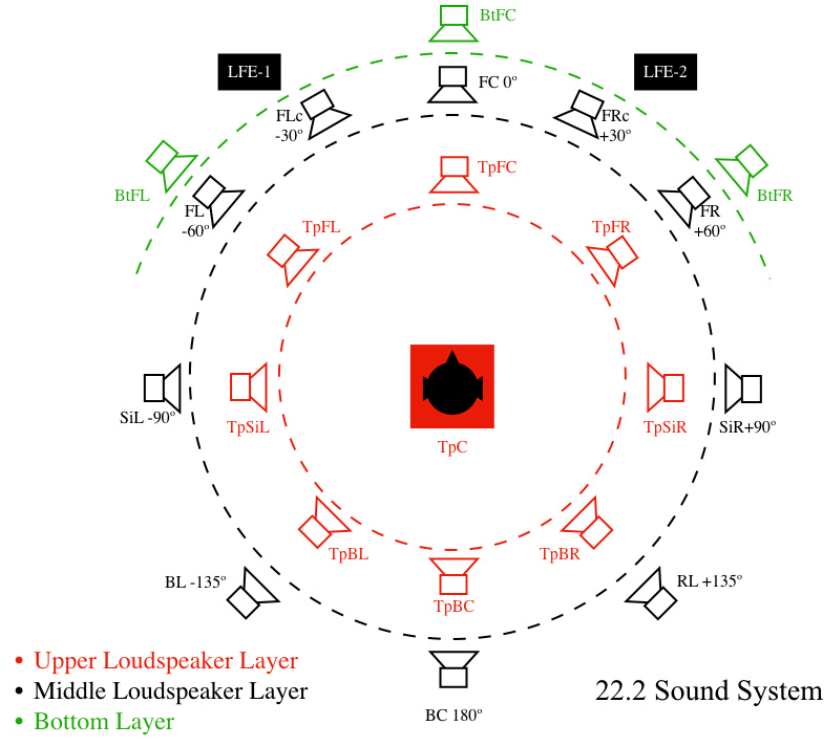


Figure 14: 22.2 Sound System with 9 Height Channels.

2.3 Spatial Hearing

Spatial hearing in human beings combines the perception of the spatial properties of a sound source, the geometric location of the source with respect to the listener, and the acoustic properties of the environment. The auditory system operates binaurally due to the fact that human beings possess two ears located on the opposite sides of the head. The cues for localization of a sound source involve the relative differences between the signals arriving at each ear. These differences are in arrival time—Interaural Time Differences (ITD), and intensity—Interaural Intensity Differences (IID). When perceiving continuous pure tones and periodic signals—which have no clear reference point in time—Interaural Phase Differences (IPD) are employed instead of ITD.

This topic has been presented in many comprehensive texts such as Blauert's *Spatial Hearing* [73], by Braasch in Blauert's *Communication Acoustics* [74], Grantham in Moore's

Hearing [75], and also in Begault's report for NASA on *3-D Sound for Virtual Reality and Multimedia* [42].

2.3.1 Localization and Localization Error

A comprehensive discussion of Localization Error or *Blur* was put forward by Letowski and Letowski in 2011 [76]. Blauert, published in 1983, remains a seminal text as attested to by Yost in his Resource Reviews for *Ear and Hearing Journal* in 1998 [77].

Localization as defined by Blauert: "*is the law or rule by which the location of an auditory event (e.g., its direction or distance) is related to a specific attribute or attributes of a sound event, or of another event that is in some way correlated with the auditory event.* Letowski and Letowski expanded upon this definition to include 'estimation error' of localization: as follows: *". . . is an estimate of the actual location of sound source in space and is characterized by a certain amount of inherent uncertainty and operational bias that results in estimation errors. The type and size of the estimation errors depend on the properties of the emitted sound, the characteristics of the surrounding environment, the specific localization task, and the abilities of the listener."*

In some respect, Localization, along with its accompanying errors, is a measure of spatial uncertainty and bias in the perception of a sound source's location.

2.3.1.1 The Minimum Audible Angle

Mills [78] and Perrot [79] define the minimum detectable difference in azimuth or elevation between locations as the *Minimum Audible Angle* (MAA). While being dependent on both frequency and direction of arrival, it has been reported by Mills, Perrott, and Kuhn [78-80] that for wideband stimuli and low frequency tones, the MAA is on the order of 1° to 2° for sounds

originating from the frontal position, 8-10° at ±90°, and 6-7° for sounds originating behind the listener Figure 15.

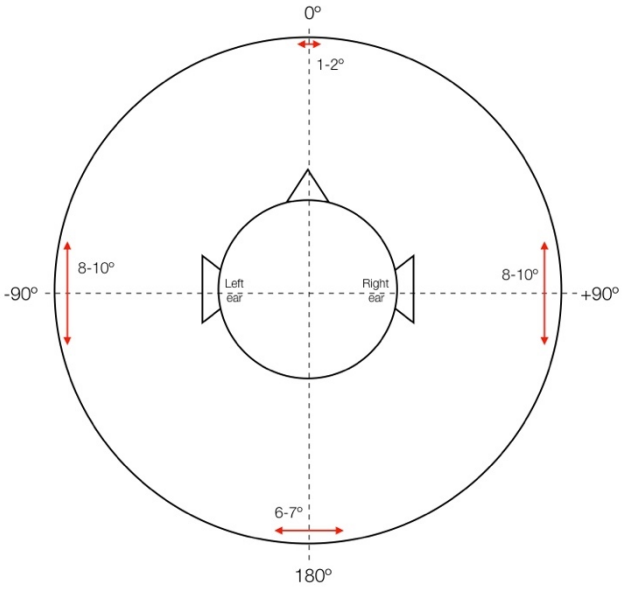


Figure 15: Localization Error [78-80] .

2.3.2 Perception in the Horizontal Plane

A sound source deviating for 0º front-centre will produce time and intensity differences on arrival at each ear. This time difference is related to the angle of incidence Ø (Figure 16), and reaches a maximum at ±90º (Figure 17) when the time difference is approximately 650µs [25, 42]. This difference in arrival time yields the Interaural Time Difference (ITD) illustrated in Figure 16. It shows a sound emanating to the right of a listener will arrive before the sound travels around the head to the left ear.

Rumsey [25] and Zhang [81] state that the human psychoacoustic system does not compare these arrival times directly due to the fact that the neurons compare the Interaural Phase Differences (IPD) while Letowski and Letowski clarifies that continuous pure tones and periodic

signals with no clear reference point in time depend on Interaural Phase Differences (IPD) for signal analysis.

The sound sources that deviate from 0º front-centre (as pictured in Figure 16, Figure 17) will also cause Interaural Intensity Differences (IID). But this occurs for frequencies with wavelengths smaller than the diameter of the head and begins in the range of frequencies greater than 1.2 kHz - 1.5 kHz. At these frequencies the head acts as an obstruction or baffle which attenuates the intensity level of sound arriving at the far ear. This attenuation increases with frequency—as the wavelengths decrease. For frequencies below 1 kHz, the wavelengths increase to the degree that they are able to bend or diffract around the obstructing head. This causes the IIDs to become less effective as the level differences at each ear decrease.

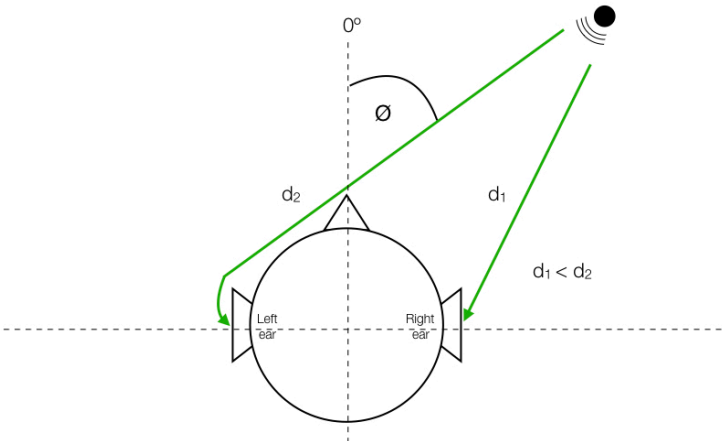


Figure 16: Interaural Time Delay (ITD).

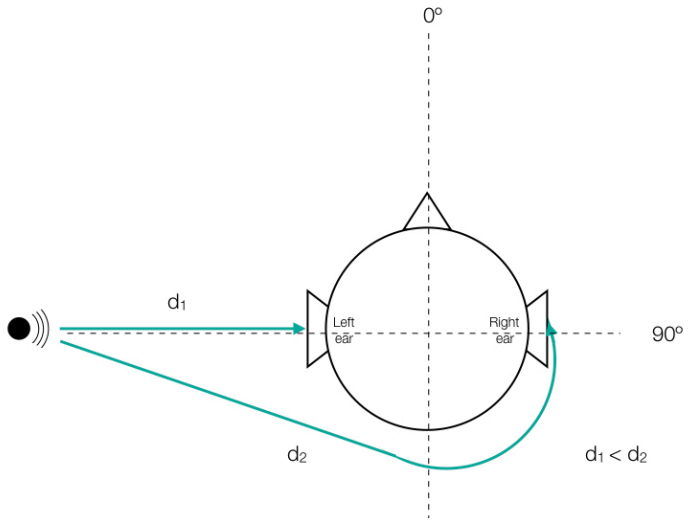


Figure 17: Maximum Interaural Time Delay (ITD).

From the above assertions this *Duplex Theory,* reported by Stevens in 1936 [82], described how each cue exhibited a frequency-dependent limitation where the range from 1 kHz - 1.5 kHz delineated the overlapping boundary of their effectiveness: below 1 kHz IIDs become less effective due to diffraction around the head—as the levels arriving at each ear become similar—and above 1.5 kHz the difference in distance to each ear from the source renders ITDs and therefore IPDs more effective.

2.3.3 Envelope Cues

Current researchers [42, 74, 81] have come to accept that the Duplex Theory does not exhibit strict frequency-bounded limitations on localization cues. It is now understood that the ITDs can be evaluated through envelop fluctuations at frequencies above the previous 1.5 kHz limit. The analysis of the fine-grained amplitude envelope provides information that helps to avoid phase ambiguities in the binaural ITD signals. Experiments by Steven van de Par and Armin

Kohlrausch [83] have shown that the human psychoacoustic apparatus is equally sensitive to ITDs across the entire frequency range when "transposed" signals are employed at high frequencies.

Begault [42] illustrates this process with the following description: "*A and B (*Figure 18*) show sine waves at the left and right ears that are below 800 Hz. Because the half period of the waves is larger than the size of the head, it is possible for the auditory system to detect the phase of these waveforms unambiguously, and the ITD cue can function. Above a critical point of about 1.6 kHz, sine waves become smaller than the size of the head, creating an ambiguous situation: the phase information in relationship to relative time of arrival at the ears can no longer convey which is the leading wavefront; i.e., whether D leads E, or E leads F in* Figure 18*. But if the sine waves are increased and decreased in amplitude (via a process known as amplitude modulation) then an amplitude envelope is imposed on the sine wave (see X and Y in* Figure 18*)."*

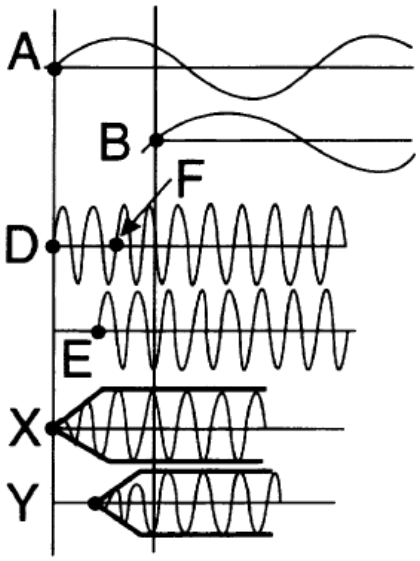


Figure 18: Transient Envelope Cue diagram from [42].

2.3.4 Median Plane Localization and Monaural Cues

When a sound source is in the median plane, the symmetry of the head renders the binaural cues (ITD, IID, IPD)—which are the main localization mechanisms in the horizontal plane—at a minimum because the signals arriving at each ear are identical. Therefore, due to head symmetry, all binaural cues would be zero and, thus, a listener should be unable to discern front and back sources. The resulting spatial ambiguity is referred to as the *Cone of Confusion* (Figure 19) as described by Wallach in 1939 [84]. This cone is the imaginary 'cone' extending outward from each ear along the interaural axis representing a sound source location that would produce identical interaural differences.

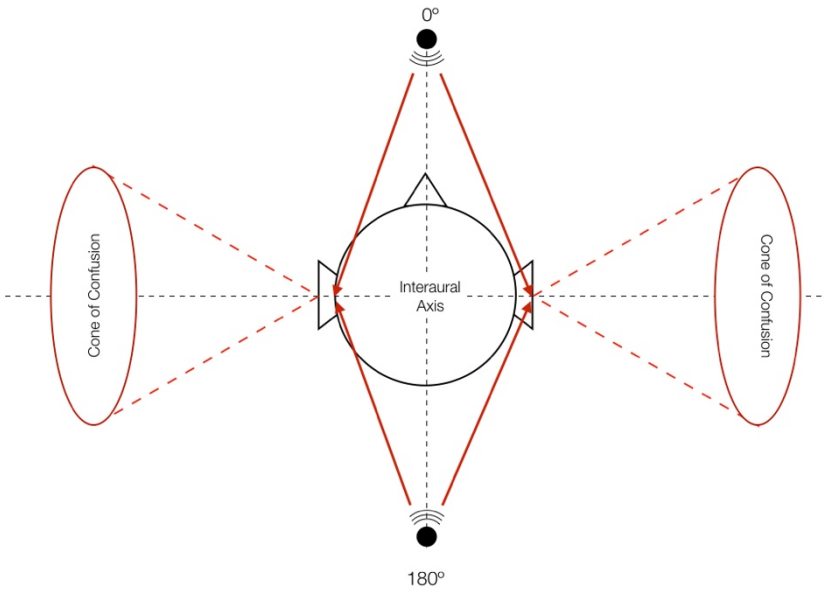


Figure 19: Cone of Confusion.

In reality, most people can localize sound sources in the median vertical plane. Steinhauser [85], Musicant and Butler [86], and Lopez-Poveda and Meddis [87] report that median localization is achieved primarily by *Monaural* cues which utilize spectral information associated with

asymmetry 1) in the head itself, 2) in the placement of the ears on the head, 3) the shape of each pinnae, and 4) the directional sound filtering of the torso and upper body. These monaural spectral cues are located in the frequency range of 4-16 kHz [88-90]. These asymmetrical cues produce peaks and troughs in the sound spectrum that are unique for each sound source's spatial location relative to the listener [91-93]. Roffler and Butler [94] state that the requirements for a sound source to be accurately localized in the median plane by a listener are: 1) The sound must be complex, 2) The complex sound must include frequencies above 7 kHz, and 3) The pinna must be present.

Conversely, lower frequency cues were reported by Algazi et al. [95] that are likely the result of sound reflecting off the torso and shoulders. And recently Lee [96] has theorized that low frequency spectral notches caused by torso reflections may be used in the creation of elevated phantom images for 3D audio reproduction systems.

Localization Blur also exists in the median plane. Blauert [73] reports that the localization blur in the forward direction for continuous speech by an unfamiliar person is on the order of 17º. While Damaske and Wagener [97] have found it to be minimally 9º for continuous speech by a familiar person (Figure 20), and 4º or white noise.

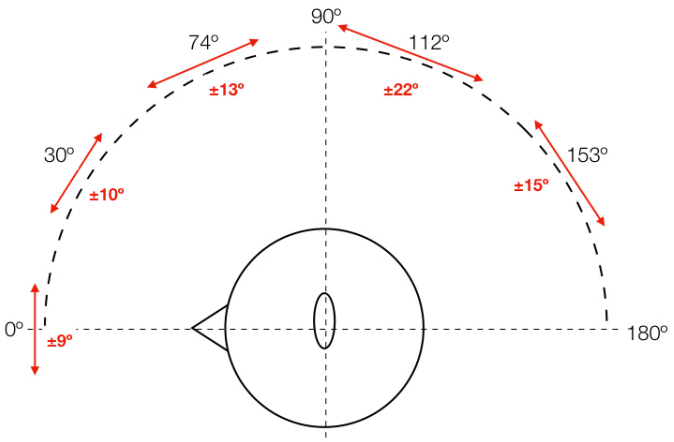


Figure 20: Localization Blur in the Median Plane for continuous speech by a familiar person [73, 97]

2.3.4.1 Directional Bands and the Perception of Height

Many researchers have found that narrow frequency bands in the high frequency region are associated with the perception of height [66]. Blauert [73] found that specific spectral bands are boosted and cut according to sound source location. He termed them "directional bands", and reports that frequencies surrounding the 8 kHz region are associated with source positions above the horizontal listening plane. Hebrank and Wright [88] also reported narrow bands corresponding to specific locations above the horizontal plane. These overhead positions were associated with a 1/4-octave peak between 7 kHz and 9 kHz. In 2015, Wallis and Lee [98] reported that "*the 1/3-octave band bursts tends to agree with Blauert's findings for 1 kHz, 4 kHz and 8 kHz*". These studies attest to the existence of directional bands in the 8 kHz region, which influence the perception of elevation.

2.3.4.2 Distance and Depth Perception

Distance describes the perceived expanse of space heard between a sound source and the listener, while *Depth* describes the perceived distribution, in the range from-front-to-back, of multiple sources presented in a sound scene. Depth can also be used to describe the depth of an

individual volumetric sound source. The extreme complexity of Distance and Depth Perception make its study much more difficult than that of Directional Hearing. The factors that contribute to distance perception include: 1) sound level decreases with distance, 2) the timbre of the sound source, 3) air absorption reduces high frequency content, 4) the perceived reverberation of the acoustic space, 5) also, the time interval between the direct sound and the first arriving reflections, 6) the attenuation and the delay of ground reflections, 7) the listener's familiarity with the sound source, and 8) the source directivity and angle of radiation, including amplitude envelope (attack and decay). Wenzel et al. have lately tackled this complex subject [99] as well as Begualt [42] and Blauert [73].

2.4 Stereo Recording

"Stereo is merely an attempt to create the illusion of reality through the willing suspension of disbelief" (Richard Heyser) [100].

Michael Williams [101] describes the situation more concretely: *"The number of different microphone systems available for stereophonic sound recording is very limited and almost without exception these systems have fixed characteristics . . . Each system has been developed to be "optimum" in a given set of circumstances; however, as recording conditions are infinitely variable, this optimum is rarely achieved . . . Microphone position is generally a compromise between a good coherent stereophonic image and the required ratio of direct to reverberant sound."*

Franssen, in his book *Stereophony* [102]—and by extension Eargle [103]—details the distortion of the stereo image of various stereo microphone arrays when played back via two loudspeakers. Which would seem to bring us back to Heyser: *stereo is the illusion of reality*.

The stereo recording concepts that dominate most conventional techniques were developed in the 1930s. In England Alan Blumlein [41] developed coincident-pair configurations while in Murray Hill, New Jersey, Harvey Fletcher, as Director of Physical Research at Bell Laboratories—and aided by the likes of William B. Snow and J.C. Steinberg—focused on spaced pairs, and two- and three- channel techniques [54, 100, 104]. The vertical-imaging arrays that were developed over the course of the current research are a direct descendent of these investigations, especially the so-called "the curtain of sound" consisting of a curtain of microphones for the capture of sound in one location coupled to a corresponding curtain of loudspeakers to listen to in another location [105].

2.4.1 Useful Angle of Acceptance

Stereophonic microphone systems have an angle within which the sound sources must be located to be accurately reproduced within the stereo auditory image when played back via stereo loudspeakers. This is referred to the *Useful Angle of Acceptance*, the *Recording Angle*, or the *Useful Acceptance Area* (Figure 21) [106]. Sound sources located outside the *useful angle of acceptance* will be reproduced directly at the left or right loudspeaker. In this case, the stereo image will not represent the location of the sources in the recording situation, thus providing a distorted representation of the stereo perspective (Figure 22).

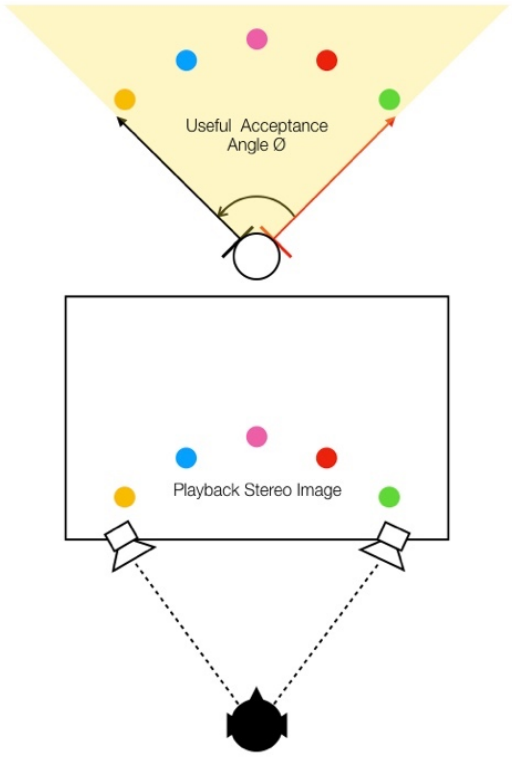


Figure 21: Stereo playback imaging of sound sources located within the Angle of Acceptance.

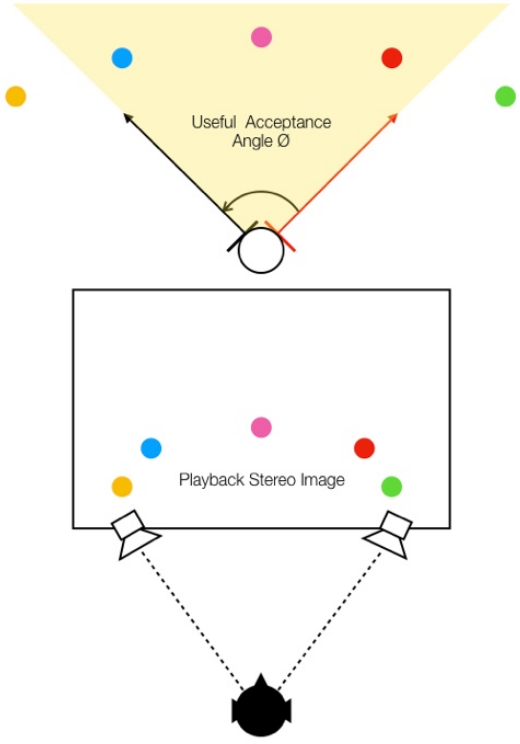


Figure 22: Stereo playback imaging of sound sources located outside the Angle of Acceptance.

Stereo microphone configurations create spatial images through either intensity differences, time differences or both. Stereo microphone systems fall under three categories:

1. Coincident arrays which are based on the principle of intensity differences.
2. Near-Coincident arrays which utilize both time and intensity differences.
3. Spaced-apart techniques which utilize both time and intensity differences.

2.4.2 Coincident Microphone Techniques

In a coincident microphone pair, directional microphone capsules are placed as close together as possible to minimize time-of-arrival differences and are both adjustable in their lateral pickup angles. Because the input to each microphone differs only in intensity—which is determined by the direction of the arrival of the sound—coincident microphone techniques are often termed *Intensity Stereo* and are also referred to as *XY* arrays.

2.4.2.1 XY Stereo Techniques

In XY stereo arrays, two directional microphone capsules are arranged one on top of the other, and oriented so as to point to the left and right of the sound stage in such a manner that the coincident capsules are set at an included angle Ø (Figure 23). The angle Ø generally varies between 80º and 135º [103]. As a sound source deviates from the center position, the difference in *intensity* between sound arriving at each of the two microphones will increase, and the corresponding phantom image of the source will be localized in the position corresponding to the sound intensity difference when played back on loudspeakers. Two of the most common configurations consist of either cardioid or bi-directional/figure-8 microphones.

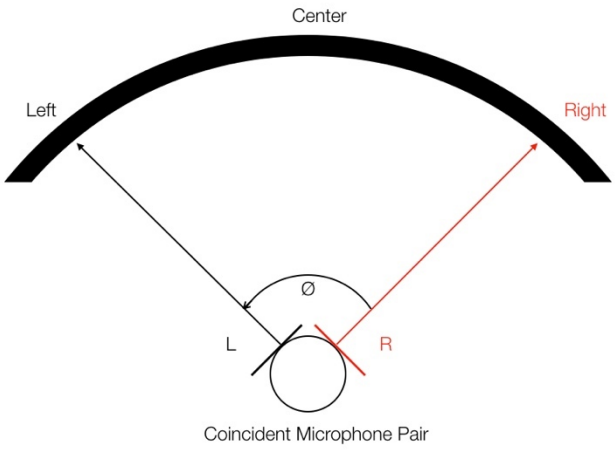


Figure 23: XY Coincident Microphone Pair.

2.4.2.2 Crossed Cardioid Arrays

Crossed cardioid arrays generally have an included angle between 90º and 135º. These arrays exhibit excellent monophonic compatibility. When set at an included angle of 90º, the array produces a centre-dominant image which many engineers do not find optimal. A wider stereo image is easily attained by increasing this angle. The use of super- and hyper-cardioid microphones may provide a good compromise between a wide stage pickup and direct-reverb ratio [103].

2.4.2.3 Blumlein Array

This array, consisting of coincident figure-8/bi-directional microphones angled at 90º, was proposed by Alan Blumlein in 1931 (Figure 24). The front quadrant pickup maps sound sources with a high degree of accuracy within the included angle. The rear pickup of the system is in reverse polarity to the front. The array also provides an excellent sense of the acoustic space due to the rear quadrant pickup of the figure-8 microphones, and a balanced distribution of the diffused field. The only drawback is when the front-quadrant sound sources are located beyond the 90º pickup angle. In this instance the original stereo image is distorted in the same manner illustrated in Figure 22.

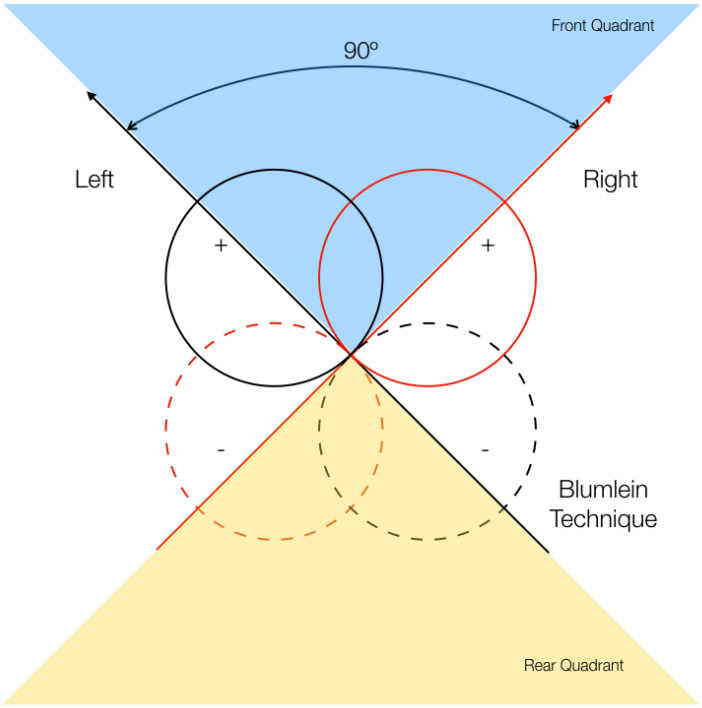


Figure 24: Blumlein Array.

2.4.2.4 Middle-Side (MS) Technique

The MS stereo technique was proposed by Alan Blumlein in his 1934 patent, but not realized until Danish Radio engineer, Holger Lauridsen put it into practice in the 1950s [100].

The *Middle* signal is a discrete monophonic pickup provided by a single microphone aimed along the centre 0º axis. The polar pattern of this microphone can be cardioid, bi-directional or omnidirectional. The *Side* information is provided by a bi-directional microphone oriented 90º to the *Middle* microphone main axis with the positive side of the capsule positioned to the left. The two microphones are arranged one on top of the other—as in all coincident techniques—to minimize the acoustic time-of-arrival differences. Because the *Side* bi-directional microphone is aimed left-to-right, with its null side on the 0º axis, the information provided is primarily ambient, with little direct sound pickup (Figure 25).

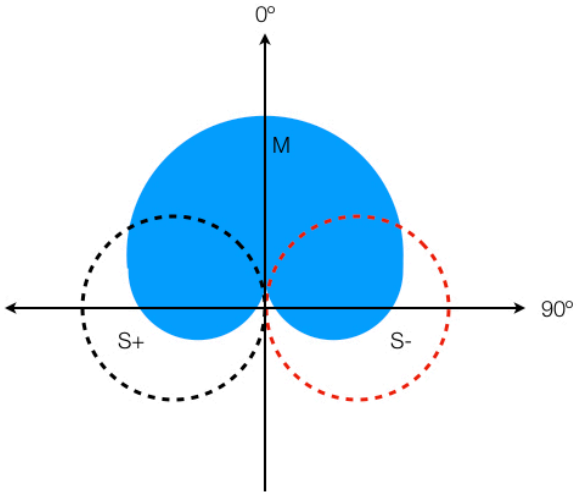


Figure 25: MS Stereo basic microphone configuration.

Deriving a stereo signal from the MS array is achieved by using a passive or active matrix (Figure 26). The *Middle* and *Side* signals are combined to derive two new signals: Mid + Side and Mid – Side. The centre monophonic signal is obtained by summing (M+S) + (M-S) = 2M. The left and right *Side* signals are obtained by (Figure 27):

1. Splitting the *Side* microphone signal into two identical discrete channels.
2. Routing the left (S+) to the left output.
3. Routing the right (S-) to the right output (Side signal in opposite polarity).

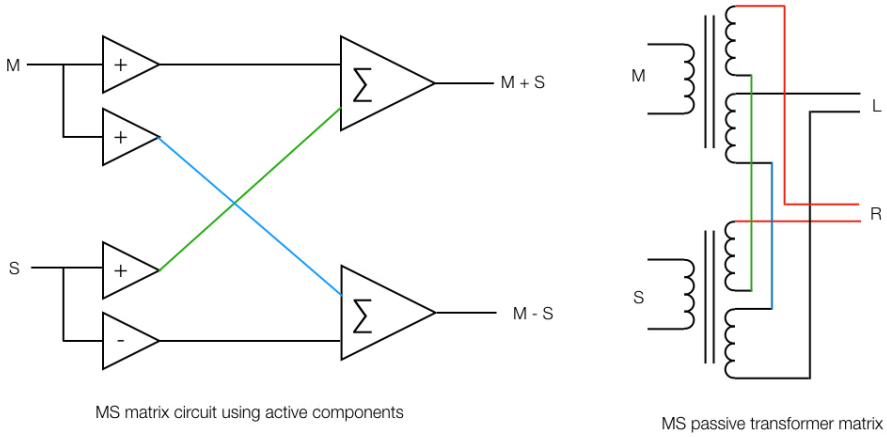


Figure 26: MS Matrix Circuits.

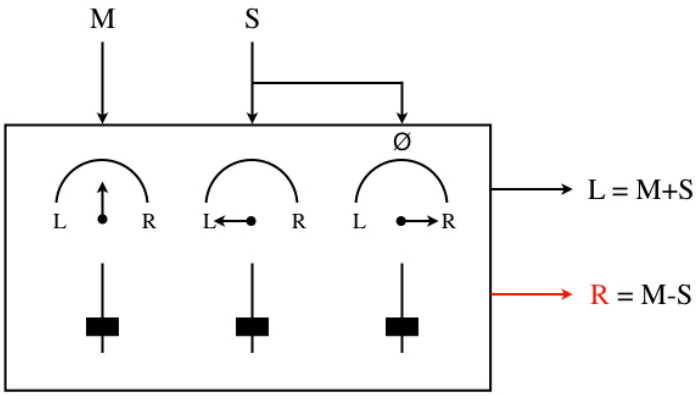


Figure 27: MS matrix circuit at recording console.

2.4.3 Near-Coincident Microphone Techniques

Near-Coincident microphone techniques employ a pair of directional microphones, which are closely spaced but splayed outward. They rely largely on intensity differences to provide the stereophonic information, but also on time differences. The spacing between the microphones in these arrays is on the order of 17-50 cm, and the included angle between the microphones ranges from 90º to 135º. These distances introduce phase differences. When compared to coincident arrays, near-coincident techniques have been described (due to these phase anomalies) as having an increased sense of 'space', more 'air' and being more 'open' sounding. However, the introduced phase differences can have a significant impact on the sound quality [100, 106, 107].

There are many examples of 'named' pairs, some of which originated with various European broadcasters, (Figure 28 displays a comparison of near-coincident microphone arrays):

1. ORTF technique: developed by the French broadcasting organization *Office de Radiodiffusion-Television Francaise*. Configured of two cardioid microphones 17 cm apart, angled at 110º.
2. DIN technique: specified by the German national standards organization *Deutsches Institut für Normung*. Configured of two cardioid microphones spaced 20 cm apart,

angled at 90º.

3. RAI technique: developed by the Italian broadcasting agency RAI, *Radiotelevisione Italiana* (known until 1954 as *Radio Audizioni Italiane*). Configured of two cardioid microphones 21 cm apart, angled at 100º.
4. NOS technique: developed by the Dutch Broadcasting Foundation *Nederlandse Omroep Stichting*. Configured of two cardioid microphones 30 cm apart, angled at 90º.
5. Olson technique: developed by Lynn T. Olson (of Audionics, Inc.) with the goal of creating a 180-degree sound field. Configured as two hyper-cardioid microphones 5 cm apart, angled at 135º.

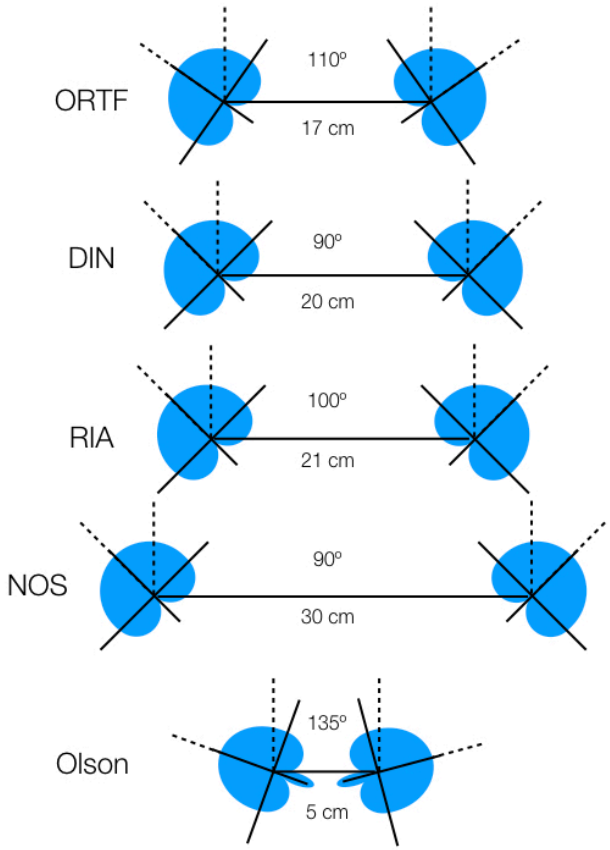


Figure 28: Comparison of near-coincident microphone arrays.

2.4.4 Spaced Microphone Techniques

Widely spaced microphone arrays were first reported by Clement Adler at the 1881 Paris Exhibition, and later provided the basis for the stereo systems investigated by Bell Laboratories in the 1930s [108]. These techniques usually employ two or more omnidirectional microphones (but cardioids can also be used) spaced apart at a distance and are commonly referred to as an *AB* pair.

In spaced arrays both time and intensity differences play a role in the creation of the stereo image during playback with the precedence effect being a primary factor. When the microphones are widely spaced, the delay between channels will be on the order of a number of milliseconds for sound sources located at the extreme left and right side of the soundstage.

This system is most commonly used in the recording of classical music ensembles [109]. AB pairs have been criticized for their lack of consistent monophonic compatibility, and lack of accuracy in the stereo image during playback due to lack of phase coherence including phase inversion at low frequencies [100], but *"not always as poor in practice as might be expected"* [108]. Richard King, in his book *Recording Orchestra and Other Classical Music Ensembles,* states: "*A properly placed AB* (pair) *can provide great impact, width, depth, low frequency response, and with the right microphones, incredible clarity and realism."*

2.5 3D Audio Recording Techniques

Many authors have proposed specific microphone arrays for three-dimensional music recording, and the great majority are concerned with the capture of acoustic ensembles in a performance space. A comprehensive summary of immersive microphone arrays was reported by Sungyoung Kim in the book *Immersive Sound: The Art and Science of Binaural and Multi-Channel Audio* [66], and William Howe et al [110].

2.5.1 William Howie's Orchestral Recording Technique Optimized for the 22.2 Multi-channel Sound System.

William Howie [111], building on the work of Kimeo Hamasaki [1], has developed a novel technique for orchestral music capture for three-dimensional audio reproduction optimized for Japanese broadcaster NHK's 22.2 Multi-channel Sound System. This array employs omnidirectional microphones for the main orchestral sound capture, directional microphones to capture floor reflections and vertical orchestral imaging, and an ambience array designed to capture many points of decorrelated reflected sound energy (Figure 29).

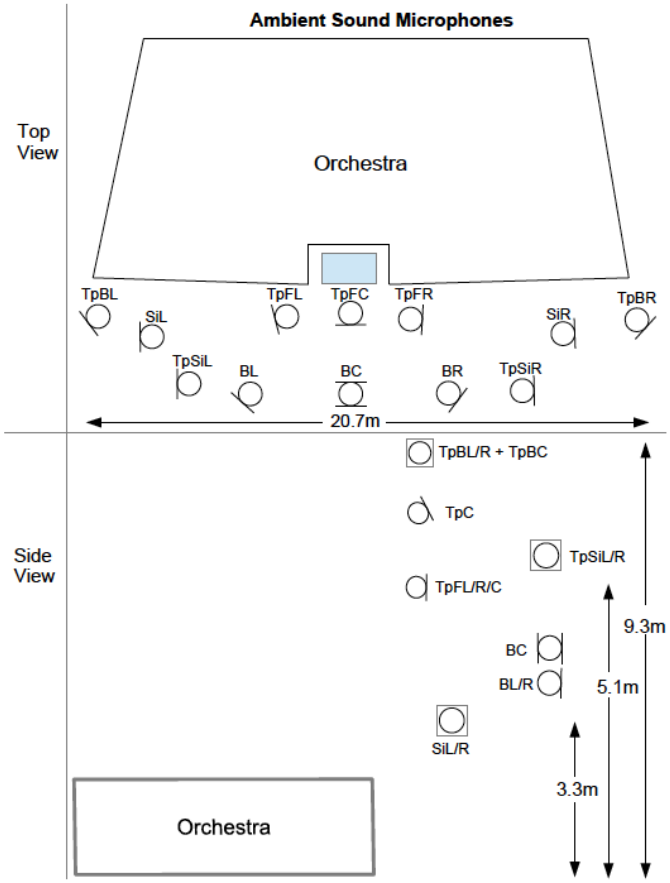


Figure 29: William Howie's Recording Technique Optimized for 22.2 Multi-channel Sound [111], (Used by permission).

2.5.2 Michael William's Arrays for Rendering 3D Sound Fields

Michael Williams has long been a researcher of microphone arrays for multi-channel sound recording. In 2012 [112] he devised a two-layer array for capturing vertical information based on his surround-capture array known as the 'Williams's cross'—a cross of four hyper-cardioids spaced 35 cm apart at angles of 45º, 135º, 225º and 335º. To integrate height information into this array he added an upper layer of three bi-directional microphones at angles of 0º, 90º and 270º pointing upward. This 7-channel array was named the M.A.G.I.C (Multi-channel Arrays Generating Inter-format Compatibility) array.

In 2013 [24] Williams proposed a 12-channel microphone array for the complete rendering of 3D sound fields (Figure 30). This array consists of a middle layer of the 'Williams's cross' combined with an optional four satellite microphones angled at 0º, 90º 180º and 270º spaced at 2.5 m; and an upper layer of four upward-facing bidirectional microphones angled at 0º, 90º 180º and 270º spaced at 1.5 m.

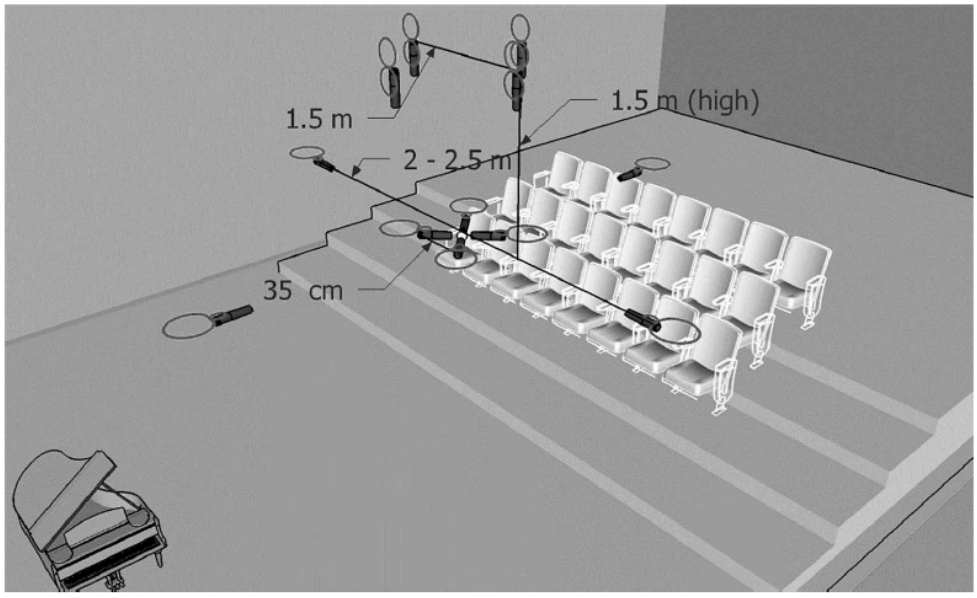


Figure 30: Williams' 12-channel M.A.G.I.C. array [66]. (Used by permission).

2.5.3 Günther Theile's OCT-9 Array

Günther Theile OCT-9 array is based on his Optimized Cardioid Triangle (OCT) [113] array which was designed to exploit the L-C-R frontal image of a three-channel stereo playback system. The array was designed to minimize inter-channel crosstalk and provide directional stability without decreasing the stereophonic quality. The OCT array consists of a 0º cardioid, and left and right microphones comprised of super-cardioids, angled at 90° and 270º spaced 8 cm to either side of the center microphone.

The original OCT array was expanded to a five-channel configuration (OCT-surround) which included two rear facing cardioids. This iteration was further expanded to the OPT-9 array for immersive, three-dimensional sound capture by the addition of four super-cardioids pointing upwards [68], and located 1m above the middle layer (Figure 31). This array was largely developed as the *'Auro-3D Main Microphone'* to be played back via the Auro-3D 9.1 loudspeaker configuration.

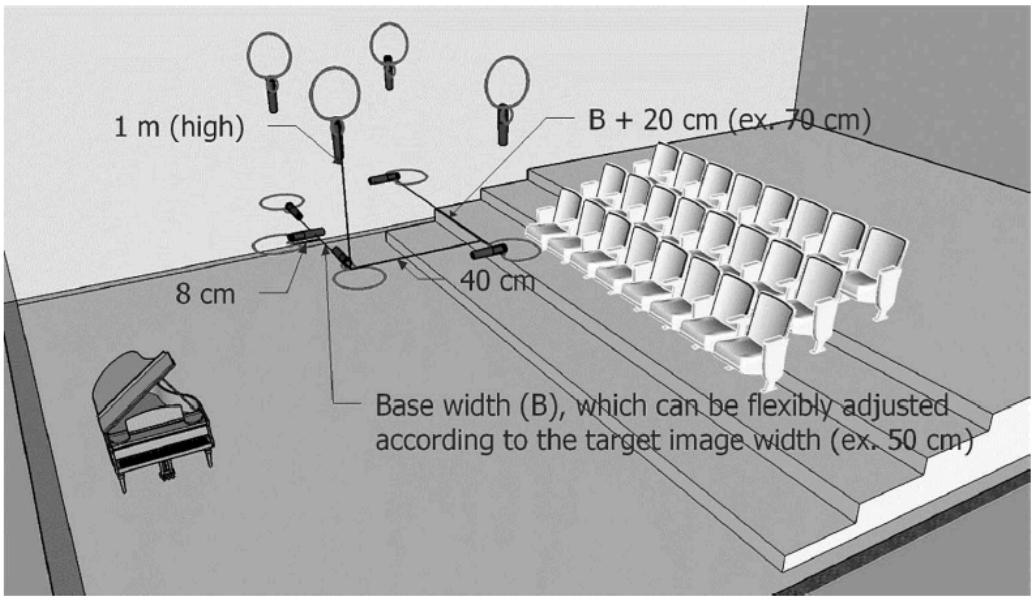


Figure 31: Günther Theile OCT-9 array [66].(Used by permission).

2.5.4 Morten Lindberg's 2L-Cube Array

Morten Lindberg, of Norway's *2L-the Nordic Sound* record label, began recording with 5.1 surround arrays in 2004. In 2008—in a collaboration with Wilfried van Baelen of Auro-3D—he began developing an immersive/3D microphone array optimized for playback in the 9.1 Auro-3D loudspeaker configuration [114].

The result was the development of the 2L-cube array (Figure 32), which is a descendent of the Decca-[1] and Mercury-tree arrays[2]. The goal of the configuration is that the 3D audio image (presented to the listener during playback in the Auro-3D 9.1 configuration) is created during the recording process with dedicated microphone techniques, not in the mixing process. Lindberg states: "*The composers and musicians should perform to the extended multi-dimensional sonic sculpture, allowing more details and broader strokes. Then immersive audio and surround sound is just a matter of opening up the faders."*

The array was designed to be used in large acoustic venues such as churches, cathedrals and large concert halls. The array is comprised of eight omni-directional microphones with the dimensions of the 'cube' varying from 150 cm (for orchestra) to 40 cm (for chamber music ensembles). Lindberg prefers larger diaphragm microphones which provide "*a more focused on-axis texture of sonic image*" [66].

[1] The stereo microphone array commonly referred to as the "Decca Tree" was originally conceived by the recording engineers at English Decca Records. The configu- ration quickly became the standard among the classical recording community.It utilized three omnidirectional microphones situated at the ends of a large T-shaped fixture. The spacing between the left and right microphones was approximately 2 meters, and the central microphone was in front of these by about 1.5 meters. Placement of the array was generally a few feet behind and about eight to ten feet above the conductor's head.

[2] The Mercury Living Presence recording technique, associated with C. Robert Fine, dates from the late 1940s (the first official Mercury Living Presence recording was made in April, 1951) and started out as a single-microphone method to make a full- range monophonic recording of a symphony orchestra. In 1955, the Mercury team decided to record in 3- channel stereo, feeding each track on a 1/2" tape directly from left, center and right microphones.

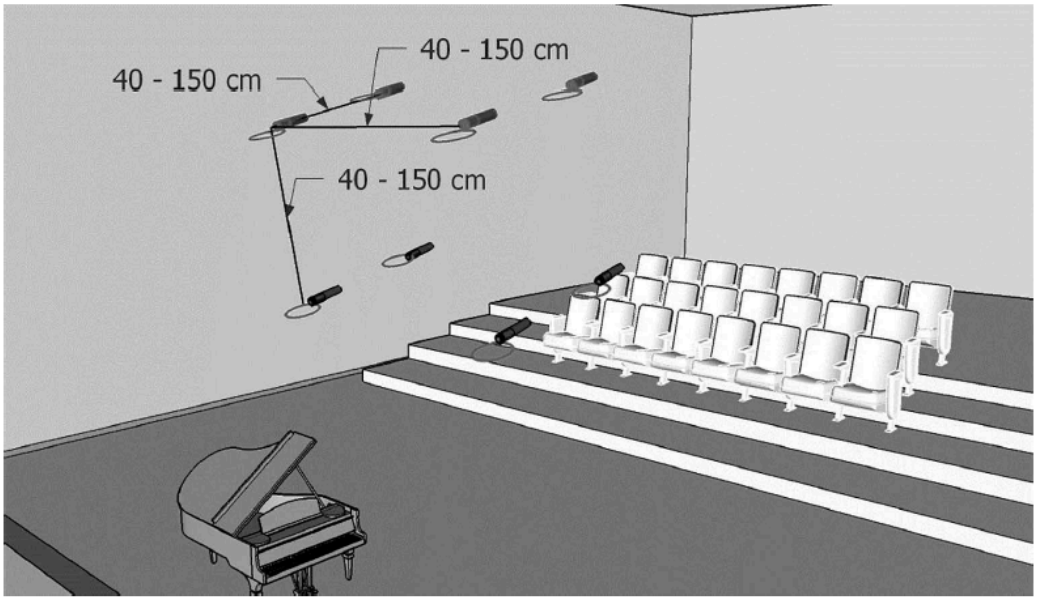


Figure 32: Morten Lindberg's 2L-Cube [66]. (Used by permission).

2.5.5 Gregor Zielinsky's Twins Cube and Twins Square Arrays

The Zielinsky arrays are based on the spaced AB pair. These arrays have the ability to capture both horizontal and vertical audio imaging. The arrays are comprised of Sennheiser 800 Twin microphones. The MKH 800 Twin is a condenser microphone containing two opposite facing, cardioid capsules. The dual signals from each transducer pair are available as discrete outputs so the directional characteristics can be determined remotely in-situ, or during post production [115].

2.5.5.1 Twins Square Array

The Twins Square is comprised of four MKH 800 Twin microphones. With each device containing two transducers, this array has the ability generate eight outputs: left, right, left-surround, right-surround, upper-left, upper-right, upper-left-surround, and upper-right-surround.

2.5.5.2 Twins Cube Array

The Twins Cube array (Figure 33) expands the Twins Square by placing a second Twins Square array behind the first—forming a cube. The Twins Cube array is constructed to mirror the 9.1 loudspeaker configuration. Zielinsky states: "*Via the cube, the signal positions itself. Every signal is reproduced by at least three, four or even more loudspeakers, and thus you can clearly hear where the signal is coming from* [116]".

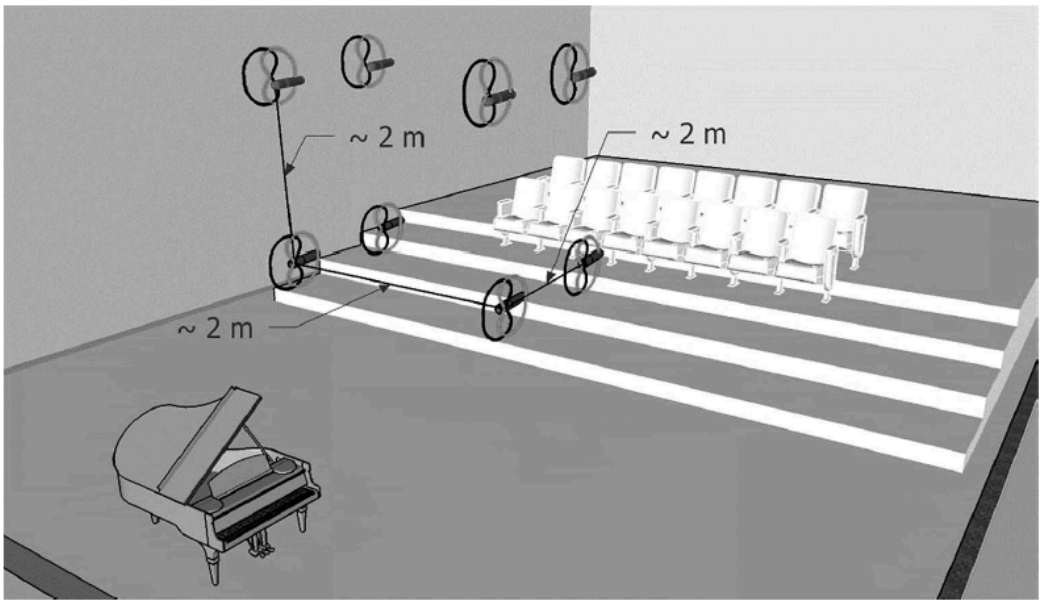


Figure 33: Gregor Zielinsky's Twin Cube array [66]. (Used by permission).

2.5.6 Helmut Wittek and Günter Theile's ORTF-3D

The ORTF-3D system was developed by Helmut Wittek and Günter Theile for Schoeps GmbH [117-119]. The array is comprised of eight super-cardioids: the first 10 x 20 cm rectangular four-microphone array forms the middle layer, and a second 10 x 20 cm rectangular four-microphone array forms the upper layer. The two rectangular arrangements are placed one on top of the other, (Figure 34). Vertical imaging is achieved by angling each layer's microphones to form vertical X/Y pairs at 90º. The middle four channels are generally routed to the left, right, left-

surround and right-surround loudspeaker positions. The upper four channels are generally routed to height loudspeaker positions: left high, right high, left-surround high and right-surround high.

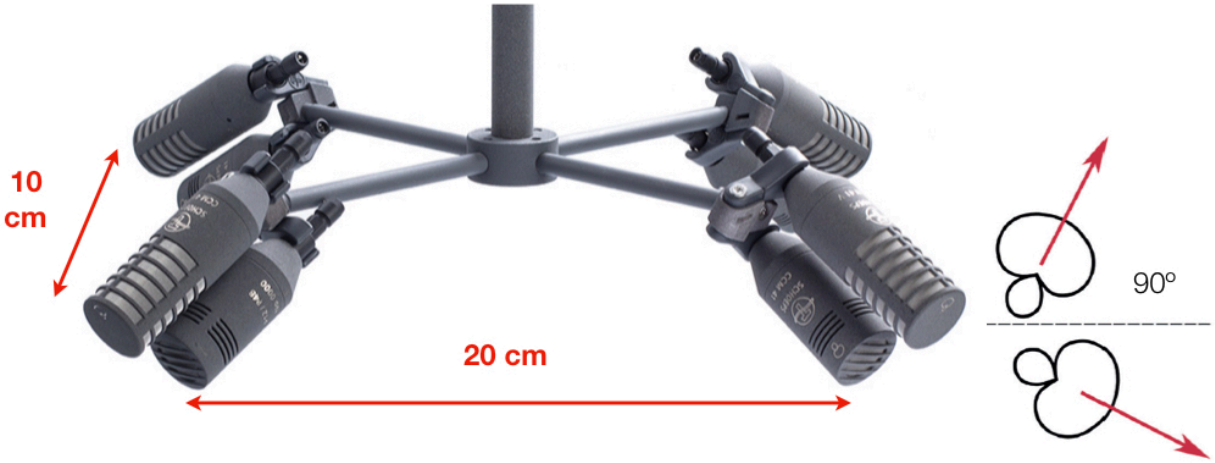


Figure 34: ORTF-3D. (Used by permission).

2.5.7 David Bowles' Microphone Array

The Bowles Microphone Array [66, 120] is comprised of a horizontal (middle) layer containing a single center-front directional microphone and four surround omnidirectional microphones, and an upper layer of four super-cardioid microphones angled upward at 30º (Figure 35). The aim of the upper layer is to capture ceiling and upper sidewall reflections. The spacing between the layers, and between the microphones within each layer, may vary. This is largely dependent upon the characteristics of the acoustic environment and the size of the ensemble. Additional microphones may be added for coverage of larger ensembles.

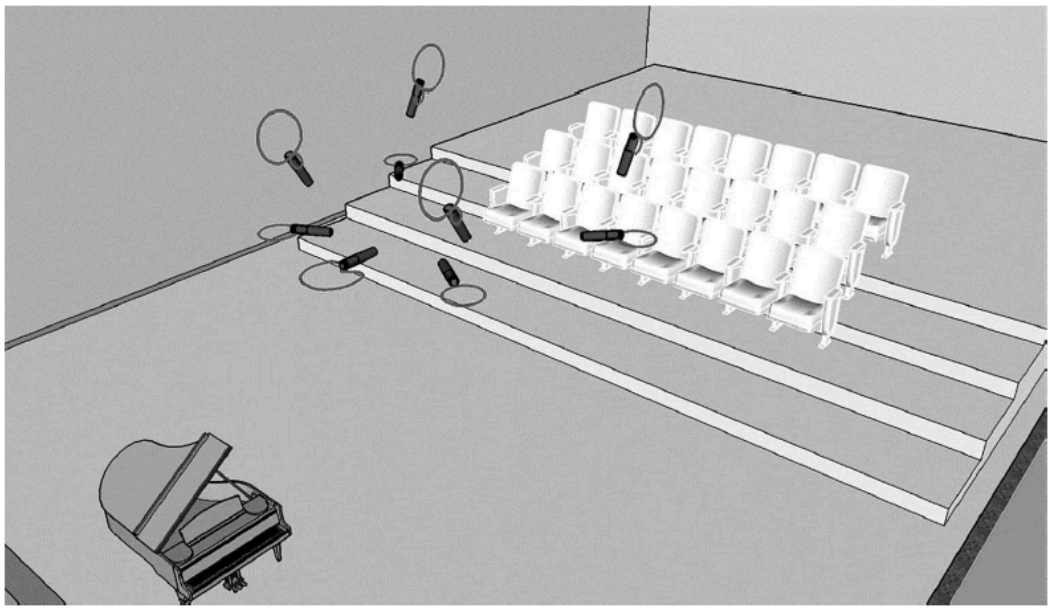


Figure 35: Bowles Microphone Array [66]. (Used by permission).

2.5.8 Paul Geluso's Z-Microphone Array

Paul Geluso's Z-Microphone Array [121] adds a vertically oriented bi-directional 'Z' microphone to a horizontally oriented traditional MS pair. The MZ pair can be decoded to stereo by the same MS matrixing described in section 2.4.2.4. The conventional MS pair will yield (after matrixing): center, M+S left, and M-S right. The MZ pair will yield (after matrixing): M+Z height at +45º, and M-Z lower level information at -45º. Due to the possibility that the Z microphone can be paired with any microphone within a stereo or surround array, the MZ technique is easily adapted to conventional horizontal configurations as shown in Figure 36.

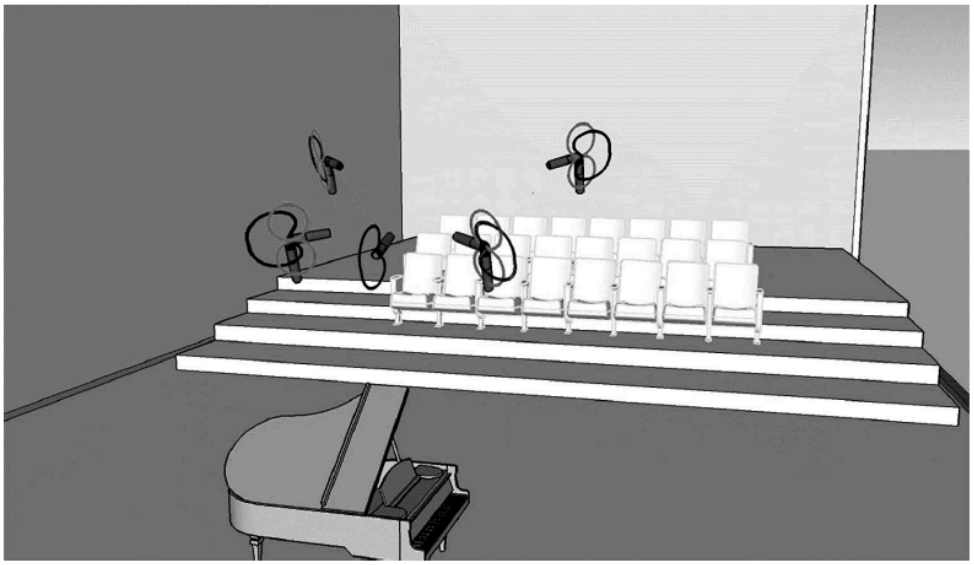


Figure 36: Paul Geluso's space MZ array incorporating four vertically oriented bidirectional microphones paired with horizontally oriented cardioid microphones [66]. (Used by permission).

2.5.9 NHK Portable Spherical Microphone Array for Super Hi-Vision 22.2 Multi-channel Audio

The NHK proposed a portable spherical microphone array with the capability of simultaneously recording 24 channels of audio. The unit consists of a spherical structure (45 cm in diameter) containing 24 omni-directional microphones. The sphere is partitioned into 24 angularly segmented sections: eight within each layer with three vertical layers (Figure 37). This corresponds to the three layers of the NHK 22.2 Multi-channel Sound System. The aim of the acoustic baffles between the sections is to achieve "*a constant and narrow beam width [122]*".

Figure 37: Proposed NHK Spherical Microphone [122]. (Used by permission).

2.5.10 Richard King's Omni-directional Height Array with Diffraction Attachments

Richard Kings technique [123] is comprised of omni-directional microphones configured in conventional 5.1 surround arrays with additional microphones to capture height/immersive information. The height-capture microphones are omni-directional with diffraction attachments (Figure 38) aimed away from the sound source. These capsules are oriented to achieve sufficient decorrelation from the direct pickup to ensure the ensemble image is not distorted in the vertical plane.

Figure 38: Omnidirectional microphone with diffraction attachment [123].(Used by permission).

2.5.11 Wieslaw Woszczyk's and Paul Geluso's 3D Sound Field Array

The 3D sound field array proposed by Woszczyk and Geluso [124] is comprised of two or more tetrahedral microphones, (Figure 39). Minimally, two tetrahedral units (left and right) are required to deliver overlapping three-dimensional sound fields. Together they create the impression of a coherent 360º immersive presentation. A multi-channel, immersive monitor configuration is unnecessary during recording due to the well-defined characteristics of a calibrated tetrahedral capsule. Therefore, headphones are sufficient to judge the spatial aspects and sound quality of the microphones in-situ.

Figure 39: Tetrahedral Microphone Capsule.

2.6 Subjective Assessment of Spatial Sound Quality

Testing for perceptual sound quality has been a common practice for decades pertaining to loudspeakers, headphones, audio and hi-fi equipment as well as performance spaces and concert halls. Prior to the 1980s many of these tests were lacking in the necessary controls and rigor to extract reliable and significant results. Floyd Toole—and later with Sean Olive—working at the National Research Council, Ottawa, Canada, (and moving to Harmon International in the 1990s), contributed much research in the areas of standardization of listening tests, subjective evaluation, the selection and training of test subjects, defining the testing environments and experimental procedures [125-129].

A comprehensive overview of perceptual testing can be found in Søren Bech's and Nick Zacharov's *Perceptual Audio Evaluation–Theory, Method and Application*. Its aim is to guide the researcher through the complete perceptual testing process. The topics covered include experimental considerations, variables and statistics; the technical aspects of the listening environments such as the placement of both the loudspeakers and the listener/test subject within

the testing environment; the training and selection of subjects; the planning, administration and reporting of testing, and test standards and recommendations [130].

The ITU-R BS.1116-1 is another source of comprehensive guidelines for subjective testing. The recommendation outlines the experimental design where subjective tests are conducted using sound systems utilizing a selected group of trained subjects. The recommendation describes a "double-blind triple-stimulus with hidden reference [131]" method also referred to as a 'Triad' test. In this method a test subject is asked to compare three randomly assigned stimuli, "A", "B", "C", and decide which two are the same. The ITU-R BS.1116-1 also recommends approaches for the selection of listening panels, familiarization or training of test subjects, post-screening of subjects, attributes of playback systems, program material, reproduction devices, listening conditions, statistical analysis, and the presentation of the results of the statistical analyses.

2.6.1 Spatial Audio Assessment

Many researchers have contributed to the subjective testing and description of immersive and three-dimensional audio quality, but as research accelerated in the 21st Century there accrued a lack of uniformity, clarity and agreement on the meaning and weighting of spatial attributes found in the literature [132, 133].

Francis Rumsey has conducted many investigations into the evaluation, description and qualification of spatial sound quality [134, 135]. Together with Jan Berg investigations were undertaken to identify, verify and correlate spatial attributes [136, 137]. In further research they proposed a systematic approach to provide statistically significant results in evaluation of different modes of spatial reproduction and different microphone techniques [138].

Le Bagousse et al. [139]—in an effort to provide a more uniform weight and meaning to spatial attributes used in previous research—generated three ‘sound families’ from the results of two semantic tests. Their study strove to provide clarification to the meanings of spatial audio descriptors in an effort to minimize biases in listening tests. In further research, the team of Bagousse, Paquier and Colomes created a ‘lexicon’ of audio quality assessment attributes [133]. The goal of their research was aimed at understanding whether the assessment of spatial technologies (multi-channel codecs, microphone arrays, and immersive audio reproducing systems) is related to an attribute of the spatial reproduction or to other attributes of sound.

Others who have contributed to the define, evaluate and quantify spatial audio attributes are Choisel and Wickelmaier [140-142], Kamekawa and Marui [27], Liebetrau et al. [143], Darcy [144], Zacharov and Pederson [145], and Zacharov and Koivuniemi [146]. Letowski [147] and Lokki [148] produced comprehensive circular layouts of attributes of audio quality showing how they relate to each other.

2.6.1.1 Nonverbal Methods of Spatial Audio Assessment

Nonverbal methods for assessing spatial perception have generated much research in the past 20 years. Before than, in 1960s and 1970s, Damaske and Wagener [97] used graphical methods to show the perceived location and distribution of sound presented to listeners. Scene based discussions for describing audio program have long been used by industry professionals, audiophiles and audio enthusiasts dating to the very beginnings of reproduced sound. Mason, Ford, Rumsey and de Bruyn [149] report that nonverbal elicitation methods may be preferable to verbal elicitation for communicating some attributes of auditory events. Ford and Rumsey furthered this research by examining graphical elicitation techniques for the assessment of spatial audio [150, 151]. They reported (for conventional two-channel playback configurations), that a graphical

elicitation technique is intuitive, and provides useful information regarding the influence of loudspeaker location, listener location and sound source location in the context of the spatial representation of complex stereo sources.

Usher and Woszczyk [152, 153] advanced this research to the depiction of spatial information beyond two-channel playback. They investigated a graphical user interface (GUI) for depicting the spatial audio image when played back via a multi-channel (5.0) loudspeaker configuration. The interface allowed for the elicitation of differences of the sound sources within multi-channel recordings of varying music ensembles.

3 EXPERIMENTS IN IMMERSIVE MIXING

This chapter is comprised of three publications that detail experiments in three-dimensional music mixing and height-channel utilization in recording of music. The first two papers discuss immersive mixing and three-dimensional image creation methodologies developed for the 22.2 Multi-channel Sound systems. The first paper discusses the techniques for expansion of the size of sonic objects in two- and three-dimensional planes, while the second paper is a study to establish effective levels of height-channel information based on the results of a listening test. The third study examines the design and implementation of early and late reflections, and of reverberant fields, in mixing.

It should be noted that at the time these studies were undertaken (2015), three-dimensional music mixing (and recording) of popular music was in its infancy. Little or no 3D software existed to aid in these endeavors, nor were many studies published on these topics. This is reflected in the conservative language found in these sections regarding the progress of research into three-dimensional mixing and recording.

3.1 Schulich School of Music Studio 22

This research was conducted at McGill University, in the Schulich School of Music's Studio 22 (Figure 40). It is an audio research laboratory designed for the production and assessment of multi-channel audio. The acoustic design is by Ben Kok of Nelissen Ingeneursbureau, Netherlands.

Studio 22 is optimized for multi-channel recording and playback, and is configured according to the ITU-R BS.2051-0 [62] recommendations for the 22.2 channel sound system detailed in section 2.2.7.6, and optimized for multi-channel recording and playback and is compliant to ITU-R BS.11161 [154].

Figure 40: Studio 22.

3.1.1 Music Studio 22 Properties

Studio 22 is a music mixing control room with an RT30 from 137-223ms (Figure 41). It has a full-range playback system comprised of two-way loudspeakers and displays a relatively flat response between 20 Hz and 18kHz. This studio is optimized for multi-channel recording and

playback, with up to 30 discrete channels and loudspeakers available in the control room infrastructure (Table 1).

Channel Number	Channel Name	Label	Azimuth	Elevation	Distance
1	Front Left	FL	60°	0°	2.1
2	Front Right	FR	-60°	0°	2.1
3	Front Centre	FC	0°	0°	2.05
4	Low Frequency Effect 1	LFE1	50°	-15°	2.35
5	Back Left	BL	135°	0°	2.1
6	Back Right	BR	-135°	0°	2.1
7	Front Left centre	FLc	30°	0°	2.1
8	Front Right centre	FRc	-30°	0°	2.1
9	Back Centre	BC	180°	0°	1.85
10	Low Frequency Effect 2	LFE2	-50°	-15°	2.35
11	Side Left	SiL	90°	0°	2.1
12	Side Right	SiR	-90°	0°	2.1
13	Top Front Left	TpFL	60°	35°	2.6
14	Top Front Right	TpFR	-60°	35°	2.6
15	Top Front Centre	TpFC	0°	35°	2.6
16	Top Centre	TpC	0	90°	2.1
17	Top Back Left	TpBL	135°	35°	2.6
18	Top Back Right	TpBR	-135°	35°	2.6
19	Top Side Left	TpSiL	90°	35°	2.6
20	Top Side Right	TpSiR	-90°	35°	2.6
21	Top Back Centre	TpBC	180°	35°	2.45
22	Bottom Front Centre	BtFC	0°	-20°	2.2
23	Bottom Front Left	BtFL	45°	-20°	2.2
24	Bottom Front Right	BtFR	-45°	-20°	2.2
25	Auro Top Front Left	U+030	30°	35°	2.6
26	Auro TopFrontRight	U-030	-30°	35°	2.6
27	Auro Top Back Left	U+110	110°	35°	2.6
28	AuroTopBack Right	U-110	-110°	35°	2.6

29	Auro Back Left	M+110	110°	0°	2.1
30	Auro Back Right	M-110	_-110°	0°	2.1

Table 1: Studio 22 loudspeaker layout and naming conventions.

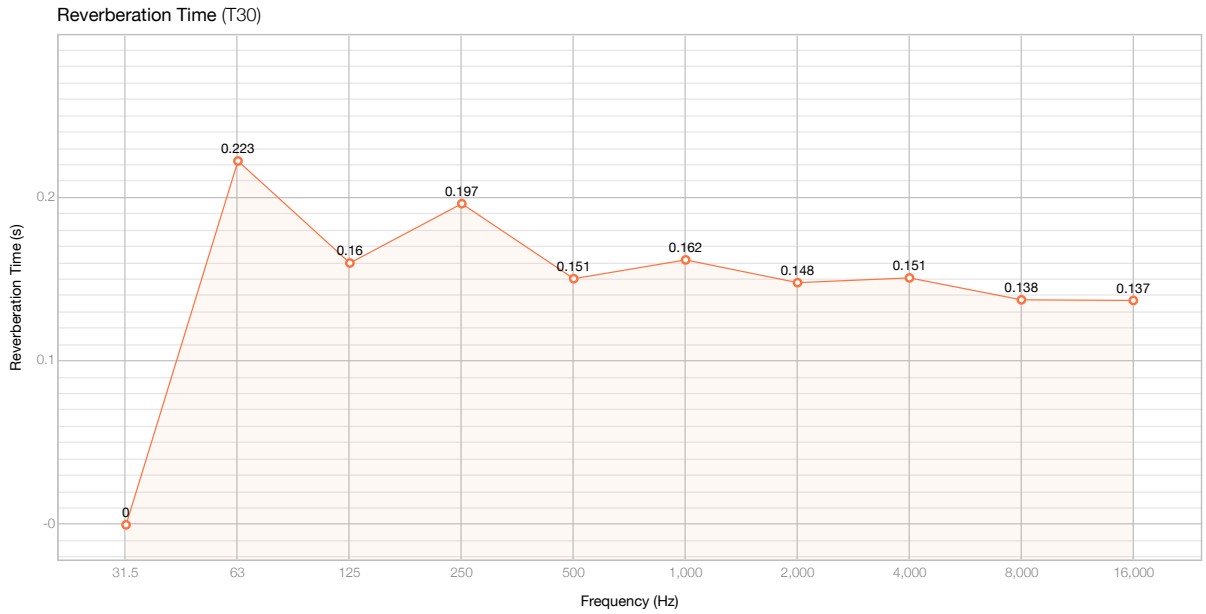


Figure 41: Studio 22 RT30.

3.1.2 Room Shape and Dimensions

The room is a hexagon with an area of 28.32 m^2 and volume of 116.67 m^3. The dimensional ratios of the studio fulfill the requirements defined in BS.1116-1[131] for a uniform distribution of the low-frequency eigentones. The room dimensions are details in Table 2.

Studio 22 Dimensions:		
Number of Walls:	6	
Front Wall:	1.95	m
Back Wall:	5.9	m
From Left Diagonal:	1.57	m
Front Right Diagonal:	1.57	m
Left Side Wall:	4.16	m
Right Side Wall:	4.16	m
Height:	4.12	m

Table 2: Studio 22 dimensions.

3.1.3 Background Noise

The continuous background noise (produced by an air conditioning system, internal equipment or other external sources), measured in the listening room area at a height of 1.2 m above the floor does not exceed NR 15, and is not perceptible (is not impulsive, cyclical or tonal in nature). Full-band Sound Pressure Level (SPL) measurement of the background noise gives the average reading of 17.8 dBA [155].

3.1.4 Loudspeaker Geometry and Configuration

The configuration of the loudspeakers is detailed in Table 1, Figure 40 and Figure 42. The left, center and right bottom loudspeakers are installed on stands near the floor level, aiming upward towards the optimum listening position. The mid- and top-ring loudspeakers are installed suspended on vertical non-resonating rails, which are hung from a ceiling grid.

The system calibration has been performed to a reference signal (pink noise at -18dBFS feeding all 22.0 loudspeakers) giving a level of 85dBA at the optimum mixing position.

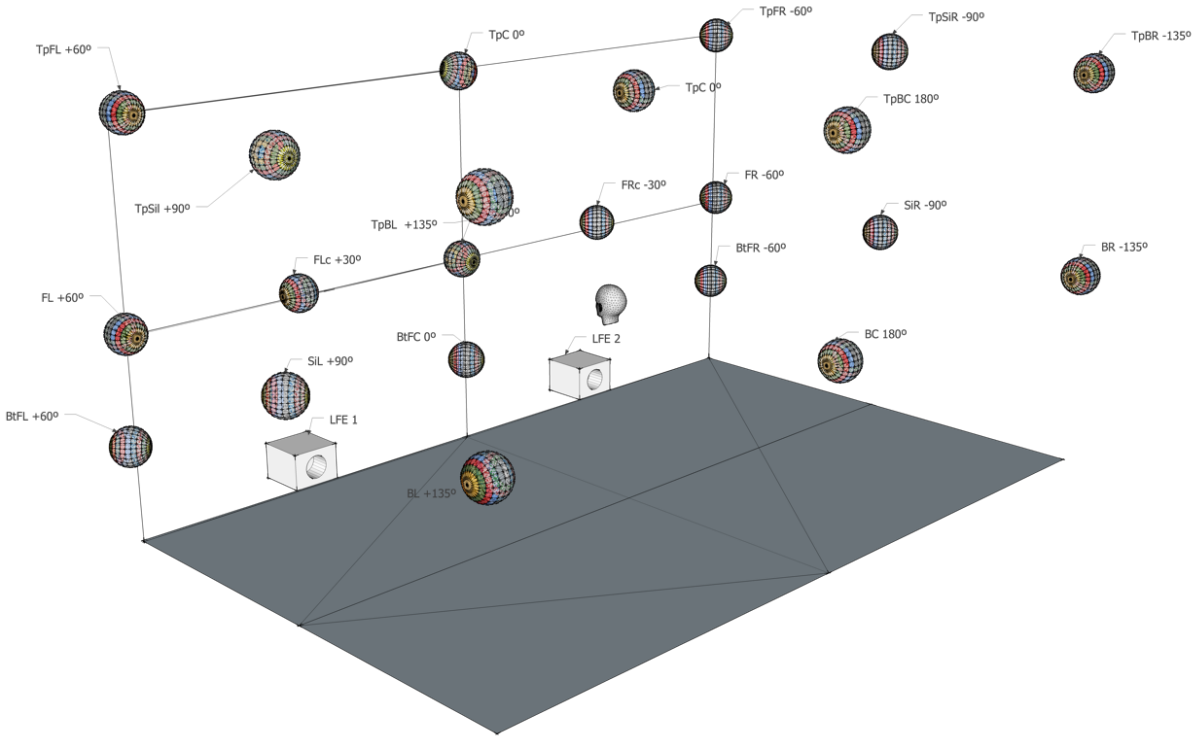


Figure 42: Studio 22 loudspeaker layout.

3.1.5 Loudspeaker System Properties

3.1.5.1 Musikelectronic Geithain ME25

- Loudspeaker type: coaxial 2-way-system in a vented box enclosure.
- Loudspeaker model: passive reference studio monitor type ME25.
- Manufacturer: Musikelectronic Geithain GmbH, Geithain, Germany.
- Number of loudspeakers: 22.
- Bandwidth: 50 Hz - 20 kHz (Figure 43).
- Maximum sound pressure level: 104 dB / r = 1m.

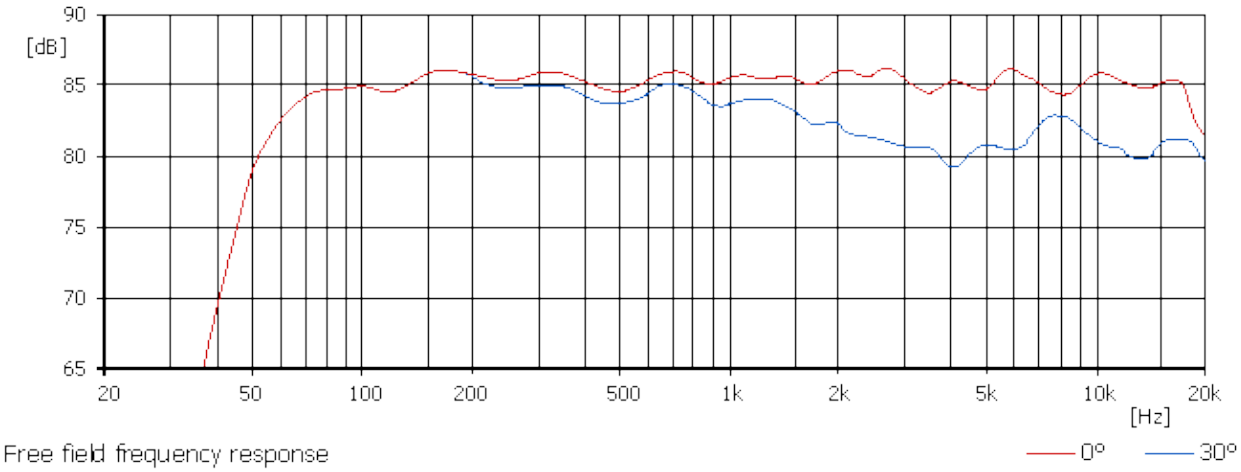


Figure 43: ME 25 free field frequency response [156].

3.1.5.2 Eclipse TD725SWMK2 Subwoofer

- Subwoofer type: R2R Twin Drive-Units, Sealed Enclosure
- Subwoofer model: TD725SWMK2.
- Subwoofer Manufacturer: Fujitsu Ten Limited, Japan.
- Number of subwoofers: 2x25cm Diameter Subwoofers.
- Overall frequency response: 20Hz – 150Hz.

3.1.5.3 Power amplification:

- Flying Mole Modular Power Amplifier 24 channels PM-162dBI x 100W.
- Flying Mole Chassis & Power Distributor DPA-M1600.

3.1.5.4 Digital to Analog conversion:

- RME M32DA Digital to Analog Converter (MADI).

3.1.5.5 Monitor level controller:

- Junger 206(4) 24-channel Analog Master Level Controller.

3.1.5.6 Level matching alignment between loudspeakers

The loudspeakers are aligned in level at a reference point using ProTools workstation.

3.1.5.7 Bass management:

- Bass management is not used.

3.1.6 The Measured Operational In-Room Loudspeaker Response [155]

The operational room response curves describe the frequency characteristic within the sound field in the listening room. The free-field frequency response of all loudspeakers is matched by the manufacturer before shipping. In room, the measured differences do not exceed the tolerance value of +/-3.0 dB in the frequency range from 250 Hz to 2 kHz.

Figure 44 shows the operational response of the loudspeakers measured in-situ using a ½" diameter omnidirectional measurement microphone pointing upwards. No frequency compensating equalization was used, and only level and time-delay were applied to compensate for small differences in the physical distance of the loudspeakers to the reference listening point.

The loudspeakers in listening room are not fully compliant with the early reflection requirement (-10 dB in the first 15 ms in the range 1-8 kHz) because of the presence of strong floor and desk reflections, and the presence of other loudspeakers.

Late energy and anomalies such as flutter echoes, tonal colorations, etc. are not present.

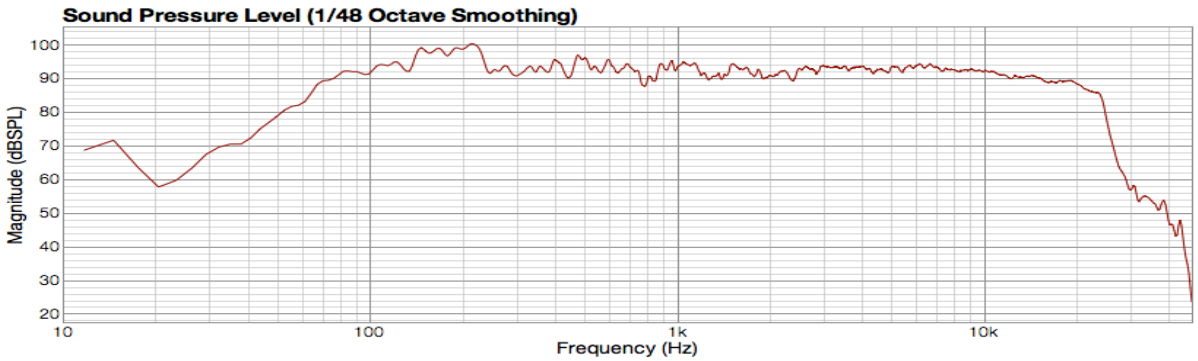


Figure 44: The averaged operational room response of all 22 loudspeakers. It represents the averaged spectral balance of sound radiation when all loudspeakers are operating [155].

3.2 Mixing Popular Music in Three Dimensions: Expansion of the Kick Drum Source Image

3.2.1 Abstract

Three-dimensional sound is being implemented in cinema, automobiles, codecs, and in new domestic listening specifications, but there is little investigation into the tools and methods needed to create music mixes in multiple dimensions. Commercial releases of popular music beyond stereo have been limited to 5.1 and 7.1 formats with no height channels present. The sound-stage architecture varies widely in these offerings, and the small number of releases has constrained the dialog for the artistic evolution of the sound-field presentation. This paper discusses evolving 3D mix architectures being developed for 22.2 multi-channel surround sound systems by McGill University's Sound Recording Program. The major topic discussed is the expansion of the size of sonic objects in two- and three-dimensional planes.

3.2.2 Introduction

The presentation of audio in three-dimensions is the next horizon in professional audio. It is being examined for delivery in cinema [2, 3, 22], home [4, 33, 157], automobiles [158, 159], headphones [160, 161], via codecs [14, 162] and up-mixing [13, 65]. The bulk of the investigations in actual recording have largely been focused in classical music recording [19, 111, 114], broadcast [21, 23], film and game sound design, and live event capture [163, 164], with little or no address of conventionally recorded popular music.

Multi-channel commercial releases of popular music have been limited to 5.1 and 7.1 formats [108] with no height channels present. The sound-stage architecture varies widely in these

offerings, and the small number of releases has constrained the dialog for the artistic evolution of the sound-field presentation.

3.2.3 Goal of Three-Dimensional Mix Investigation

The goal of the discussed mix investigation was to discover and develop effective mixing techniques for the presentation of popular music reproduced in a playback array that included height channels. The aims were to develop methods for expanding the size of sonic objects into two- and three-dimensional planes; methods and strategies for the design and implementation of early and late reflections in a three-dimensional sound field; strategies and architectures for the design of three-dimensional reverberant fields; and best practices for the distribution of the audio spectrum in hemispherical multi-dimensional playback systems. The exploration of image expansion were undertaken in the context of their function within a multitrack 3D mix for popular music from sources described in 3.2.5. The discussions of this paper are limited to the expansion of the sonic image of the kick drum.

3.2.4 Test Environment

This research was conducted at McGill University in the Schulich School of Music's Recording Studio 22 (3.1).

3.2.5 Experimental Design

The source material used for this investigation was a commercially released track recorded in the early 1980s that reached the top 10 of the U.S. Mainstream Rock chart in Billboard magazine. The sources were largely recorded via direct injection (DI) of electrical signal into the

console, and those recorded with microphones contained little or no natural ambience as outlined in Table 3.

Instrument Type:	Track #:	Instrument:
Acoustic Drum Kit	1	Kick drum
	2	Snare drum
	3	High Hat
	4	Overhead left
	5	Overhead right
	6	Pluck gtr dbl bass
Drum Machine (DI)	7	Conga
	8	Claps
	9	Cowbell
Samples (DI)	10	Orchestra hits
	11	Drum rolls
Guitars recorded via direct injection (DI)	12	Bass guitar
monophonic	13	Guitar chords A
monophonic	14	Guitar chords B
monophonic	15	Pluck Guitar A
monophonic	16	Pluck Guitar B
	17	Guitar doubling bass
Synthesizers (DI)	18	Melody line synth
	19	String lines
	20	Emulated piano
Vocals	21	Lead vocal
	22	Lead vocal double
	23	Background vocal 1
	24	Background vocal 2

Table 3: Track listing for 1980s multi-track used for experiment.

The processing and plugins used in this mix study were conventional tools developed for monophonic and stereo use. Commercially available reverberation, delay, and processing plug-ins were used in the design of the three-dimensional spatial architecture. Multi-channel reverberation from McGill University's Virtual Acoustics Technology Laboratory (Space Builder) was the only

dedicated multi-channel tool employed. It was used in the design of the hemispherical spatial architecture.

The approach taken for this investigation was to create a believable three-dimensional presentation of the source material within the 22.2 array using existing tools, and afterwards examine the strategies and techniques that proved effective. Critical listening assessment was done by faculty and graduate students of the McGill Sound Recording department. The mix platform was Protools 10, and all processes (excluding Space Builder) were performed within the DAW.

The requirements set out for the creation of an effective 3D sonic image from a monophonic source as perceived from the mix position were:

1. Defined localization within the hemispherical environment.
2. Image size should be appropriate to musical function.
3. Image expansion should encompass optimally three, and minimally two-dimensional planes.
4. The image must have a coherent immersive aspect that places it believably within the hemispherical environment.

The following explanation details the development of the image of the kick drum.

3.2.6 Techniques Used to Expand the Kick Drum Image

3.2.6.1 Original Monophonic Source Track

The monophonic source image of the kick drum was located in the middle center loudspeaker of the 22.2 array. Conventional equalization and compression were employed as in standard mix practice, (Figure 45).

3.2.6.2 Expanding Image into Two Dimensions

The original source track was bussed to a separate channel strip within Protools and assigned to the center loudspeaker in the lower LCR, (Figure 46). This increased the perceived size of the image and expanded it into two dimensions. Separate equalization and compression were used on this track.

3.2.6.3 Increasing Immersive Content

The original source track was bussed to a third channel strip within Protools and sent of the center rear loudspeaker of the middle ring (Figure 47). Discrete compression and equalization were again used on this channel. The purpose of this technique was to add an immersive aspect to the kick drum. No listeners reported localization of the kick drum image from the rear. This technique expanded the image of the kick drum into three dimensions, and also served to distribute low frequency energy to multiple loudspeakers throughout the array. Localization of the image to the front center was maintained, image size was increased to three dimensions, but the result was judged to lack both the desired immersion and power appropriate for believability in the context of listener experience with an actual instrument in a room.

3.2.6.4 Expanding Perceived Size and Power

Figure 48 illustrates the addition of the left and right subwoofers to the kick drum image. The monophonic source track arrived there via a stereo send and was panned center. This localized the image downward and added extended low frequency response within the sound field.

3.2.6.5 Focusing Impact of Image

At this stage in the development of the 3D image, the kick drum was judged to have acceptable size, but lacked detail. The solution was to buss the original source track to a fourth channel strip and add a high degree of compression. This track was panned center in the narrow left and right loudspeakers of the middle ring and blended with the expanded 3D image, (Figure 49).

4.1.8 Surround Image Expansion

The final step was to expand the immersive aspect of three-dimensional image of the kick drum. The original monophonic source track was bussed to the middle rear left and right surrounds, (Figure 50). The final addition improved immersive perception, distributed additional low frequency energy, and added weight to the image without detracting from the frontal localization.

3.2.6.6 Hemispherical Integration

The last step in the creation of the kick drum image was to integrate it into the global spatial architecture. This was achieved by the addition of the Space Builder multi-channel reverberation, (Figure 51).

The final immersive, three-dimensional image of the kick drum within the 22.2 sound field contained information from nine discrete loudspeakers: (1) FC, (2) FLc, FRc, (5, 6) LFE 1, 2, (7) BC, and (8, 9) BL and BR. Separate processing (equalization and compression) was used on the track information routed to loudspeakers 1-6.

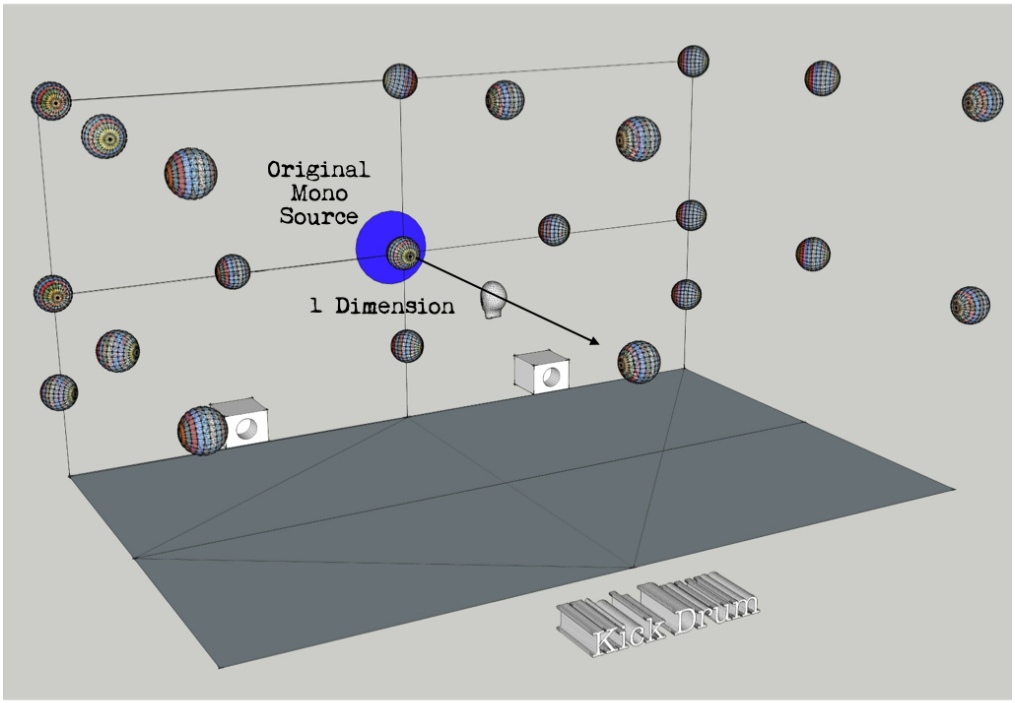


Figure 45: Original monophonic source image.

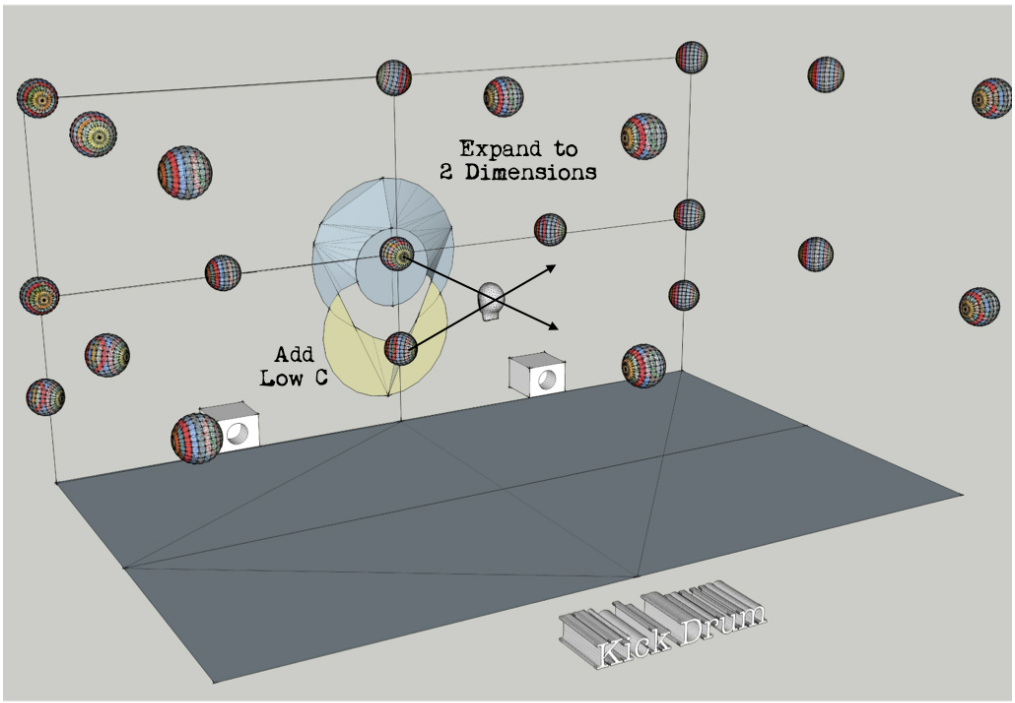


Figure 46: Low center loudspeaker added to image.

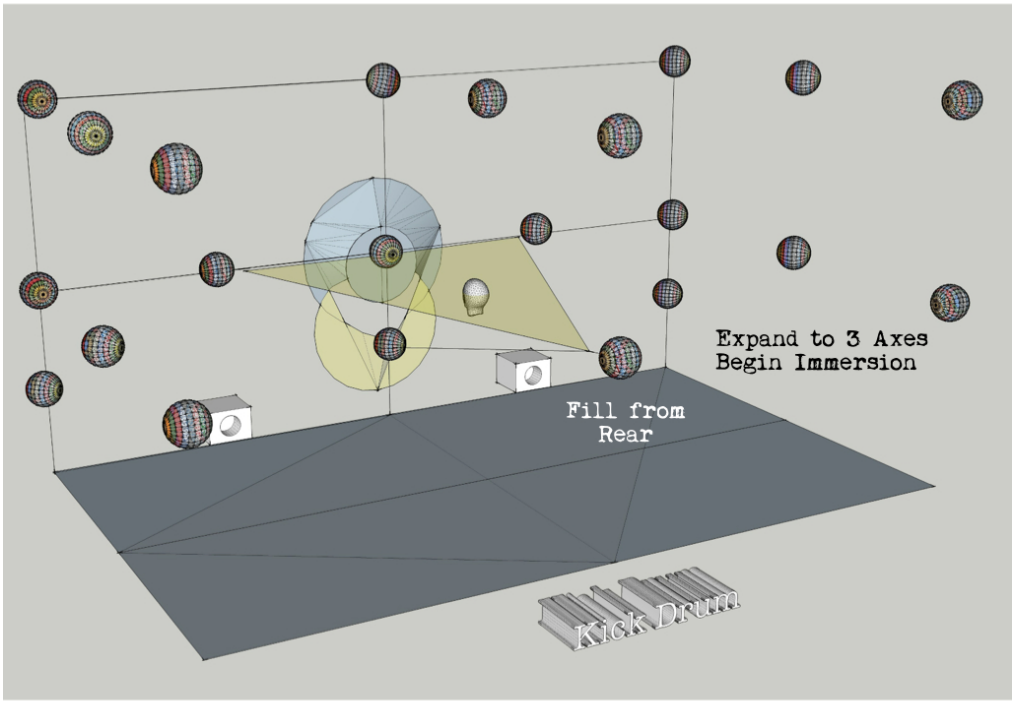


Figure 47: Rear middle center loudspeaker added to image.

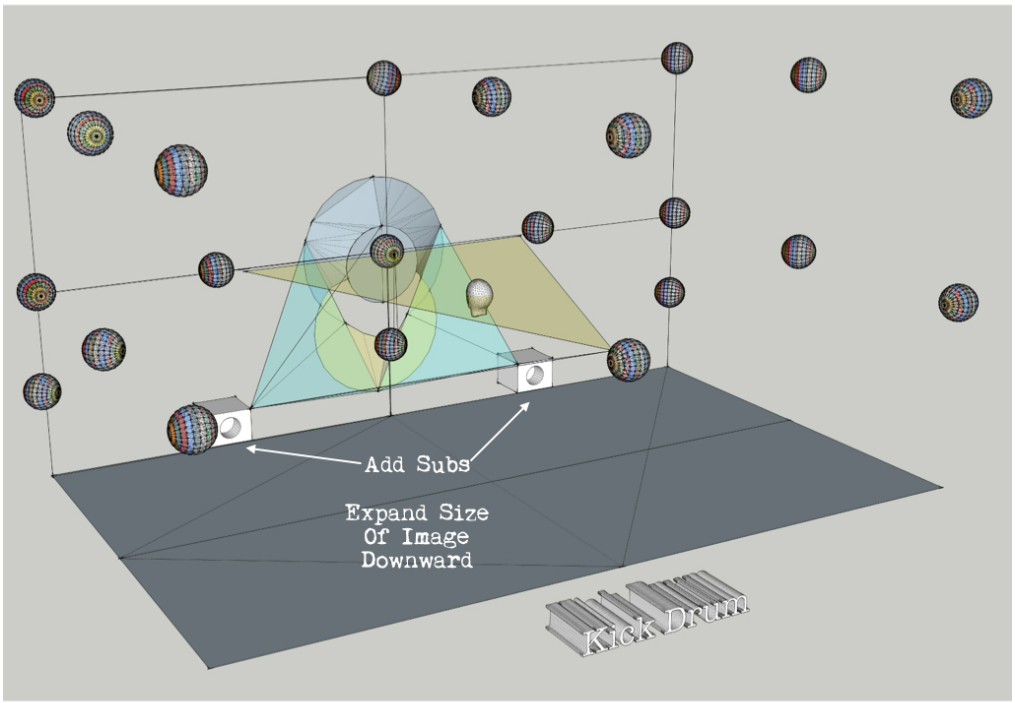


Figure 48: Subwoofer loudspeakers expand image.

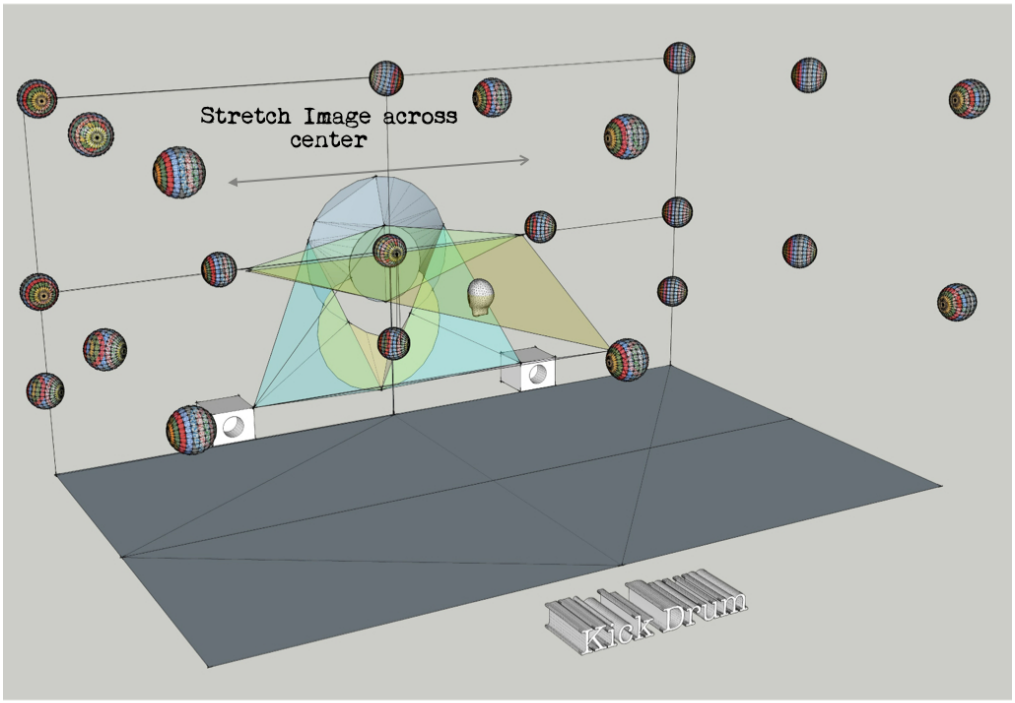


Figure 49: Narrow left and right middle loudspeakers adds focus to image.

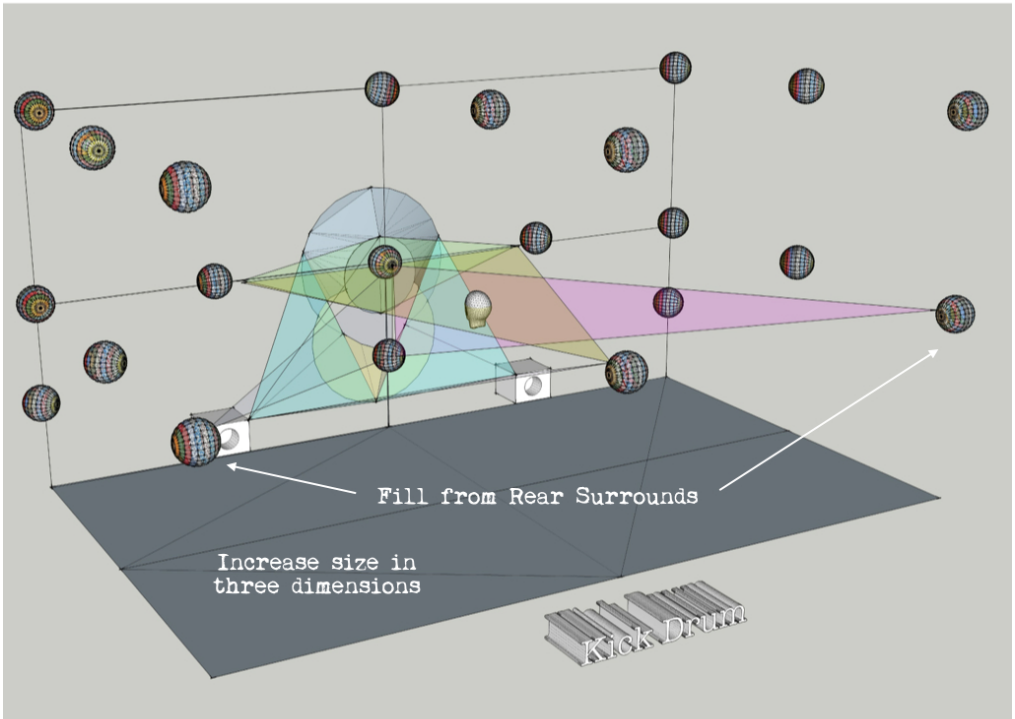


Figure 50: Middle left and right rear surround speaker increase immersive content of image.

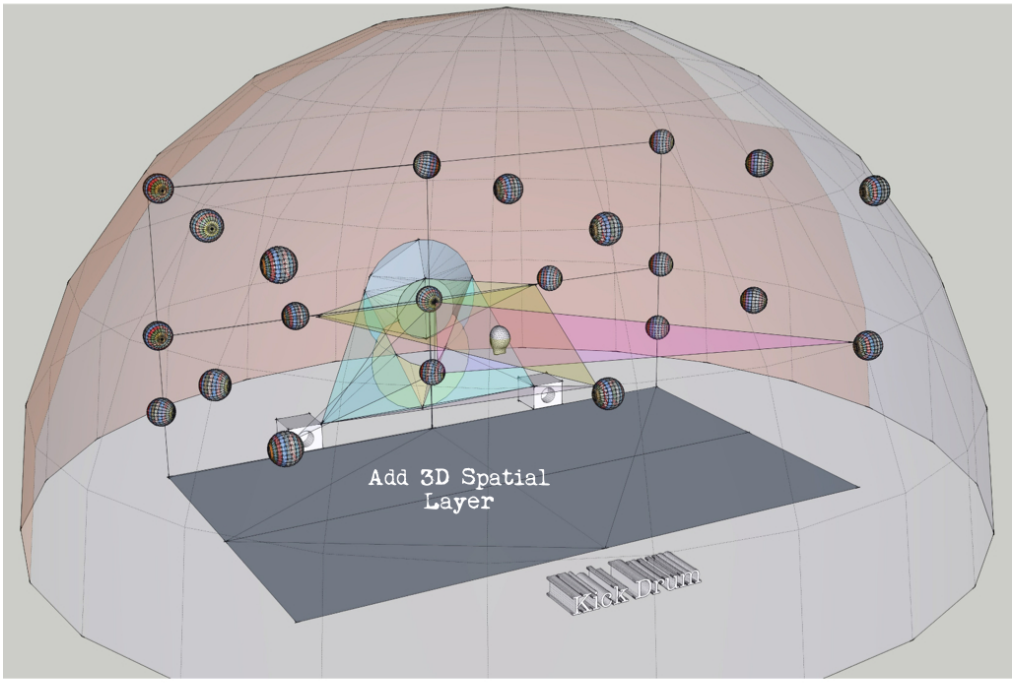


Figure 51: Expanded image is placed in hemispherical acoustic.

3.2.7 Conclusions

The extent of image development to achieve three-dimensional believability and immersion is far greater than that required for conventional stereo presentation. It was found that improved believability was achieved when the source image was spread into three distinct planes during playback, essentially sound radiation in the X, Y, and Z axes. It was also the observation of the author, that low frequency power and immersion increased when spread between multiple loudspeakers that were not necessarily those that created the source localization.

While there has been much development in the area of immersive reverberation, there is still the lack of multi-dimensional tools for the creation and processing of 3D volumetric audio images. For these tasks the user must still rely on the use of conventional monophonic and stereo tools which necessitates a large time investment and tedious workflow. This workflow is not

commercially viable, and would be improved with multi-channel (4-8) equalization, compression and panning processors.

3.2.8 Future Work

The above experiment made it clear to the investigator that mixing and recording of popular music for three-dimensional presentation is in its infancy. Every aspect of this type of capture and presentation will require much work to be understood and mastered.

Future work to propel this area forward will include the development of studio recording techniques and practices for three-dimensional capture of sound sources, the continued development and improvement of multi-channel processing tools for the manipulation of the source tracks, and the development of spatial processing handling early and late reflections, reverberation, and other effects processes.

In the current climate, much effort for 3D sound creation is being focused on post-processes to generate the immersive experience. It is the opinion of the author that a fundamental understanding of the basic principles derived from the practice of both mixing and recording in loudspeaker arrays with height channels will aid, enhance, and help define the development and architecture of future 3D tools.

3.3 Immersive Content in Three Dimensional Recording Techniques for Single Instruments in Popular Music

3.3.1 Abstract

"3D Audio" has become a popular topic in recent years. A great deal of research is underway in spatial sound reproduction through computer modeling and signal processing, while less focus is being placed on actual recording practice. This study is a test in establishing effective levels of height-channel information based on the results of a listening test. In this case, an acoustic guitar was used as the source. Eight discrete channels of height information were combined with an eight-channel surround sound mix reproduced at the listener's ear height. Data from the resulting listening test suggests that while substantial levels of height channel information increase the effect of immersion, more subtle levels fail to provide increased immersion over the conventional multi-channel mix.

3.3.2 Introduction

This study is a test in establishing playback levels of height-channel information that are considered to be effective. Eight discrete channels of height information were presented in conjunction with an eight-channel discrete multi-channel mix of solo acoustic guitar. The latter is presented in one horizontal plane at the listener's ear level as front L/C/R, rear L/C/R, and side channels positioned at +/-90 degrees (Figure 52). The ring of height channels copies the number and placement of the middle ring of loudspeakers, positioned 1.5 meters above it. This configuration comprises the middle and top layers of the 22.2 SMPTE standard 2036-2 [72] that was developed by the Japanese broadcaster NHK [70].

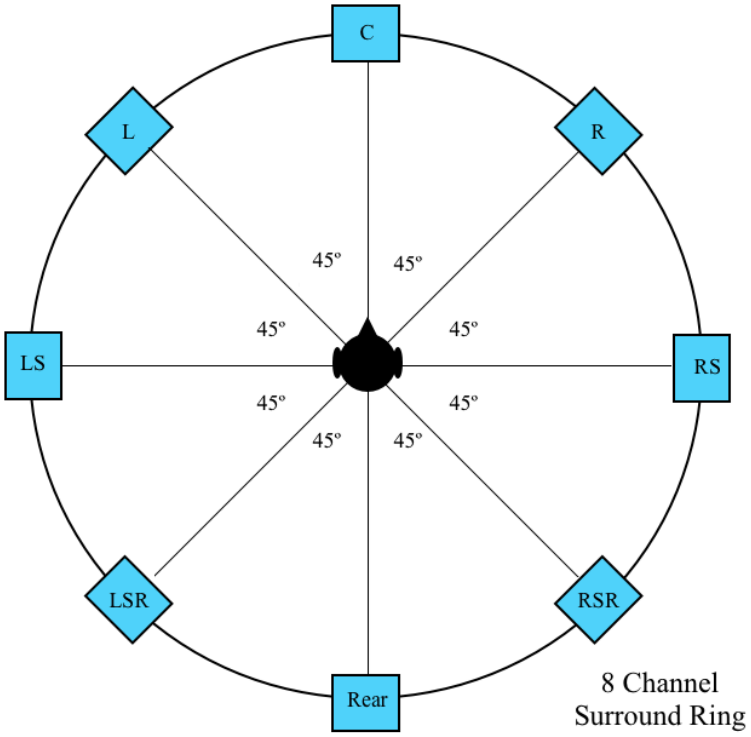


Figure 52: Eight Channel Surround Loudspeaker Ring. (Height channels copy first ring of loudspeakers 1.5 meters above).

3.3.3 Test Design

3.3.3.1 Ambient Recording Configuration

For the ambient audio component of the study, the experiment was designed such that microphones were placed in the recording studio with positioning and spacing that mirrored the number of loudspeakers in each playback "ring" of the control room; i.e. eight microphones for each ring of eight loudspeakers; sixteen microphones in total (Figure 53).

In this test, an acoustic guitar was recorded in the center of the studio. In early test recordings, several distances to the source and microphone heights were compared. After the auditioning of several radii by the authors, the radius decided upon was 1.22 m. The mid-microphone ring was placed 1.54 m from the floor (corresponding to the control room mid-ring loudspeaker height), and the high ring was positioned at 2.44 m. All microphones were pointed at the guitar (Figure

54). An additional close microphone was used to capture the direct sound of the instrument, as would be common in popular music production. This microphone was placed for optimum sound quality as determined by the recording engineer, and carefully balanced into the center channel of the multi-channel mix presented at ear level. The microphones used for this recording are listed in Table 4. All were cardioid or sub-cardioid types.

All microphones were recorded through a Sony SIU 100 interface, which provided microphone pre-amplification (DMBK-S101 cards) and A/D conversion. Careful attention was paid to match the input gains of the sixteen microphones in the mid and high rings within ±1dB. The session was recorded at a 96kHz sample rate.

	Mid Ring	High Ring
L	Schoeps CMC62U / MK 4	Schoeps CMC62U / MK 21
C	Schoeps CMC62U / MK 4	DPA 4011-TL
R	Schoeps CMC62U / MK 4	Schoeps CMC62U / MK 21
LS	Schoeps CMC62U / MK 4	ADK HA-TL-II Cardioid
RS	Schoeps CMC62U / MK 4	ADK HA-TL-II Cardioid
LSR	Schoeps CMC62U / MK 4	Schoeps CMC62U / MK 21
REAR	Schoeps CMC62U / MK 4	DPA 4011-TL
RSR	Schoeps CMC62U / MK 4	Schoeps CMC62U / MK 21
Close	ADK C-LOL-67 capsule, ADK HA-TL-II body	

Table 4: Microphones used in recording.

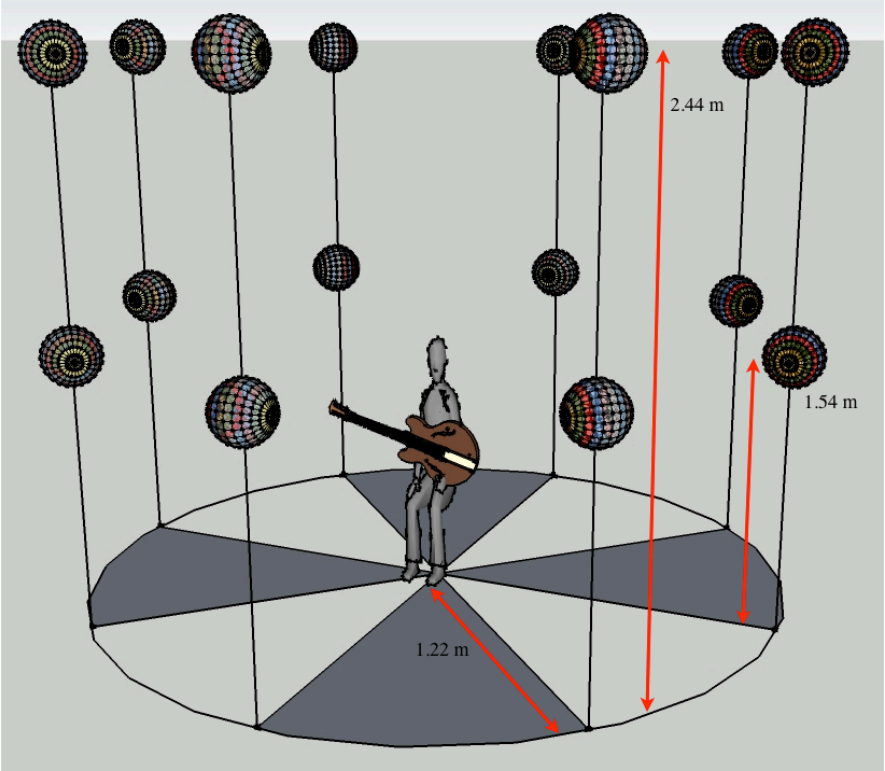


Figure 53: Recording studio microphone positions corresponding to control room loudspeaker positions.

3.3.3.2 Recording Studio

The recording studio was rectangular in shape (11 m x 7 m) with an RT30 of ≈0.7 seconds (Figure 55, Figure 56). The ceiling height was 5.7m. The wall treatment was a combination of absorption and diffusion with the upper walls and ceiling being more reflective. The studio exhibited a uniform frequency response from 40 Hz to 10kHz (Figure 57).

3.3.3.3 Test Environment Control Room Monitor Environment

This research was conducted at McGill University in the Schulich School of Music's Studio 22 (See section 3.1).

Figure 54: Recording studio microphone positions.

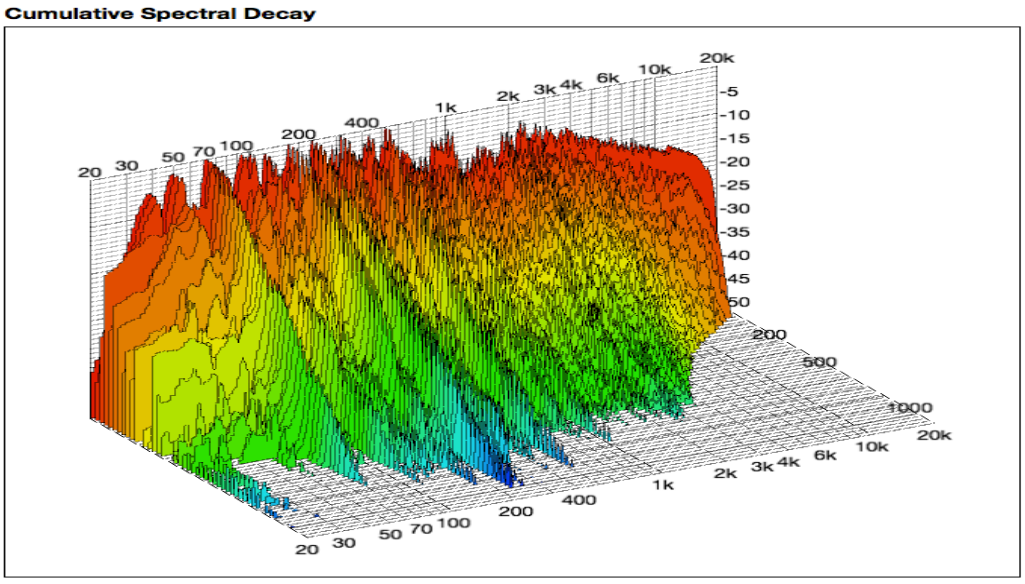


Figure 55: Spectral plot of recording studio (10 second sine sweep).

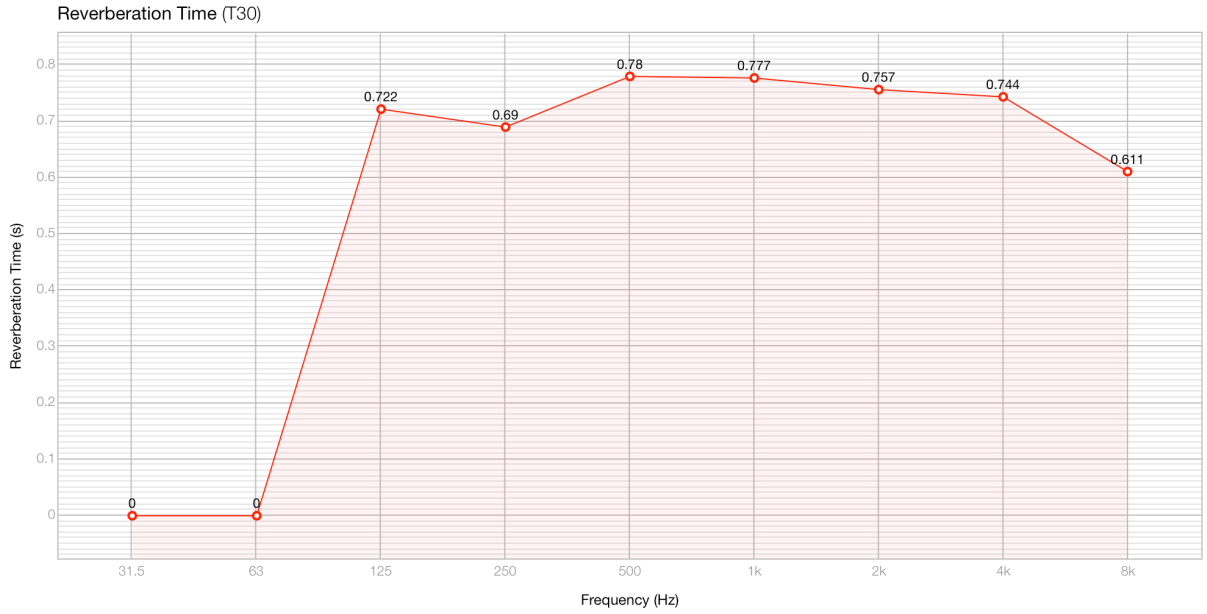


Figure 56: Reverberation time (RT30) of recording studio.

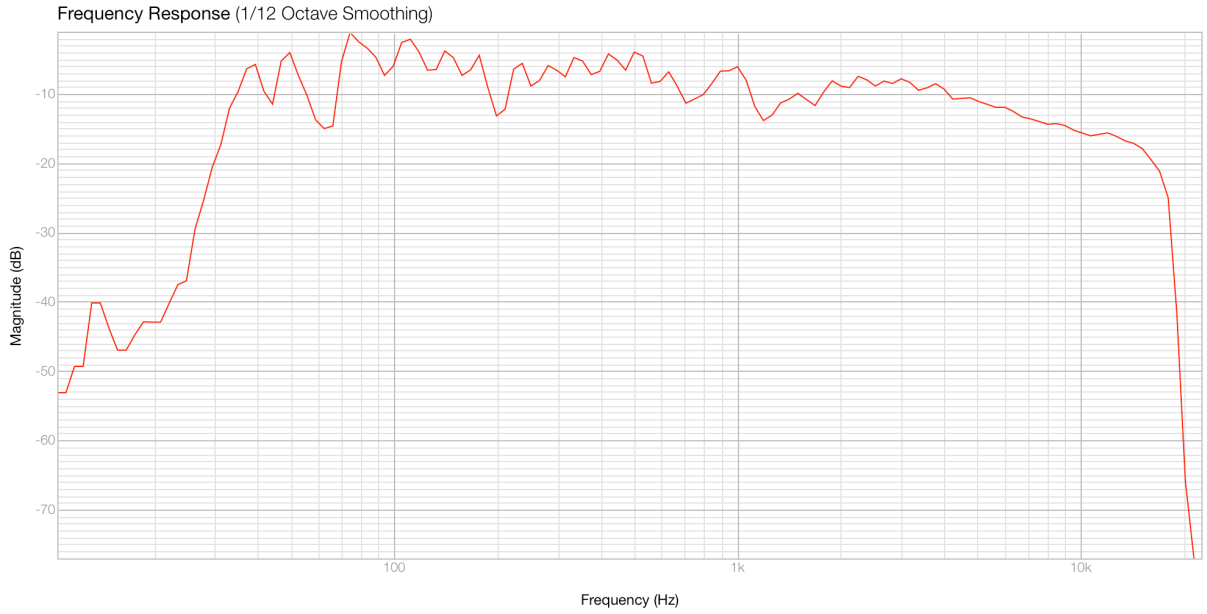


Figure 57: Frequency response of a full-range loudspeaker in the recording studio.

3.3.3.4 Musical Material

The music used for testing was played on a 1959 Harmony Monterey arch-top acoustic guitar (Figure 58). A cyclical chord progression in E major was played using primarily open voicing.

Figure 58: 1959 Harmony Monterey Archtop guitar.

3.3.4 Testing Software & Methodology

Testing was achieved using a software patch developed in Max MSP. The testing software managed audio playback, data collection, and treatment order shuffling.

The treatments to be evaluated for immersion consisted of eight discrete channels of height information presented at five different volume levels, in conjunction with an eight-channel discrete multi-channel mix of solo guitar. The five height channel levels in dB were 0, -6, -16, -22, and -144 (no signal) measured at the listening position relative to a 0dB signal played back through the middle loudspeaker layer. These levels were determined by the authors and a select group of expert listeners to be fairly equal steps between “full immersion” and “very subtle” immersion. All five

upper ring levels were presented randomly in combination with the main ring, and without repetition in each subsequent trial. Listener-ranked preference and treatment-presentation order was captured in the resulting data.

The software patch presented users with a graphical user interface (GUI) allowing for basic control of audio playback, as well as a rating "slider" in order to rate each treatment's "immersiveness", as experienced by the listener. The GUI's sliders were completely without scale or numeric indicators and were labeled only as "less immersive" and "more immersive" from left to right (Figure 59). Each slider allowed for an immersive rating of 0-100, and the default starting position for each trial was 50. Additionally, users were able to vary any slider in a given trial, regardless of which of the five levels was selected for playback, allowing for flexibility in adjustment during the test.

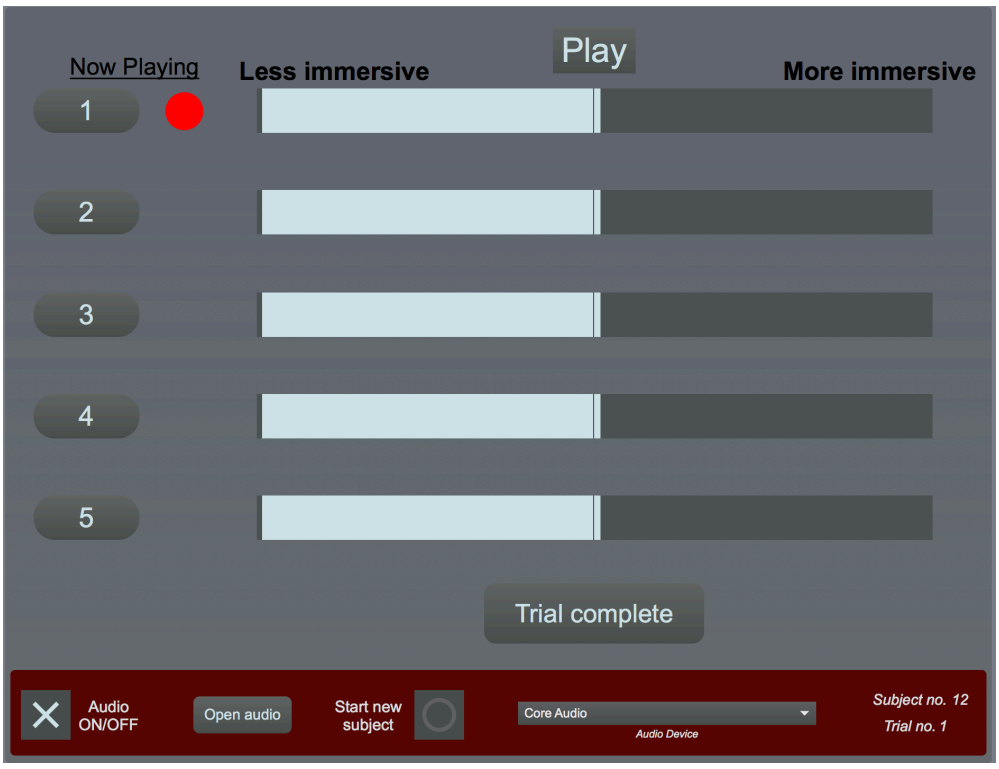


Figure 59: Graphical User Interface (GUI).

3.3.4.1 Subjective Preference Question

In addition to being asked to rate their impression of immersiveness in each presentation, the subjects were asked to select a personal preference from each of the five treatments presented in each trial. These preferences were noted on a questionnaire provided to each subject after the listening session.

3.3.5 Subjects

Thirty test subjects were drawn from the students and staff of the graduate program in Sound Recording at McGill University. All subjects had significant musical training, averaging more than 14 years; and averaging over nine years of experience in music recording and production. The subject pool was composed of individuals specializing in recording, production, and mixing.

3.3.6 Results and Analysis

3.3.6.1 Immersion Ratings

An analysis of variance was performed on the immersion ratings elicited by each of the five height channel levels. Prior to the analysis, the normality of each group was verified using a one-sample Kolmogorov-Smirnov test. All five showed as normal. Differences were found between the group means ($p < 0.05$). The means for each height level were 0 dB: 76; -6 dB: 62; -16dB: 46; -22dB: 45; -144 dB: 45. Tukey's HSD test revealed a significant difference between the immersion of the 0 dB and -6 dB height channel levels. The three lower levels (-16, -22, -144) were significantly different from the two higher levels but were not different from each other (Figure 60).

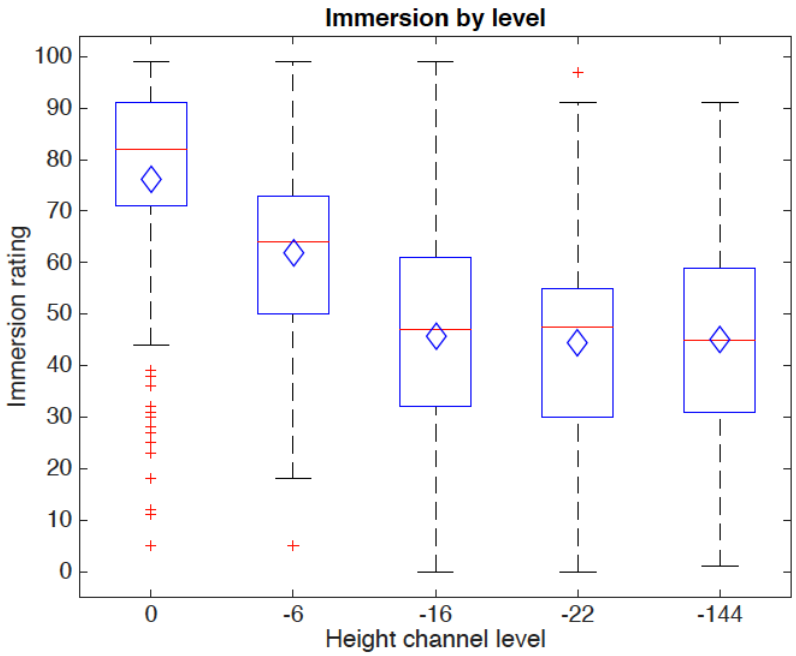


Figure 60: Immersion ratings grouped by height channel levels

3.3.6.2 Preference Question Data

To analyze preference choices, the five height channel playback levels were split into two groups. The first group was deemed "non-immersive" and consisted of the three height levels that produced no differences in immersion ratings (-16, -22, and -144dB). The second group was deemed "immersive" and consisted of the two choices linked with high immersion (0 and -6 dB).

Immersive stimuli were preferred significantly more often than non-immersive stimuli ($p < 0.05$, binomial test, Figure 61). Data were excluded from one subject who forgot to complete the questionnaire

3.3.6.3 Subject Consistency Scores

Subjects varied considerably in the consistency of their preferences. Some subjects chose the same immersion level repeatedly from trial to trial, while others shifted their preferences over the course of the test.

The consistency of each subject was gauged by the variance in his or her preferences. To measure consistency, each preference choice was assigned to one of three groups, and each group was associated with a numeric immersion level. Immersion level 0 contained the 0 dB height channel choice; immersion level -1 the -6 dB choice; and immersion level -2 the -16 dB, -22 and -144 dB choices. Variance in height-choice levels was then calculated for each subject. This variance, multiplied by -1, was referred to as the subject's consistency score (Figure 62).

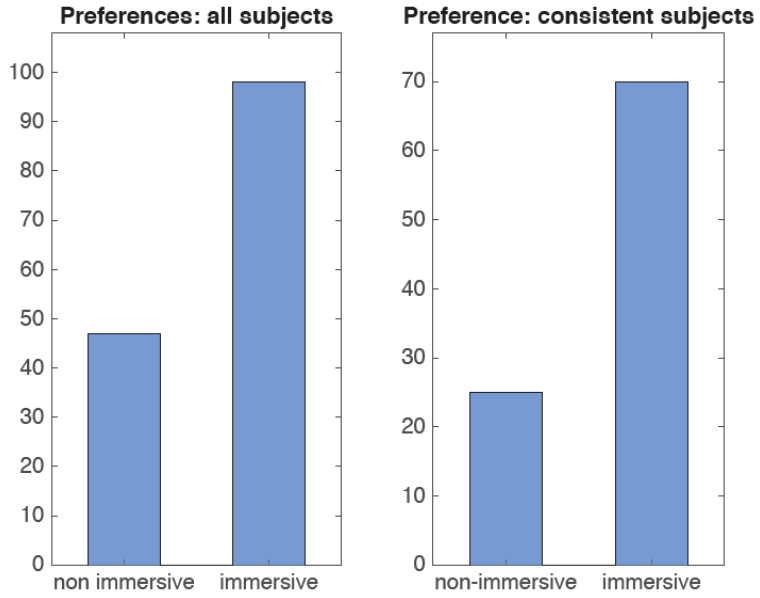


Figure 61: Preference choices for all subjects: Immersive stimuli were preferred in 98 out of 145 trials.

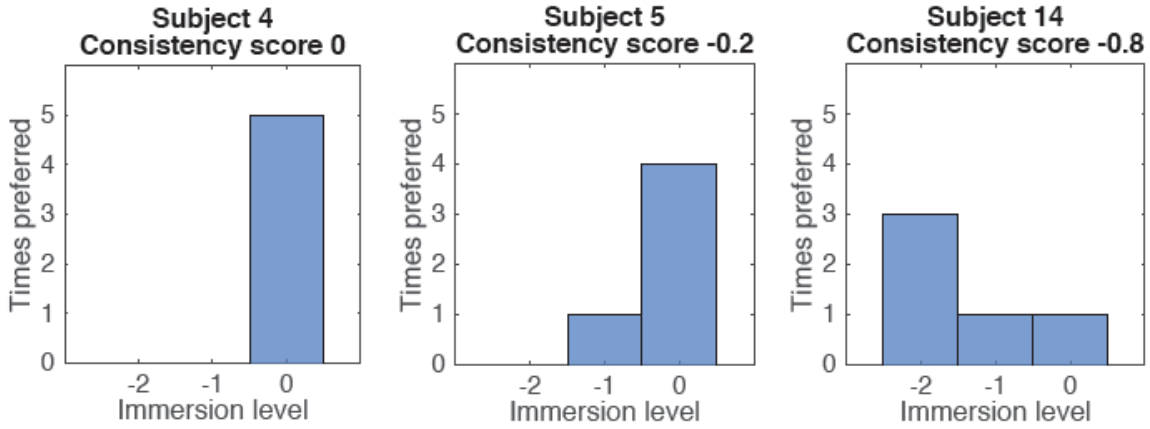


Figure 62: Consistency Scores: Examples of three subjects exhibiting high, medium and low consistency scores.

3.3.6.4 Preferences for Consistent Subjects

Consistency scores were used to divide the subjects into consistent and inconsistent groups. A score of -0.3 was used as a cut-off. Nineteen subjects were at or above this cut-off and deemed consistent. Ten subjects were below cut-off and deemed inconsistent.

In examining the consistent group, a trend toward preference for immersion became clearer. Among consistent subjects immersion was preferred in 77% of trials, versus 68% in both groups combined. Both results were significant ($p < 0.05$, binomial test).

The statistical results of the Immersion Ratings test provided significant results for the perception of immersive content, and provides a baseline for the minimum level at which height channels can be perceived in this particular test scenario.

3.3.7 Conclusions

The findings of this study were:

1. There is a minimum level of height information below which subjects could not differentiate added height content. These levels, -16dB, -22dB, provided the same perceived immersion as the mid eight-channel loudspeakers with no additional immersive content (-144dB).
2. The subjects could perceive three distinct content levels during testing: 0dB (immersive), -6dB (immersive), and the "-16, -22, and -144dB" group (little or no-immersive-content).
3. The level of the immersive content (from all height channels) needs to be substantially louder to be perceived, ≥ 10dB.
4. The Preference Question results suggest that subjects preferred a more immersive environment than the more subtle levels of immersion, when given the choice.

3.3.8 Possibilities for Future Work

Research will continue in the design and implementation of immersive mixing methodologies. These studies should include the exploration of immersive architectures for early and late reflections, the design of immersive reverberant fields and the implementation of multi-channel impulse responses for reverberant fields, as well as the other techniques for the expansion of audio images into three dimensions. Further studies should be undertaken to develop microphone arrays and recording techniques that directly provide stable three-dimensional images for popular music mixing and reproduction.

3.4 Three Dimensional Spatial Techniques in 22.2 Multi-channel Surround Sound for Popular Music Mixing

3.4.1 Abstract

Current multi-channel spatial mixing practices are largely limited to the construction of three-dimensional space using two-dimensional panning tools (meant for 5.1, 7.1, etc.), and those designed for common stereo production. A great deal of research is currently underway in spatial sound reproduction through computer modeling and signal processing, with little focus on actual recording and mixing practices. This investigation examines the design and implementation of early and late reflections, and reverberant fields in 22.2 multi-channel sound system mixing based upon research in listener envelopment. The techniques discussed will include the expansion of spatial elements into three dimensions using conventional tools, and the implementation of multi-channel impulse responses for reverberant fields. Listening tests were conducted reviewing the final music mix with positive results reported for listener immersion.

3.4.2 Experimental Design

The source material used for this investigation was a commercially released track recorded in the early 1980s. The sources were largely recorded via direct injection (DI), and those recorded with microphones and containing little or no natural ambience.

The approach taken for this investigation was to create a believable three-dimensional presentation of the source material within the 22.2 Multi-channel Sound System, and afterwards empirically examine the strategies and techniques that proved effective. Critical listening assessment was provided by faculty and students of the graduate program in Sound Recording at

McGill University. The mix platform was Protools 10, and all processes (excluding the proprietary Space Builder unit) were performed within the DAW.

The processing and plugins were conventional tools developed for monophonic and stereo use. Commercially available reverberation, delay, and processing plug-ins were used in the design of the three-dimensional spatial architecture. Multi-channel reverberation from McGill University's Virtual Acoustics Technology Laboratory (Space Builder) was the only dedicated multi-channel tool employed. It was used in the design of the hemispherical spatial architecture.

3.4.3 Test Environment

This research was conducted at McGill University in the Schulich School of Music's Studio 22. The loudspeaker configuration of the 22.2 multi-channel sound system is detailed in Figure 63 and section 3.1.

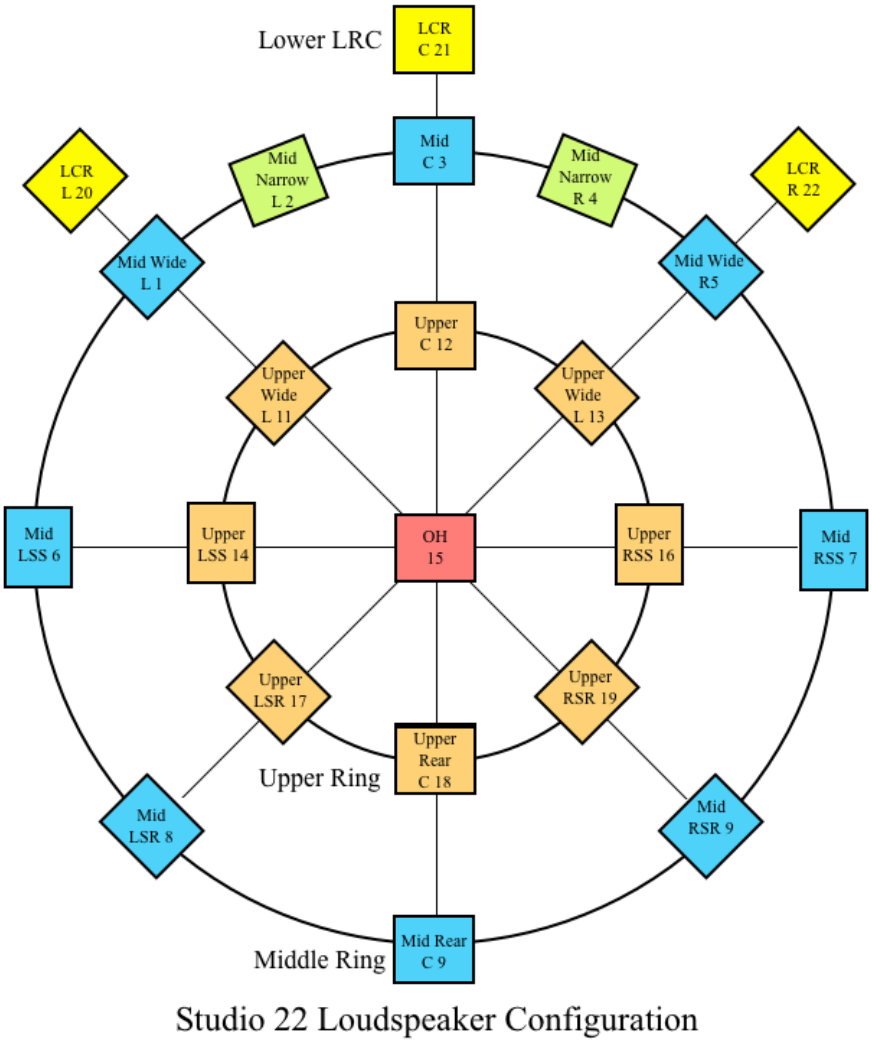


Figure 63: Studio 22 loudspeaker configuration.

3.4.4 Construction of the Virtual Acoustic

The goal of the mix investigation was to create listener envelopment (LEV), wide audio source width (ASW), and achieve the impression of three-dimensionality in the final musical presentation. The scientific (physical and perceptual) basis for the construction of the mix was taken from research findings on listener envelopment.

Nyberg and Berg [165] sum up the work of Beranek [166]: '*Envelopment is defined as the subjective impression of being enveloped by reverberant sound in a concert hall—reverberant*

sound defined as sounds arriving at the ear 80ms after the direct sound', and Bradley and Soulodre [167] '*ASW appears when reflections are present in the 80ms window from the initial sound. After this window the reflections becomes late arriving sound energy and affect the LEV of the sound. The degree of late arriving reflections, after 80ms, diminish the ASW of the sound and make the LEV more present*.'

Nyberg and Berg [165] summarize that '*There is no clear agreement on the position or the size of the time window in which these reflections arrive at the listener. The proposed lower limits are 80ms, 105ms and 150ms. However, there is a majority of work indicating that sound energy after 80ms creates LEV*'.

The above conclusions were applied in the construction of the virtual acoustic envelope of the mix. The placement of delays and reverberations followed the guidelines:

1. Discrete delay times and reverberation pre-delay times were largely kept between 50ms to 100ms, which falls within the range indicated in the above researcher.
2. The reverberation times ranges from 0.9 to ~3.0 seconds which exceeds the minimum time window for the creation of LEV.
3. Lateral reflections are used to increase LEV.
4. Discrete left and right reverberations and delays are used to maintain ASW.
5. Spatial information arrives at the listener from all loudspeaker rings/directions.

3.4.5 Placement of Discrete Delays

Eleven discrete delays were used in total (Table 5, Figure 64). The delays were positioned to provide balanced spatial information from all directions. In an effort to achieve a wide ASW, 93ms and 78ms delays (Early Reflections 1, 2) were placed at the Mid Wide L1 and Mid Wide R5

loudspeaker positions. The depth of the center-front image was defined by Early Reflection 3 at 32ms, and longer Echo Delays 4 and 5, which were 237ms and 474ms respectively. These delays were used primarily for elements that appeared in the center front position. The side delays provided lateral reflections to improve LEV. The rear delays (8-11: Upper and middle rings) completed the 360º immersion of the listener. Multiple source tracks were routed to these delays. Delays were also routed to each other (and selected reverbs) to increase the complexity and blend of the perceived reflections.

3.4.6 Placement of Discrete Reverberations

Seven stereo reverberation plug-ins were used to create the immersive acoustics for this mix under investigation (Table 6, Figure 65). The front sound stage was created using a small room (0.9 second decay, identified as light pink) located between Mid Wide L1 and R5, and a longer hall (1.1 second decay, light purple) placed in the Upper Left 11 and Right 13 positions. The aim was to give the successive layers of spatial information an increase in the perceived height and depth of the acoustic space.

Morimoto et al. [168], among others have found that ASW is negatively affected by high values of interaural cross-correlation (IACC). To reduce this effect, to maintain a wide front image, and to increase the lateral de-correlation plus increase the LEV as well as ASW, discrete left/right reverbs were employed in the front left-to-right sound stage, and on the sides.

Description:	Delay Time (ms):	Position:	Loudspeaker(s)
Early Reflection 1	93	Wide Middle Left	Mid Wide L1
Early Reflection 2	78	Wide Middle Right	Mid Wide R5
Early Reflection 3	32	Middle Center	Mid C3
Echo Delay 4	237	Narrow Middle Left	Mid L2
Echo Delay 5	474	Narrow Middle Right	Mid R4
Side Early Reflection 6	75	Left Middle Side	Mid LSS 6
Side Early Reflection 7	49	Right Middle Side	Mid RSS 7
Rear Early Reflection 8	56	Middle Left Rear	Mid LSR 8
Rear Early Reflection 9	61	Middle Right Rear	Mid RSR 10
Rear Early Reflection 10	50	Upper Left Rear	Hi LSR 17
Rear Early Reflection 11	44	Upper Right Rear	Hi RSR 19

Table 5: Parameters of discrete delays.

Description:	Pre-Delay Time (ms):	Decay time (sec):	Position:	Loudspeaker(s)
Small Room	25	0.9	Wide Front L-R	Mid Wide L1 and R5
Short Hall	53	1.1	Upper Wide L-R	Hi Left 11, and Hi Right 13
Medium Room Left	33	1.3	Left Front	Mid Wide L1, and Mid LSS 6
Meduim Room Right	33	1.3	Right Front	Mid Wide R5 and RSS 7
Medium Hall	72	2.5	Left Side	Mid Wide L1 and Mid LSR 8
Medium Hall	66	2.5	Right Side	Mid Wide R5 and Mid RSR 10
Long Hall	53	1.4	Rear Surround L-R	Mid LSR 8 and Mid RSR 10

Table 6: Parameters of stereo reverberations.

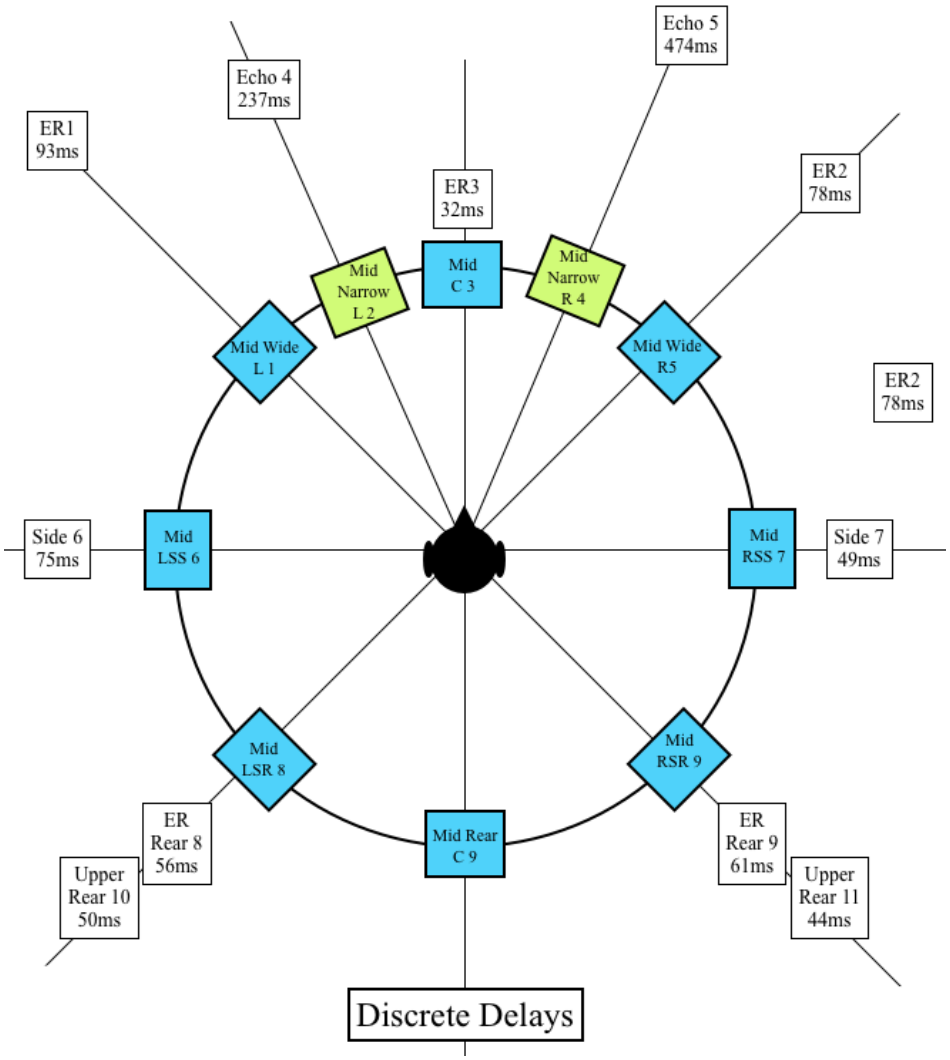


Figure 64: Positioning of discrete delays.

The front 'far' left and right positions were executed with two medium halls (light green, 1.3sec decay). The left was placed between Mid Wide L1, and Mid LSS 6, and the right between Mid Wide R5 and RSS 7. The side reverbs (dark purple) were executed with medium halls of 2.5sec decays. The left was positioned at Mid Wide L1 and Mid LSR 8, and the right at Mid Wide R5 and Mid RSR 10.

A long hall with a decay time of 1.4 seconds and a pre-delay of 53ms served to define the rear wall of the acoustic (lime green). It was assigned to Mid LSR 8 and Mid RSR 10. This reverb

completed the surround environment, and balanced the reverberant information arriving from all directions.

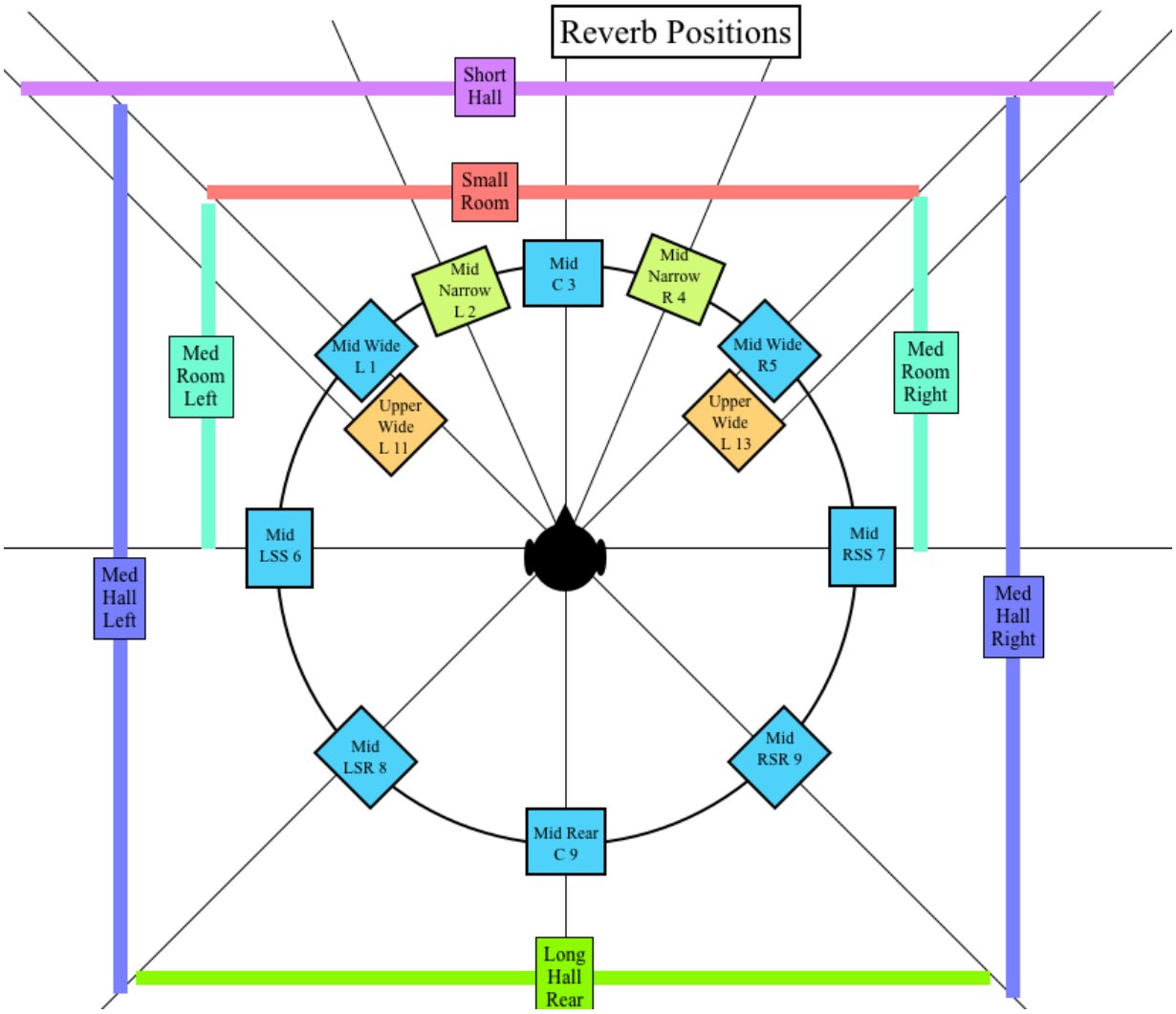


Figure 65: Reverberation positions.

3.4.7 Multi-Channel Impulse Response Layer

The final hemispherical layer was applied using McGill University's virtual acoustic technology. Space Builder is a 24-channel convolution reverb. It consisted of three multi-channel impulse responses that combined the measurements from two small churches and one small hall.

The estimated decay time was approximately 3.0 seconds. The Space Builder returns were routed to the 22 loudspeaker channels.

3.4.8 Mix Evaluation

An informal assessment of the mix as a three-dimensional presentation of the musical material was made by the graduate students and staff of the McGill Sound Recording Program. The resultant mix was judged to be successful, and listening tests were then carried out to gauge (1) Immersion, (2) 3-Dimensionality, and (3) to determine the expansion of the sweet spot to the left and right of the center mix position. The results were meant to establish a baseline for the continuing research into the three-dimensional presentation of popular music.

3.4.9 Test Subjects

Twenty-five test subjects were drawn from the students and staff of the graduate program in Sound Recording at McGill University, and audio professionals from the Montreal area. The subject pool was composed of individuals specializing in recording, production, and mixing. All had significant musical training.

Subjects were asked to listen to the final mix within the 22.2 environment, and answer a simple questionnaire rating *Immersion* and *3-Dimensionality* on a scale of 1 to 10, (‘1’ being minimum and ‘10’ being maximum). Subjects were also asked to determine the extent they felt the sweet spot extended to the left and right of the center listening position. This was done by moving around the area of the center listen position. Subjects were instructed to make a subjective assessment of at what point the musical presentation was not longer viable. Subjects were provided with white adhesive tape to mark these points on the mix desk.

3.4.10 Results and Analysis

The mean rating for *3-Dimensionality* was 7.22/10 with a standard deviation of 1.08 (Figure 66, left). The mean *Immersion* rating by the twenty-five subjects was 7.62/10 with a standard deviation of 1.09 (Figure 66, right).

The width of the sweet spot (Figure 67, Figure 68) extended to the left of the center position by a mean of 54.18 cm (22 subjects) with a standard deviation of 22.23 cm. The right-of-center extension of the sweet spot had a mean of 57.86 cm (21 subjects) with a standard deviation of 23.03 cm. It can be seen that there is asymmetry in the left and right estimations made by the test subjects. Contributing factors to this asymmetry could be source material within the mix itself, or the amount of spatial processing dedicated to source elements present in each side of the mix.

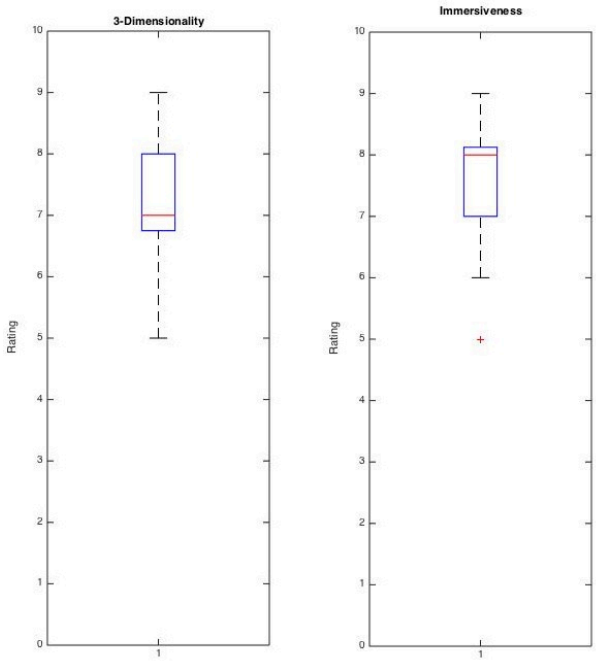


Figure 66: Subjects ratings of 3-Dimensionality and Immersion.

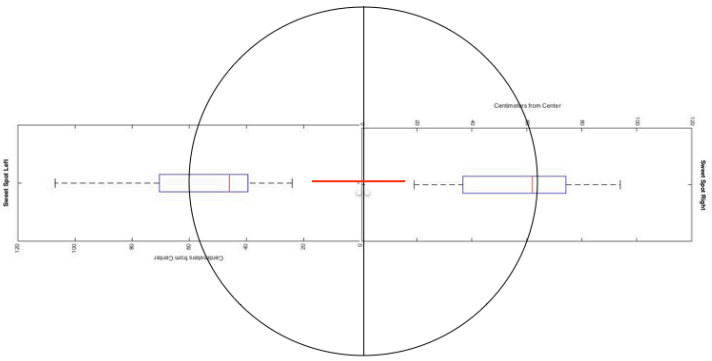


Figure 67: Sweet Spot: asymmetrical left-to-right.

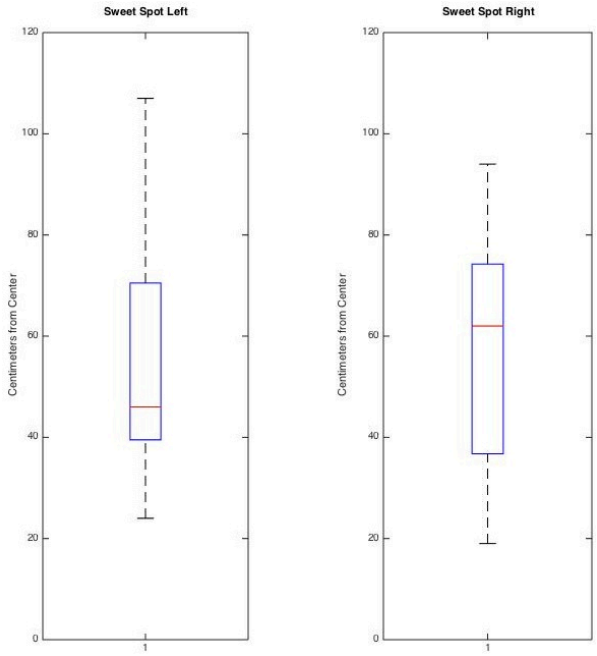


Figure 68: Sweet Spot Estimations, Left and Right.

3.4.11 Conclusions

The subjects' evaluation of *3-Dimensionality* and *Immersion* suggest that it is possible to create believable three-dimensional immersion using conventional stereo spatial tools in a 22.2 multi-channel playback environment from monophonic multi-track source material.

It was beyond the scope of this paper to answer the second research question: ‘How can conventional mix tools be implemented to expand monophonic sources into three dimensions?’. This will be addressed in a future paper. The subjects test ratings would suggest that it is possible.

Following the guidelines developed from the sited research, the subjects’ ratings suggest that it is possible to use spatial tools designed for stereo reproduction to create believable three-dimensional playback. The workflow, however, using conventional tools, is inefficient and time-intensive, and may not be commercially viable. (The development of the above techniques and the completion of an acceptible 3D music presentation took approximately 8 weeks of 8-10 hour days).

Regarding the fourth research question: ‘Can three-dimensional immersion beyond the ‘sweet spot’ of the mix position be expanded using one-dimensional spatial tools in a 22.2 multi-channel playback environment?’, the results appear to be mixed. Some subjects reported a very wide sweet spot, between 80-100cm of center, but the results here were wide ranging, and asymmetrical left-to-right. These results are inconclusive.

One of the primary obstacles that hindered the examination of the effectiveness of the created virtual acoustic was the channel limitation imposed by the Protools HD platform. The use of Space Builder required the 64 input channels allowed only in Protools HD. However, the number of voices (within the HD platform) needed for this investigation quickly exceeded the limit of the HD hardware. This necessitated the recording some of the virtual acoustic elements to audio tracks, and then switching to ‘Native’ operation, (which allowed unlimited channel count). Because many of the virtual acoustic elements were recorded containing multiple music sources, this negated the possibility of testing the effectiveness of the virtual acoustic construction via the subtraction of spatial components (discrete delays and reverberations) to determine where the ‘illusion’ of the immersive virtual space ‘broke down’.

An observation that surprised the author (who has decades of commercial production experience) was the ineffectiveness of many proven mix techniques employed in commercial two-channel stereo delivery.

Tools to increase productivity would include:

1. Simple, real-time non-rendering 3D panning tool.
2. Multi-channel (4-8 channel), real-time non-rendering equalization and compression processors.

3.4.12 Future Work

Future work to propel this area forward would include the development of studio recording techniques and practices, the continued development of multi-channel mixing tools for the manipulation of the source tracks, spatial processing handling early and late reflections, reverberation, and other effects processes.

In the current climate, much effort for 3D sound creation is being focused on post-processes to generate the immersive experience. It is the opinion of the author that a fundamental understanding of the basic principles derived from the practice of both mixing and recording in loudspeaker arrays with height channels will aid, enhance, and help define the development and architecture of future 3D tools.

The author has spent hundreds of hours mixing conventionally recorded studio tracks for the 22.2 playback environment and has identified the following list of insufficiencies in such a workflow [169-171]:

1. Mixes lacked realistic dimensionality and impact.
2. The created 3D images were often incoherent and broke down quickly outside the sweet-spot [169].
3. The sweet-spot itself was small [169, 170].

4. Upon repeated listening, the experience becomes less impressive, and faults become more noticeable.
5. The workflow is complex and impractical when working in the 22.2 Multichannel sound system as most 3D tools focus on delivering content with 7.1 or 9.1 beds neglecting bottom channels.
6. Current mix tools are insufficient for delivering a realistic and compelling three-dimensional volumetric audio images.

4 MICROPHONE ARRAYS FOR VERTICAL IMAGING AND THREE-DIMENSIONAL CAPTURE OF ACOUSTIC INSTRUMENTS

4.1 Abstract

This study compares a selection of microphone arrays for music recording to create vertical and three-dimensional images. The three-dimensional images are generated by routing direct signals from the microphones to discrete loudspeaker channels. Most published investigations addressing three-dimensional microphone arrays have focused on the capture of ensembles in live situations. These techniques prioritize the direct sound in the middle loudspeaker layer, ambience in the height layer and employ no bottom layer loudspeakers. Three separate arrays were investigated in this study: 1. Coincident, 2. M/S-XYZ, and 3. Non-coincident or Five-point capture. Solo instruments of the orchestral string, woodwind, and brass sections were recorded with each array. Two Triad/ABX listening tests were conducted to determine if the subjects could distinguish the arrays from a level-matched mono/point-source presentation, and if they could distinguish the arrays from one another. The results of the listening tests strongly suggest that the

subjects could discern the difference between the three arrays and the mono/point-source signal, and also from one another.

4.2 Introduction

This study compares a selection of microphone arrays for music recording to create vertical and three-dimensional images of single musical instruments. The direct signals from the microphones are routed to discrete loudspeaker channels in front of the listener. Most published investigations addressing three-dimensional microphone arrays have focused on the capture of ensembles in live situations. These techniques prioritize the direct sound in the middle loudspeaker layer, ambience in the height layer and employ no bottom layer loudspeakers (section 2.5). Notable exceptions are Hamasaki et al. [172] and Howie et al. [18].

One of the most daunting challenges in creating believable 3D mixes is expanding mono/point-source audio into a three-dimensional audio image. There is a need to develop techniques to expand the image of the captured source into a coherent vertical representation of the original source during playback. The method employed in this study was to assign direct sound (from microphones aimed directly at the source) to the high and low front channels of the 22.2 multi-channel playback system.

This study compares a selection of microphone arrays for music recording to create vertical and three-dimensional images without the use of added processing. These arrays expand upon the standard stereo practices (see section 2.4) of coincident and spaced pairs, as well as the M/S technique. This study focuses on single instrument capture in recording studio situations.

For the listening tests—in an effort to best understand the information collected by each array—the arrays are presented with the discrete microphone channels level-matched for equal-

volume. The individual arrays were also level-matched to one another for equal loudness. No effort was made to optimize the instrumental presentation of the arrays for aesthetic reasons.

Two Triad/ABX listening tests were conducted [137, 141, 173]. The aim of the tests was to determine:

1. If the subjects could distinguish the three arrays from a level-matched mono/point-source presentation.
2. If the subjects could distinguish the arrays from one another.
3. To determine if the subjects could perceive a vertical image.
4. To determine if these arrays provided a more immersive experience to the listeners than the mono signal.

4.3 Test Design and Methods

4.3.1 Microphone Arrays

The microphone arrays are presented in Table 7. The arrays were setup using either three, four, or five microphones.

Mic:	Position:	Microphone:
Coincident	Left	Schoeps CMC62U / MK 4
	Right	Schoeps CMC62U / MK 4
	Up	Schoeps CMC62U / MK 4
	Down	Schoeps CMC62U / MK 4
M/S-XYZ	Center	DPA 4011
	Horizontal Side	Schoeps CCM8
	Vertical Side	Schoeps CCM8

Non-Coincident	Center	DPA 4011
	Left	Schoeps CMC62U / MK 4
	Right	Schoeps CMC62U / MK 4
	High	Schoeps CMC62U / MK 4
	Low	Schoeps CMC62U / MK 4

Table 7: Microphones used in recording.

4.3.1.1 Coincident Array

The coincident array (Figure 69) consisted of four cardioid condenser types (Schoeps CMC62U/MK4). The configuration consists of two 90°-coincident stereo pairs oriented on the horizontal (Left/Right) and vertical (Up/Down) axes.

Figure 69: Coincident Array.

4.3.1.2 M/S-XYZ Array

The M/S-XYZ array (Figure 70) is an expansion of the M/S configuration to capture the *x, y* and *z* axes—with the second figure-8 capsule capturing the vertical axis. The mid-microphone was a DPA 4011 cardioid, and the vertical and horizontal figure-8 'side' microphones were Schoeps CCM8 types. (Matrixing from MS to XY for playback was performed according to standard practice [100, 108], see section 2.4.2.4.

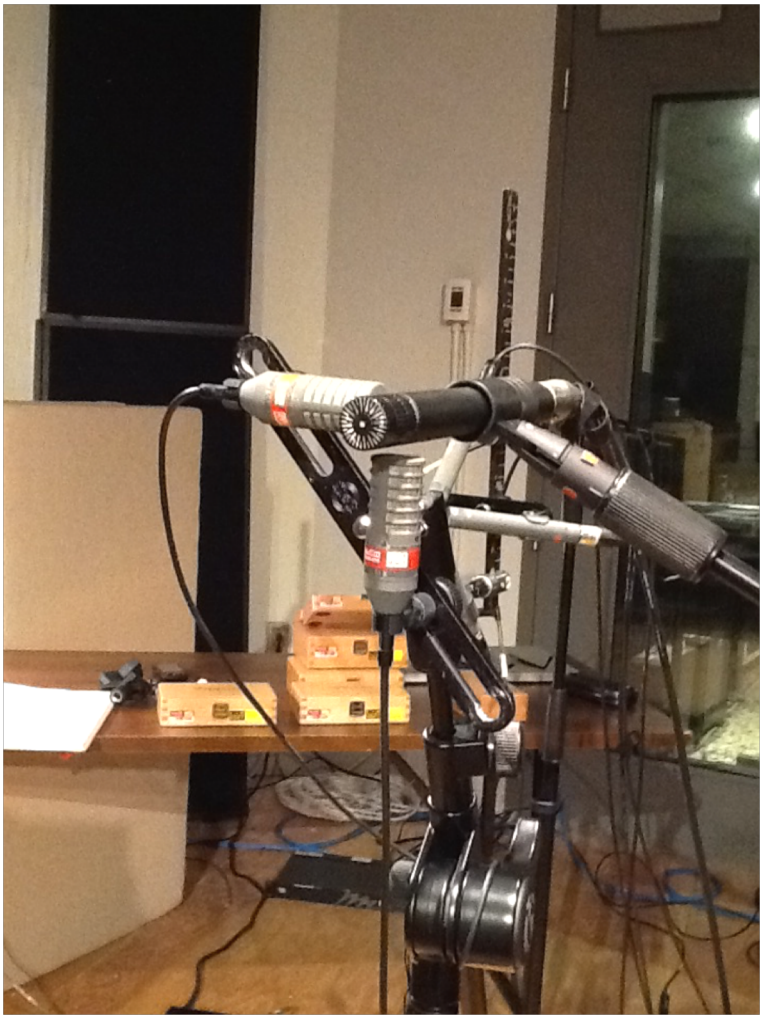


Figure 70: M/S-XYZ Array.

4.3.1.3 Non-coincident/Five-point Capture Array

The Non-coincident/Five-point capture array (Figure 71, Figure 72) used five cardioid condenser capsules arranged center, left, right, high, and low relative to the instrument. The center capsule was the DPA 4011 of the M/S-XYZ array, the other four were Schoeps CMC62U/MK4. The location and the angle of microphones were set for each instrument by the judgment of the recording engineer listening in the recording studio. All capsules were aimed at the instrument.

Figure 71: Microphone locations in Non-coincident/Five-point capture array for cello recording.

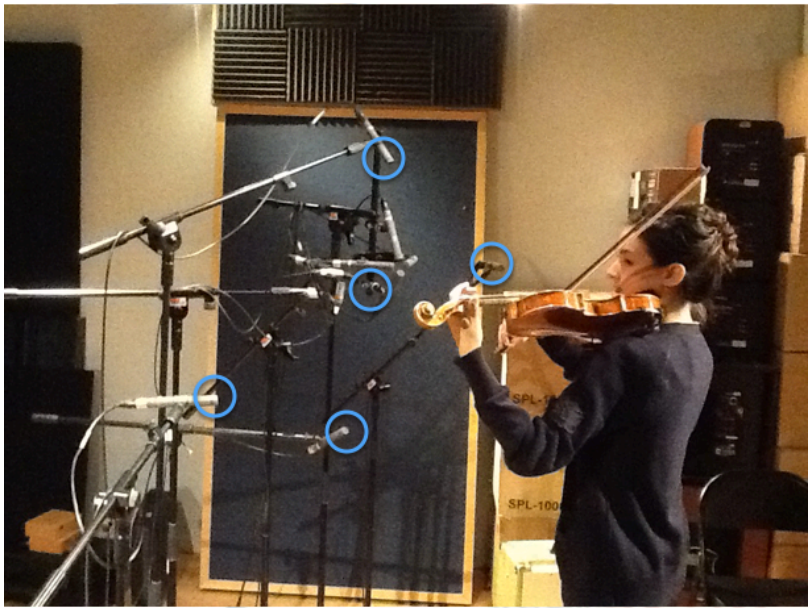


Figure 72: Non-coincident/Five-point capture array.

All microphones were recorded through a Sony SIU 100 interface, which provided microphone pre-amplification (DMBK-S101 cards) and A/D conversion. Careful attention was paid to match the input gains and the recording level of the microphones. The session was recorded at a 96kHz sample rate. Recording was made directly into Avid Protools 10 DAW.

Figure 73, Figure 74 and Figure 75 displays the frequency response of an individual Schoeps MK4 microphone in: the coincident 4 microphone array, in a stereo pair, and a single microphone. The measurements were taken in a semi-anechoic chamber.

While Figure 73 (coincident 4-mic array) displays a boost around the 10kHz region compared with the dual and single capsule measurement, this response was not found to interfere with the recordings used in this research. It should be stated that coincident microphone arrays have been used since the inception of stereo in the 1950s, and have been employed in countless recordings with great success. The use of four microphones in the coincident array was deemed of sufficient sound quality for this investigation.

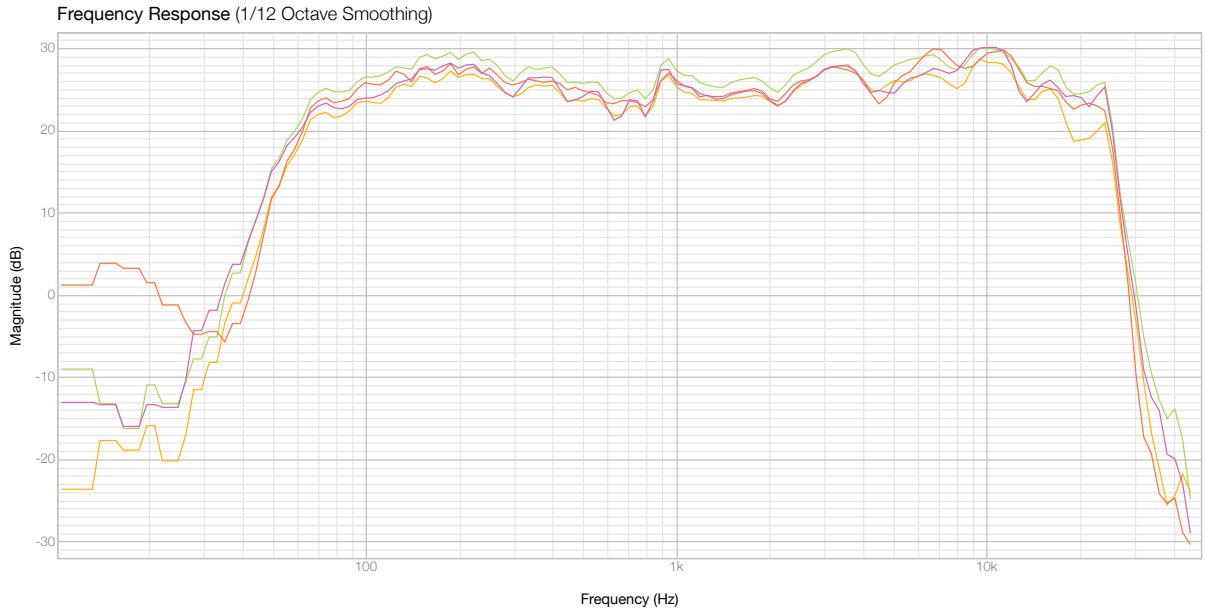


Figure 73: Measurement of Schoeps MK4 capsules in coincident 4-mic array.

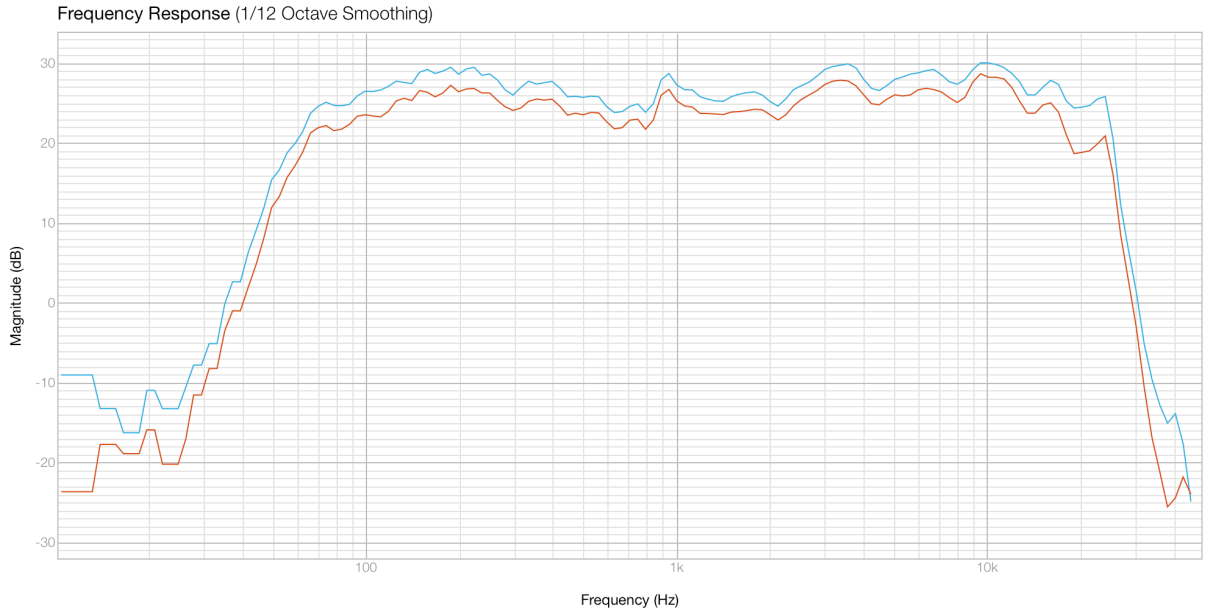


Figure 74: Measurement of Schoeps MK4 capsules in coincident stereo pair. (While the response curves are almost identical, the difference in level is due to the fact that the measurements were taken separately and overlayed at a later date),

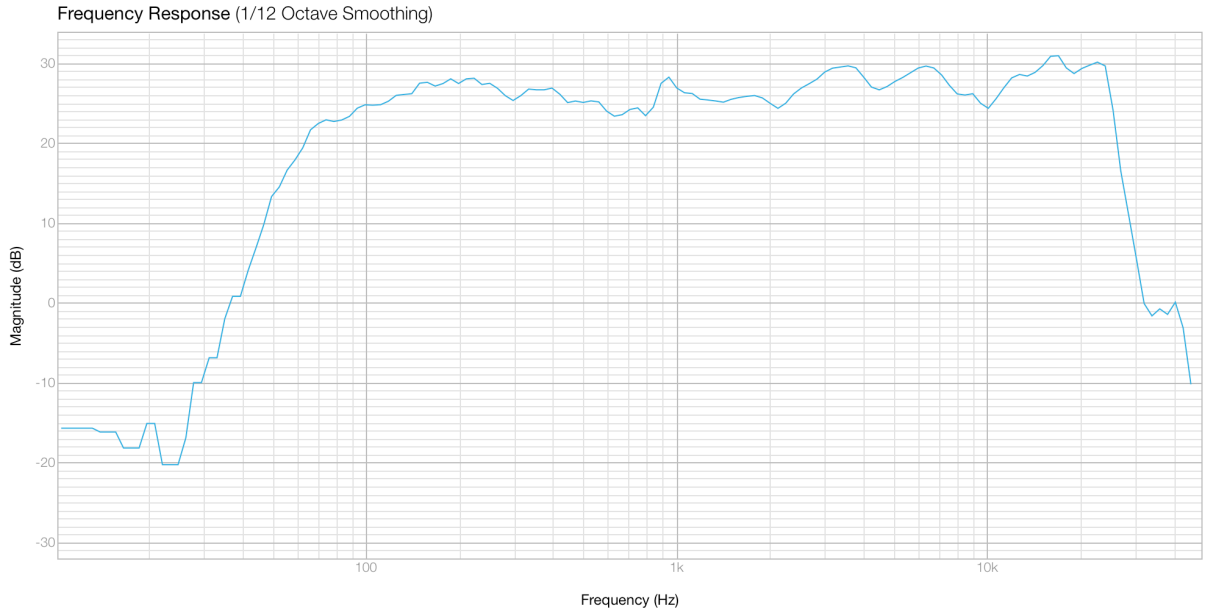


Figure 75: Measurement of Schoeps MK4 single capsule.

4.3.2 Recording Studio

See section 3.3.3.2.

4.3.3 Control Room Monitor Environment

This research was conducted at McGill University in the Schulich School of Music's Studio 22. See section 3.1.

4.4 Array Level Matching

4.4.1 Intra-Array Level Matching

Intra-array level matching of the discrete microphone channels was achieved by the normalization function within Protools. Channels were peak-normalized to -3dB 0dBFS. This

process was judged to be effective due to the fact that each channel contained the same recording, albeit a different perspective. The accuracy of this process was verified by checking each channel after normalization in an LUFS meter integrated over 20 seconds. Results were typically ±0.2 LUFS.

4.4.2 Inter-Array Level Matching

The level matching between the individual arrays (and also the center mono channel which is used as a control) was achieved by playing back signals from each array through loudspeakers in the control room and capturing the acoustic signal in the control room via a Neumann KU-100 binaural head located at the mix position. The binaural input was monitored through an LUFS meter with an integration window of 20 seconds. The playback levels of each array and the mono control were adjusted to an accuracy ±1dB. Loudness matching was also judged by three experienced engineers and deemed consistent. No participants of the listening tests reported any loudness differences.

4.5 Test Subjects

The twenty-two listening test participants were drawn from the students and staff of the graduate program in Sound Recording at McGill University, and the graduate and undergraduate performance program of the Schulich School of Music. All subjects had significant musical training in performance, and/or music recording and production, (Figure 76, Figure 77, Figure 78, Figure 79).

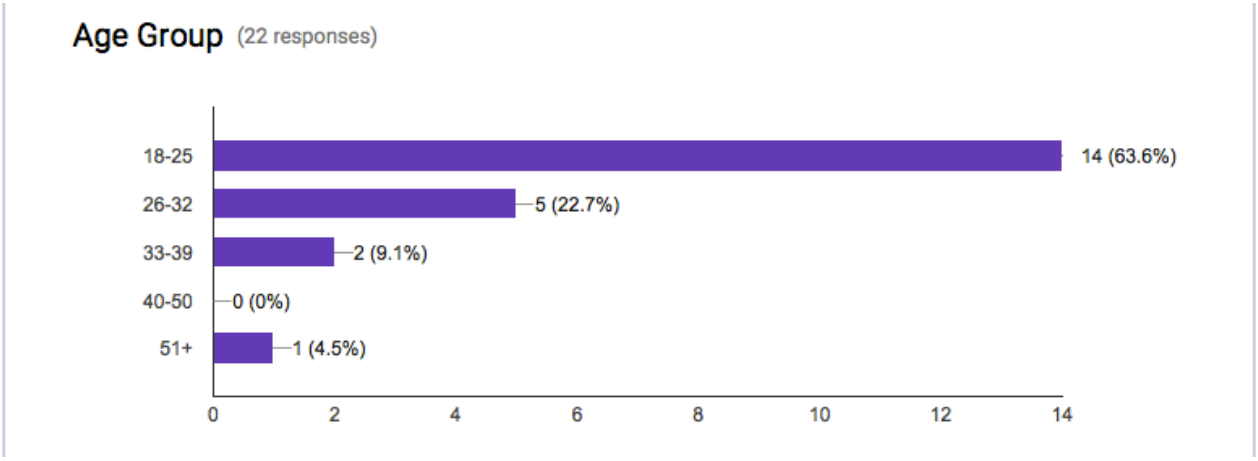


Figure 76: Age of subjects.

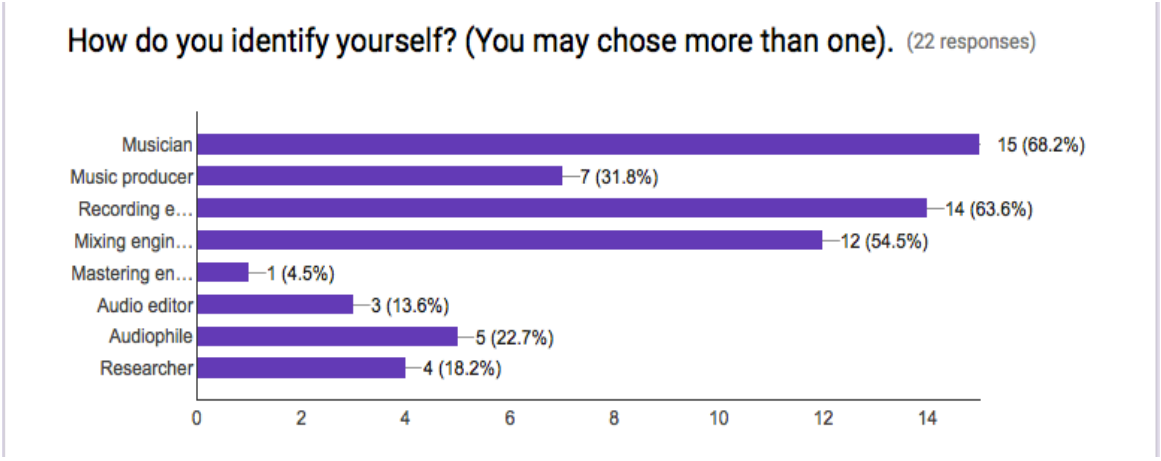


Figure 77: How subjects identify themselves.

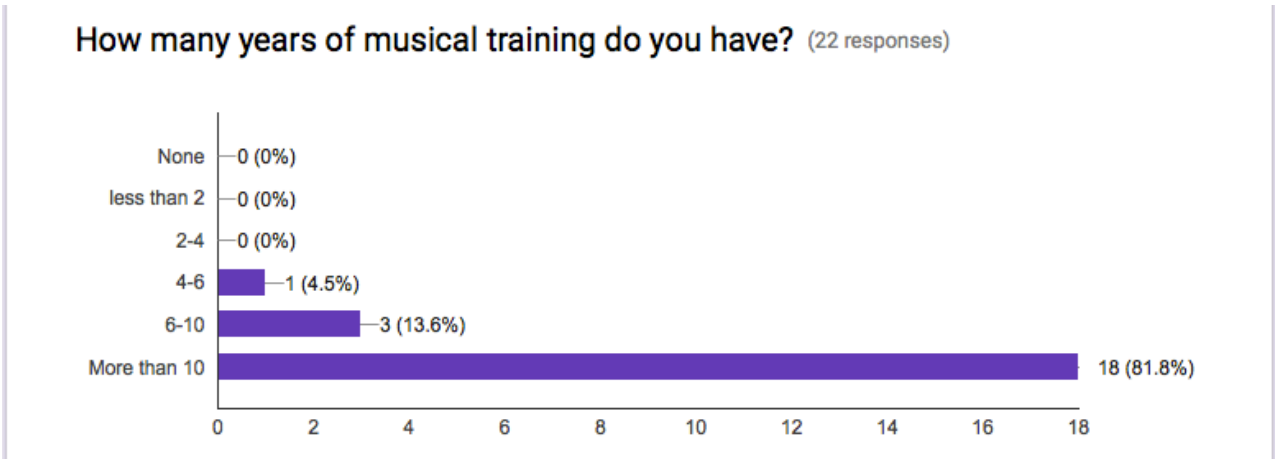


Figure 78: Subjects' years of musical training.

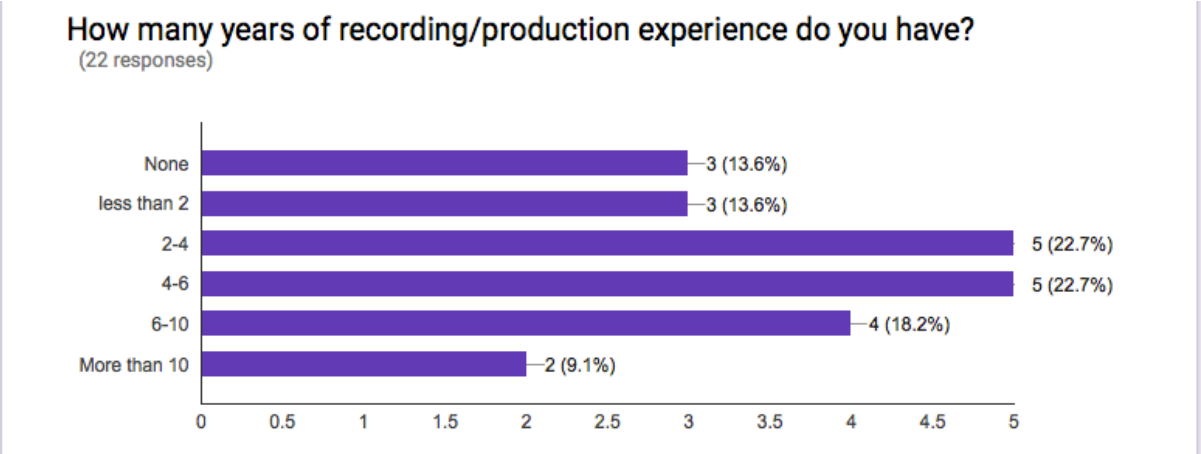


Figure 79: Subjects' years of audio experience.

4.6 Test Material

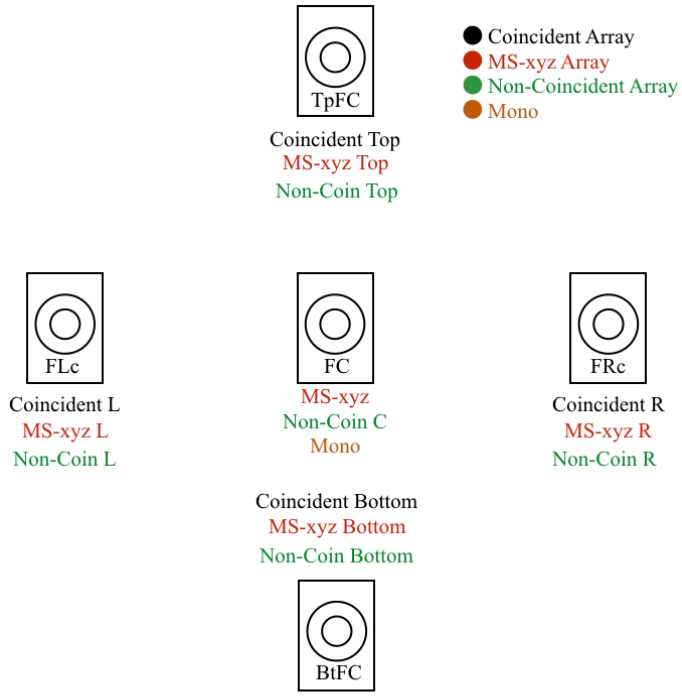
Orchestral instruments were recorded individually with each of the described microphone arrays. During recording the instruments were positioned in the center of the recording studio (3.3.3.2). An attempt was made to draw instruments from each family, to select those possessing both simple and complex radiation patterns, and to cover all registers [174, 175]. Other factors that determined selection were the quality of the performance, quality of the instrument, and the stability of localization of the recorded image (determined by performer's movement during recording). The instruments selected were:

1. Brass: Trumpet. Mid to high register; largely simple radiation pattern from the bell.
2. Strings: Viola. Mid to high register; complex radiator.
3. Strings: Cello. Low to mid register; complex radiator.
4. Woodwinds: Tenor Saxophone. Low to mid register; complex radiator.

4.7 Test Procedure

4.7.1 Playback of Microphone Arrays for Testing

Each microphone within the array was assigned to a discrete loudspeaker in the test environment corresponding to the region of capture during recording (Figure 80). The 'M' microphone from the MS-XYZ array was used for the mono signal.



Microphone-to-Loudspeaker playback for testing

Figure 80: Microphone-to-loudspeaker position during tests.

4.7.2 Test Familiarization and Training

The test was conducted in four stages:

1. Familiarization and training.
2. A training experiment.
3. The main experiment.
4. Questionnaire for demographic and qualitative observations.

The familiarization stage consisted of oral instructions of the Triad/ABX test, the test interface, and listening to the test signals. Pairwise comparisons were made with a simple Triad/ABX test. Subjects were seated in the mix position of Studio 22. For each trial, a musical excerpt was played on a continuous loop. Playback of the different arrays was time aligned due to the simultaneous recording of all arrays, which provided for seamless switching between stimuli. The test took approximately 30 minutes to complete, which is in line with ITU recommendations for the similar "double-blind triple stimulus with hidden reference" methodology [131].

The training experiment was to allow the subjects to familiarize themselves with the test interface and the testing procedure. It provided a simpler version of the main experiment. The task was to learn how to differentiate between a level-matched mono signal and one of the arrays. The training test consisted of six trials. The array configurations and pickup patterns were explained. The instruments' sound radiation patterns were explained. The subjects were not instructed to ignore differences in timbre nor to focus on spatial differences.

The main experiment consisted of 24 trials. The task was to differentiate the arrays from one another. Each listener was allowed to adjust the playback volume. The average listening level was 76dB.

4.8 Results

4.8.1 Results of Training Tests

The training test consisted of the six trials shown in Table 8. The task was to distinguish the level-matched mono signal from one of the arrays.

Trial:	Instrument:	Mono	Coincident	Cross	MS-xyz
1	Viola	x2	x		
2	Tenor	x2		x	
3	Trpt	x2			x
4	Viola	x	x2		
5	Tenor	x		x2	
6	Trpt	x			x2

Table 8: Training test trials.

All subjects scored 100% correct on trials 1, 2, 4, 5 and 6. The exception was Trial 3 where the trumpet was presented, 2 x mono to 1 x MS-XYZ. Here 20 subjects (90.9%) chose correctly, with 2 (9.1%) choosing incorrectly (Figure 81).

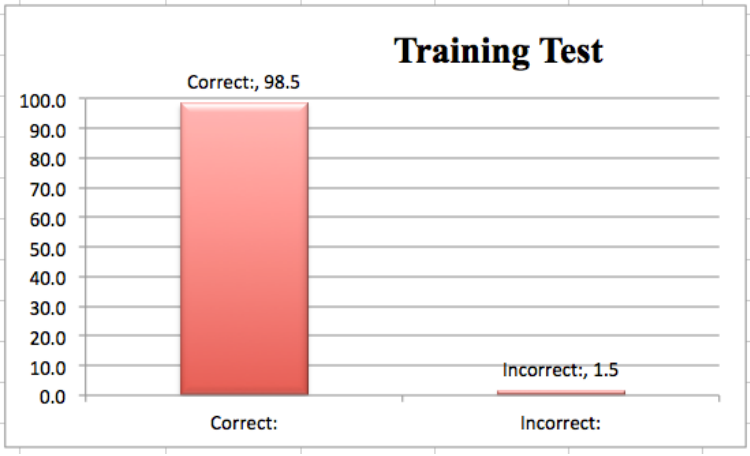


Figure 81: Results of training tests.

4.8.2 Results of Main Test

The main test consisted of the twenty-four trials shown in Table 9. The task was to distinguish the microphone arrays from one another.

TRIAL #:	Instrument:	Coincident	Cross	MS-xyz
1	Viola	x2	x	
2	Cello	x2		x
3	Tenor Sax	x	x2	
4	Trpt		x2	x
5	Viola	x2		x
6	Cello	x	x2	
7	Tenor Sax		x2	x
8	Trpt		x	x2
9	Viola	x	x2	
10	Cello		x2	x
11	Tenor Sax	x		x2
12	Trpt	x2	x	
13	Viola		x2	x
14	Cello	x		x2
15	Tenor Sax		x	x2
16	Trpt	x2		x
17	Viola	x		x2
18	Cello		x	x2
19	Tenor Sax	x2	x	
20	Trpt	x	x2	
21	Viola		x	x2
22	Cello	x2	x	
23	Tenor Sax	x2		x
24	Trpt		x2	x

Table 9: Main test trials.

The results of the main test were similar to those of the training test with 98.82% correct in distinguishing the arrays from one another with only 1.18% incorrect identifications in the Triad/ABX test (Figure 82).

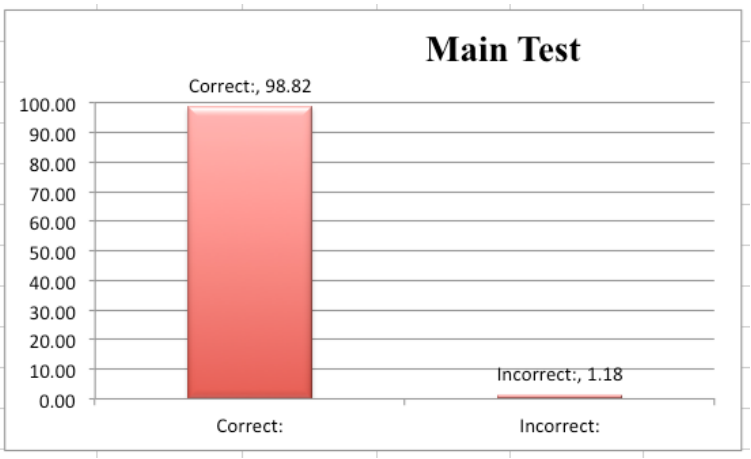


Figure 82: Results of main test

4.8.3 Questionnaire for Qualitative Observations

The final stage of testing was for the subjects to answer a questionnaire to provide demographic information and qualitative information about the test itself, and the observed characteristics of the microphone arrays overall.

4.8.4 Subject Rating of the Training and Main Test

To determine the effectiveness of the familiarization and training portion of the test, the subjects were asked to offer their observation on its efficacy.

Twenty (90.9%) of the subjects reported that the training helped as opposed to only two (9.1%) reporting that it did not (Figure 83).

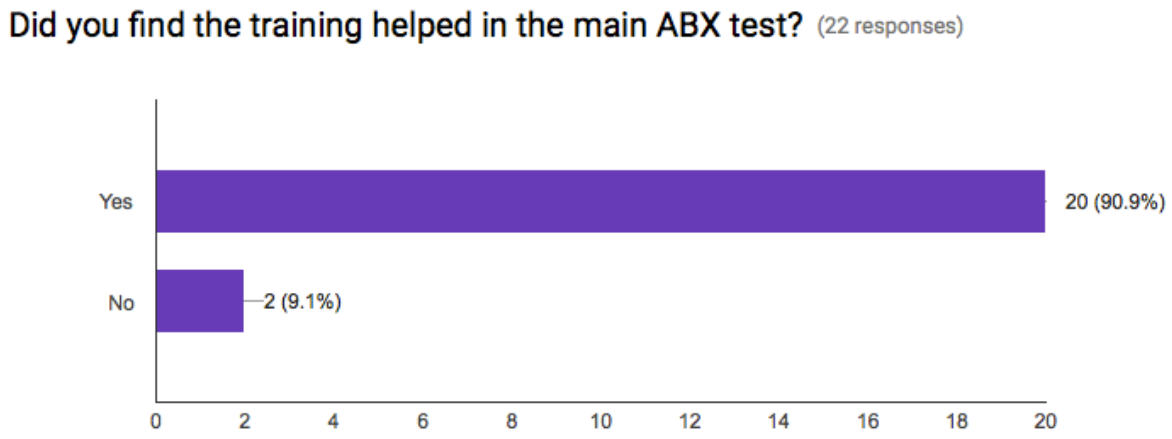


Figure 83: Subject rating of training test.

The subjects were asked to rate the difficulty of the Triad/ABX test on a 1-5 scale; 1 being 'easy' and 5 'difficult'. Sixteen (72.7%) rated the test as 'easy' (1), and six (27.3%) subjects rated it as '2' (Figure 84). None rated it as moderate of difficult.

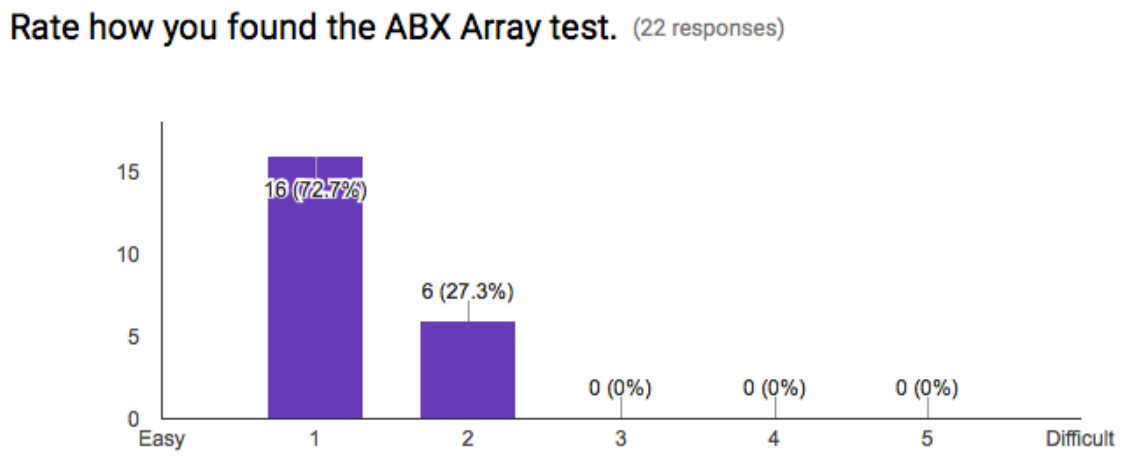


Figure 84: Subject rating of main Triad/ABX test.

4.8.5 Subjects Qualitative Observation of Arrays

This study was not designed to extract qualitative information about the individual arrays. But it was decided that some useful information might be gathered once the subjects had completed the testing.

Four qualitative questions were asked concerning the arrays as a whole. Observations were rated on a 1-5 scale. The questions were:

1. Do you feel the microphone arrays provided a vertical audio image compared to the mono/point-source audio? (‘1’ being *No Verticality*, and ‘5’ being *Extended Verticality*). Figure 85.
2. Do you feel the microphone arrays provided a greater depth of image compared to the mono point-source audio? (‘1’ being *Same depth as mono*, and ‘5’ being *Greater depth than mono*). Figure 86.
3. Do you feel the microphone arrays provided a more immersive image compared to the mono point-source audio? (‘1’ being *Same immersion as mono*, and ‘5’ being *Greater immersion than mono*). Figure 87.
4. Would you provide a general impression of the size of the images provided by the microphone arrays in comparison to the mono/point-source audio. (‘1’ being *Small*, and ‘5’ being *Large*). Figure 88.

There was a ‘Comments’ section provided at the end of the questionnaire, and three of the twenty-two subjects stated that they were not thinking about the qualitative aspects of the arrays but were focusing on the ABX nature of the test. A few other subjects also mentioned this to the author after completion of the tests.

Figure 87 displays the results of the question concerning the verticality of the image presented by the microphone arrays as a whole compared to the level-matched mono/point source image. 86.4% of the subjects rated the array images having an extended vertical image compared with the mono source as opposed to 13.6% rating them the same. 77.3% rated the verticality in the 3-5 range.

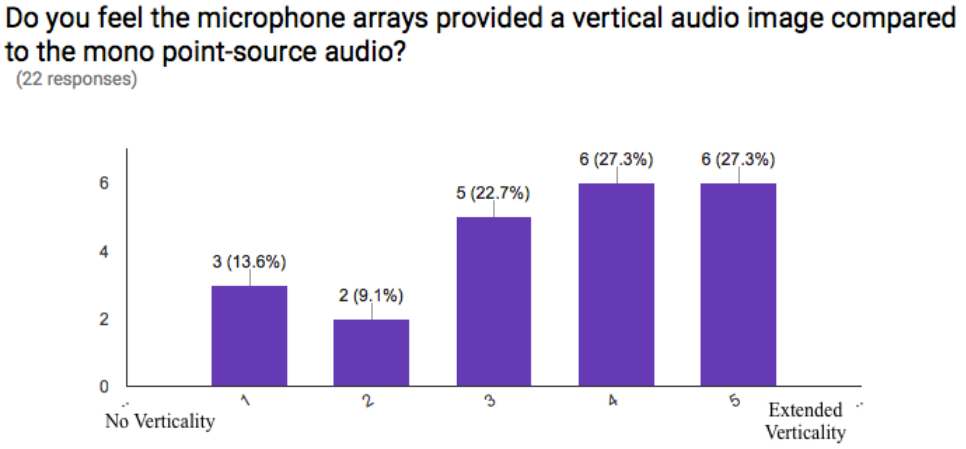


Figure 85: Subjects rating for audio image verticality.

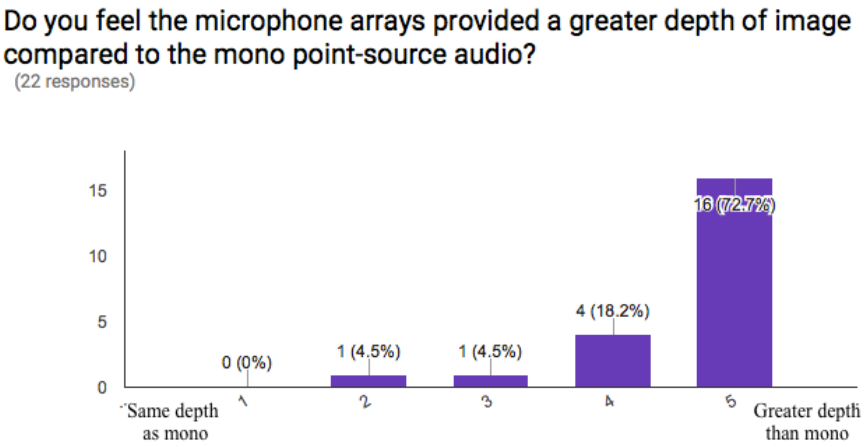


Figure 86: Subjects rating for depth of audio image.

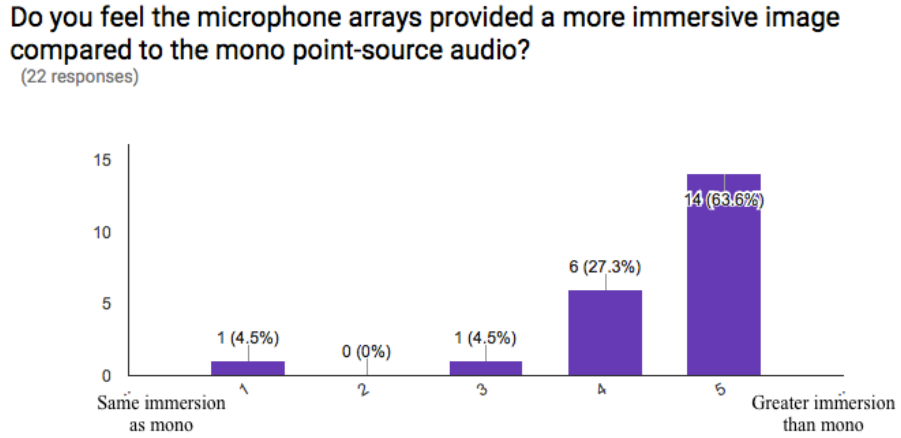


Figure 87: Subjects rating for immersion of array image compared to mono image.

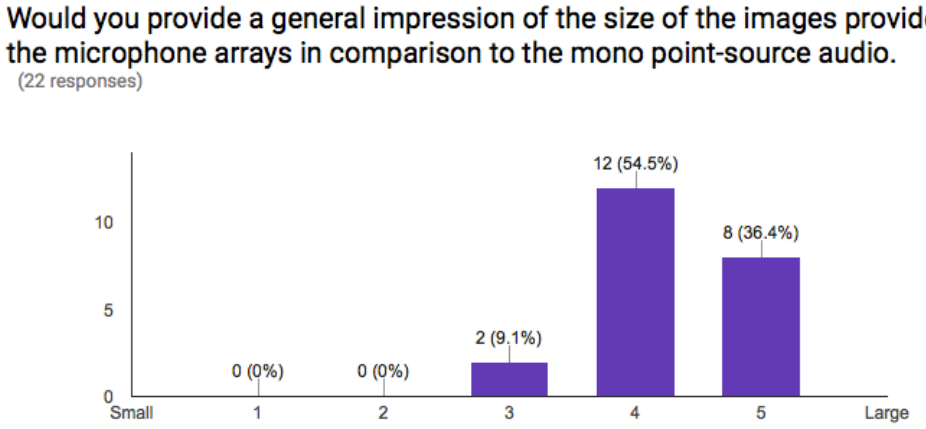


Figure 88: Subjects rating for size of audio image.

4.9 Discussion of Results

The effectiveness and value of the training of subjects has been reported by many authors, such as Bech [130, 176], Olive [129], and is detailed in ITU-R BS.1116-1 [131]. It has also been reported that training can elevate the prowess of subjects and improve their reliability in testing. Most of the subjects had participated in previous listening tests conducted by the Sound Recording department, but five were orchestral musicians who had never participated in testing before. The clarity and significance of the data coupled with the ratings provided by the subjects does support the value of training and familiarization with test procedures and methods.

4.9.1 Training Test

The subjects discerned the mono/point-source signal from the arrays in 98.5% of the trials. The data strongly suggests that a listener can easily identify a sound source containing extra-dimensional information from a mono source when the signals are played through loudspeaker systems that include height channels.

4.9.2 Main Test

The subjects discerned the different arrays from each other in 98.82% of the trials. The data strongly suggests that a listener can easily identify the arrays from one another when the signals are played through loudspeaker systems that include height channels, and the arrays are assigned from each microphone to a loudspeaker that roughly approximates its position in the original capture.

4.9.3 Perceived Verticality of Audio Image Captured by Arrays Compared to Mono

The majority of the subjects (86.4% overall, 77.3% in 3-5 range) rated the array images having an extended verticality over the mono source with only 13.6% rating them the same (Figure 85). Even in this informal survey of subjects, the results suggest that the 3D microphone arrays do provide vertical imaging of the captured audio.

4.9.4 Perceived Depth of Audio Image Captured by Arrays Compared to Mono

90.9% of the subjects rated the arrays in the 4-5 range as having greater depth than the mono image. No subjects rated the mono images of equal depth (Figure 86). The results suggest that the 3D microphone arrays do provide greater depth of image than the mono capture.

4.9.5 Perceived Immersion of Audio Image Captured by Arrays Compared to Mono

90.9% of the subjects rated the arrays in the 4-5 range as having greater immersion than the mono image. One subject rated the mono images of equal immersion to the arrays (Figure 87). The results suggest that the 3D microphone arrays do provide more immersive image than the mono capture.

4.9.6 Perceived Size of Audio Image Captured by Arrays Compared to Mono

90.9% of the subjects rated the arrays in the 4-5 range as having greater size than the mono image. No subjects rated the mono images of equal depth (Figure 88). The results suggest that the 3D microphone arrays do provide greater size of image than the mono capture.

Although the subjects had not been asked to examine the arrays for three-dimensional criteria prior to testing, their combined observations point to good performance by the arrays in capturing vertical imaging and enhanced dimensionality when played through loudspeaker systems that include height channels.

4.10 Conclusion

One of the goals of this study was to lay the foundation for future work on the development of microphone arrays for 3D and immersive capture. The significance of the results would establish the subjects' ability to differentiate the arrays from mono and the arrays from one another, which is what the author sought to establish. The significance of the vertical and lateral information was an added bonus, and points the way toward future testing, and array development.

4.11 Future Work

Understanding the methods and techniques involved in capturing audio in three dimensions is in its infancy. There is much work to be done. Ongoing work in this area include:

1. Qualitative perceptual testing of the above arrays using spatial descriptors to better understand the information offered by each array.
2. Testing of the above arrays that involves the subjects drawing their perception of the arrays on *vertical/horizontal* and *depth* grids to better understand the spatial information offered by each array.
3. Improving the microphone arrays for increased precision.
4. The examination and possible expansion of the lower channels of the 22.2 multi-channel sound system.
5. The development of mixing tools for immersive/3D audio.

5 Subjective Graphical Representation of Microphone Arrays for Vertical Imaging and Three-Dimensional Capture of Acoustic Instruments

5.1 Abstract

This investigation employs a simple graphical method in an effort to represent the perceived spatial attributes of three microphone arrays designed to create vertical and three-dimensional audio images. Three separate arrays were investigated in this study: Coincident, M/S-XYZ and Non-coincident/Five-point capture. Solo instruments of the orchestral string, woodwind, and brass sections were recorded. Test subjects were asked to represent the spatial extent of the perceived audio image on a horizontal/vertical grid and a graduated depth grid, via a pencil drawing. Results show that the arrays exhibit a greater extent in every dimension—vertical, horizontal and depth—compared to the monophonic image. The statistical trends show that the spatial characteristics of each array are consistent across each dimension. In the context of immersive/3D mixing and post-

production, a case can be made that the arrays will contribute to a more efficient and improved workflow due to the fact that they are easily optimized during mixing or post-production.

5.2 Introduction

Research into human perception of spatial audio has been increasing rapidly in the last few years due to the emergence of virtual and augmented reality technologies. It has been asserted that the ambiguity of language may not provide the most efficient means to interpret subjective assessment of spatial sound reproduction [149]. The exploration of a nonverbal means to interpret how we perceive spatial imagery has been employed as a mechanism that may more closely represent our perceptions as opposed to verbal descriptors [150-153].

This investigation utilizes a simple graphical method in an effort to represent the perceived spatial attributes of three microphone arrays (designed to create vertical and three-dimensional audio images), and a mono/point-source signal. Test subjects were asked to represent the spatial attributes of the perceived audio image on a provided horizontal/vertical grid and a graduated depth grid, via a pencil drawing. The subjects were positioned in the listening/mix position in 22.2 playback environment configured according to the ITU-R BS.2051-0 [62].

Three separate arrays were investigated in this study:

1. Coincident.
2. M/S-XYZ (also called Double MS-Z [121].
3. Non-coincident or Five-point capture configurations.

Solo instruments of the orchestral string, woodwind, and brass sections were recorded. The subjects were positioned in the listening/mix position in 22.2 playback environment configured according to the ITU-R BS.2051-0 [62].

5.3 Test Design and Methods

See section 4.3

5.4 Array Level Matching

See section 4.4

5.5 Test Subjects

Ten test subjects were drawn from the students and staff of the Graduate Program in Sound Recording at McGill University. Subjects had, on average, over ten years of musical training, over six years of training in music recording and production, and all had experience working in situations that included height channels.

5.6 Test Material

Orchestral instruments were recorded individually with each of the described microphone arrays. An attempt was made to draw instruments from each family, to select those possessing both simple and complex radiation patterns, and to cover all registers [174, 175]. Other factors that determined selection were the quality of the performance, quality of the instrument, and the stability of localization of the recorded image (determined by performer's movement during recording). The instruments selected were trumpet, viola, cello, and tenor saxophone.

5.7 Test Procedure

5.7.1 Playback of Microphone Arrays for Testing

See Section 4.7.1

5.7.2 Test Familiarization and Training

The test was conducted in three stages:

1. Familiarization with test material.
2. Oral explanation of the horizontal/vertical grid and depth grid, and the task of graphically representing the audio image.
3. Questionnaire for demographic and qualitative observations.

The familiarization stage consisted of oral instructions of the layout of the horizontal/vertical grid (Figure 89), the depth grid (Figure 90) and listening to the test signals. The loudspeakers used for playback provided a physical reference for the graphical representations drawn by each subject. The playback loudspeaker positions were: -30º left, +30º right, +45º high, and -30º low, referenced to the centre loudspeaker. The depth grid was rated on a scale of 1-10.

The experiment consisted of 16 trials (Table 10): one for each iteration (Mono-point/source, Non-coincident/Five-point, Coincident, MS-XYZ) of the four instruments (trumpet, viola, tenor saxophone, and cello) presented.

The task was to provide a pencil drawing of the outer edges of the perceived audio image on the horizontal/vertical grid, and the perceived depth of the image on the depth grid. The loudspeakers in Figure 80 are mirrored on the Horizontal/Vertical Grid. The depth grid has no specific units, but only there to provide a subjective scale for the perceived depth. Each listener was allowed to adjust the playback volume. The average listening level was 76dB SPL.

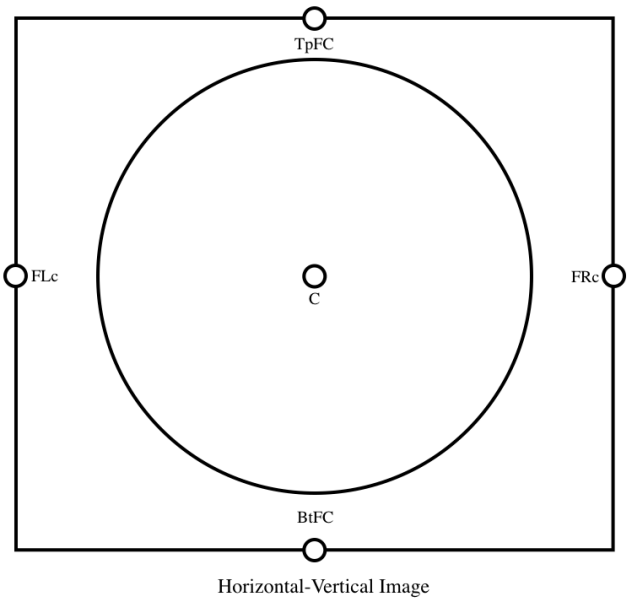


Figure 89: Horizontal/Vertical Grid.

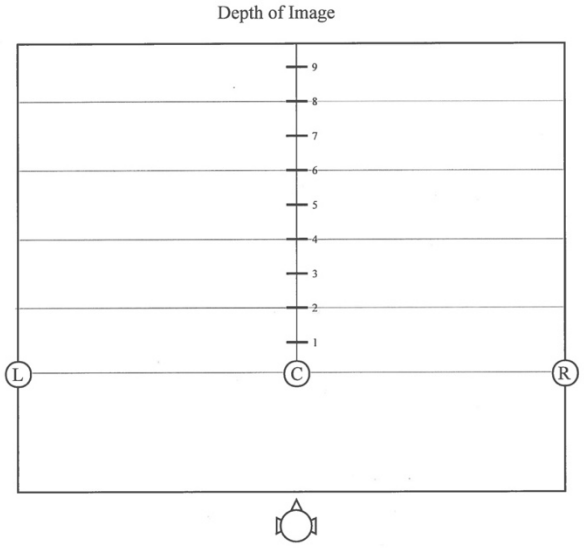


Figure 90: Depth Grid.

Trial:	Instrument/Array
1	Cello mono
2	Trpt Non-Coincident
3	Cello Coincident
4	Viola MS-xyz
5	Trpt Coincident
6	Tenor MS-xyz
7	Viola Mono
8	Cello MS-xyz
9	Tenor Non-Coincident
10	Trpt MS-xyz
11	Cello Non-Coincident
12	Tenor Coincident
13	Viola Coincident
14	Tenor Mono
15	Viola Non-Coincident
16	Trpt Mono

Table 10: Drawing trials

5.8 Results

5.8.1 Graphical Results of Test

The averaged maximum extents from the array drawings are mapped in Figure 91. The results of the drawing trials are shown in 5.12. Figures in 5.12.1 show the results of the 'Mono' trials, Figures in 5.12.2 show the drawings of the Coincident arrays, Figures in 5.12.3 the Non-coincident arrays, and Figures in 5.12.4 display the MS-XYZ arrays. The agglomerate drawings of each array and the mono playback are shown in 5.12.5.

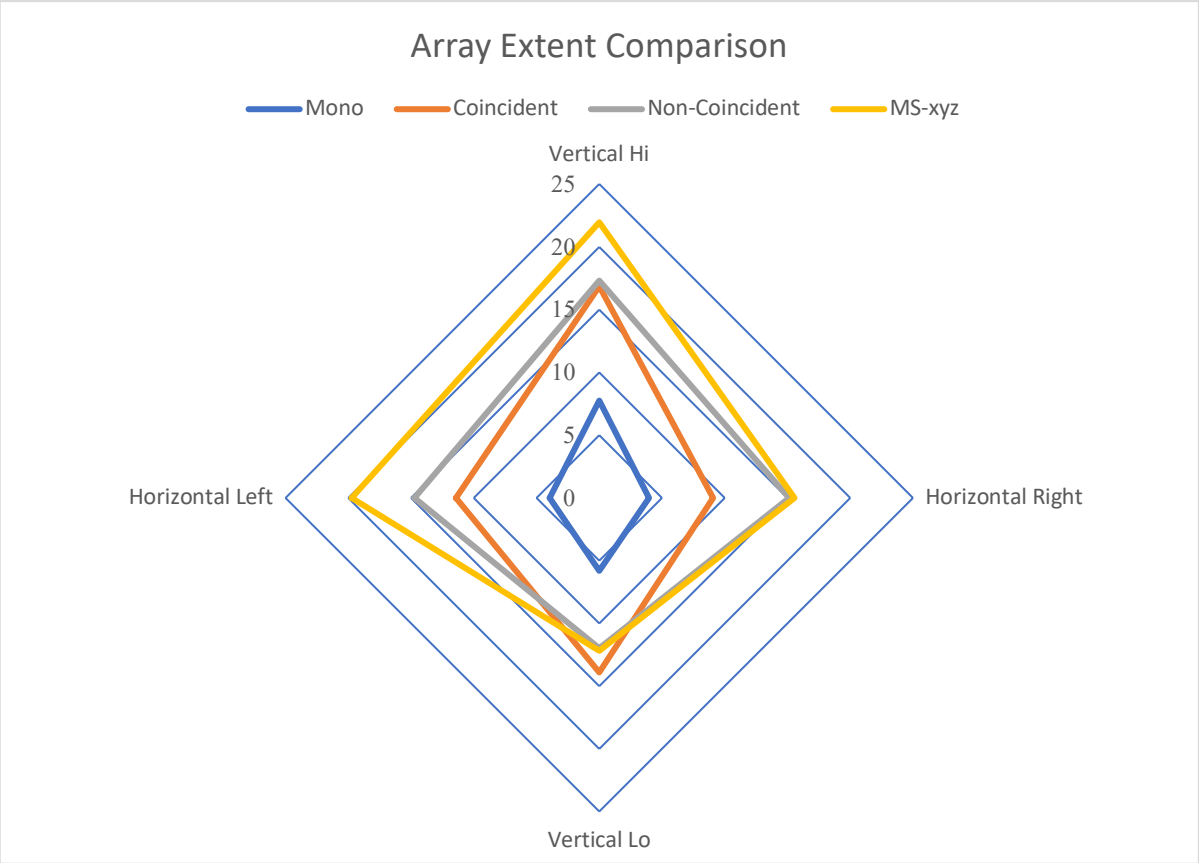


Figure 91: Averaged maximum extent comparison between microphone arrays.

5.8.2 Assessment of Test by Subjects

In the questionnaire completed after the test, the subjects were asked if they felt their drawings were a 'good' representation of the perceived audio image presented by the loudspeaker playback. All ten of the subjects answered in the affirmative.

The subjects were also asked to rate the difficulty of the test on a 1-5 scale (Figure 92). 90% of the subjects rated the test moderate '3' to easy '1', with only one subject rating it difficult '4'.

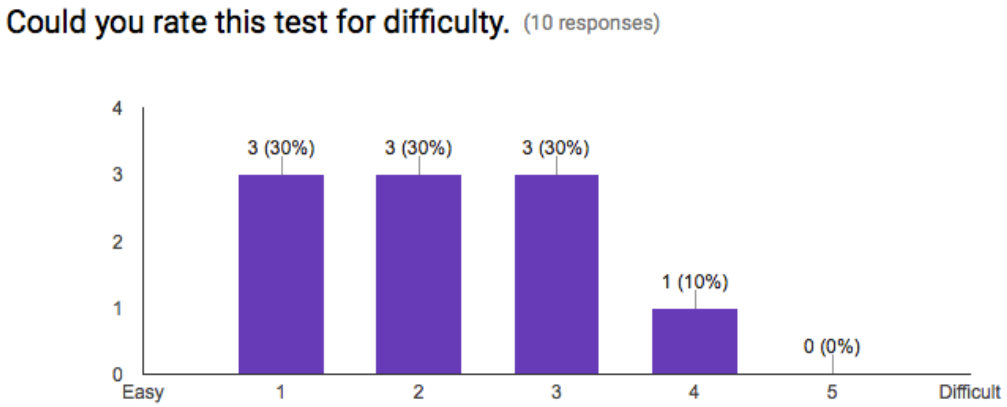


Figure 92: Subjects rating of test difficulty.

5.9 Discussion and Analysis of Results

5.9.1 Discussion of Maximum Extents

Figure 91 shows that the three microphone arrays deliver a greater horizontal and vertical extent than a single monophonic image. Conventional close capture techniques used in studio recording—employing one or two microphones—have proven adequate for mono or stereo reproduction. Within this paradigm the sound image typically captures more than enough information for an appropriately sized sound image, but there is little suggestion in the current study that it provides the information required to create a vertical audio image.

A complete sonic 'picture' is comprised of a sound source and the accompanying acoustic. The fact that the majority of musical instruments do not radiate sound in all directions with equal intensity [174] coupled with the complexity of the acoustic of the recording space [177] presents a sound image that is both spectrally complex and dynamic in time and timbre. This reality limits the amount of information that can be delivered by close capture techniques that employ one or two microphones.

The complex radiation patterns of musical instruments may contribute to the assymetrical horizontal- left and right extents exhibited by the Coincident and MS-XYZ arrays (Figure 91). In these arrays the capsules are in close proximity and are capturing the same area of radiated instrument-sound and room acoustic. The Non-coincident array does not exhibit this asymmetry. Here the capsules are more widely spaced and capture a larger picture of the instrument itself and a broader picture of the room acoustic. This increased instrument and acoustic information may provide a more averaged/symmetrical audio image.

The microphone arrays investigated in this study appear to provide a more complete sonic picture than delivered by current single and stereo microphone techniques.

5.9.2 Discussion of Graphical Representations

5.9.2.1 Mono/Point-Source Representation

Examining the graphical representations of the mono playback (Figure 98, Figure 99, Figure 100, Figure 101, and Figure 114) it can be seen that the reproduced audio image is primarily localized at/and confined to the centre loudspeaker. The author's theory for the representations extending down toward the BtFC loudspeaker are early reflections from the desk/console located in front of the mix/listening position.

The subjects' representations of the mono images confined to specific loudspeakers are in agreement with author's findings in [169]. In this investigation, it was found that mono/point-sources require extensive processing when up-mixing to create 3D audio images.

5.9.2.2 Coincident Array

The representations of the coincident arrays can be seen in Figure 102, Figure 103, Figure 104, Figure 105, and Figure 115. It can be observed that the shape of a cross emerges from the drawings

of the trumpet and tenor saxophone, while the more complex radiation patterns of the cello and viola [174, 175] create more irregular patterns. The emergent shape of the cross does echo the physical configuration of the array itself.

5.9.2.3 Non-Coincident Array

The non-coincident array representations are displayed in Figure 106, Figure 107, Figure 108, Figure 109, and Figure 116. These representations are the most irregular. This does reflect the more separated nature of the microphone placement. Again, the radiation patterns of the specific instruments appear to impact their perceived audio image. The trumpet and saxophone images are more confined and smaller than the more complex radiating string instruments.

5.9.2.4 MS-XYZ Array

The MS-XYZ arrays are displayed in Figure 110, Figure 111, Figure 112, Figure 113, and Figure 117. While the other arrays have all microphones aimed at the source, this array has only the M microphone pointed at the source. Here the figure-8 microphones are collecting horizontal and lateral information. This can be seen in the circular aspect exhibited by all representations.

In examining the MS-XYZ against the Non-Coincident representations, a case could be made that the MS-XYZ representations are of similar shape, but larger than the Non-Coincident drawings.

5.9.3 Statistical Analysis

5.9.3.1 Horizontal-Vertical Grid Mapping

The loudspeakers used for playback provided a physical reference for the graphical representations drawn by each subject. The maximum extent of each representation was mapped

against the physical layout of the playback loudspeaker positions: -30º left, +30º right, +45º high, and -30º low.

5.9.3.2 Depth Grid Analysis

The depth grid provided a rating from 1-10. Each subject's maximum was used in the analysis below.

5.9.3.3 Horizontal Left

There was a significant difference in the left boundary between microphone techniques ($F(3,144) = 8.01$, $p < .001$) (Figure 93). Post-hoc pairwise t-tests revealed that all microphone techniques were significantly different in their left boundary. The mono technique had the narrowest spread to the left at 3.95deg (SD = 2.07). The coincident spread significantly more to the left at 11.42deg (SD = 8.57). The non-coincident technique spread significantly more to the left at 14.73deg (SD = 9.06). The MS-XYZ spread the furthest to the left at 19.73 (SD = 7.69).

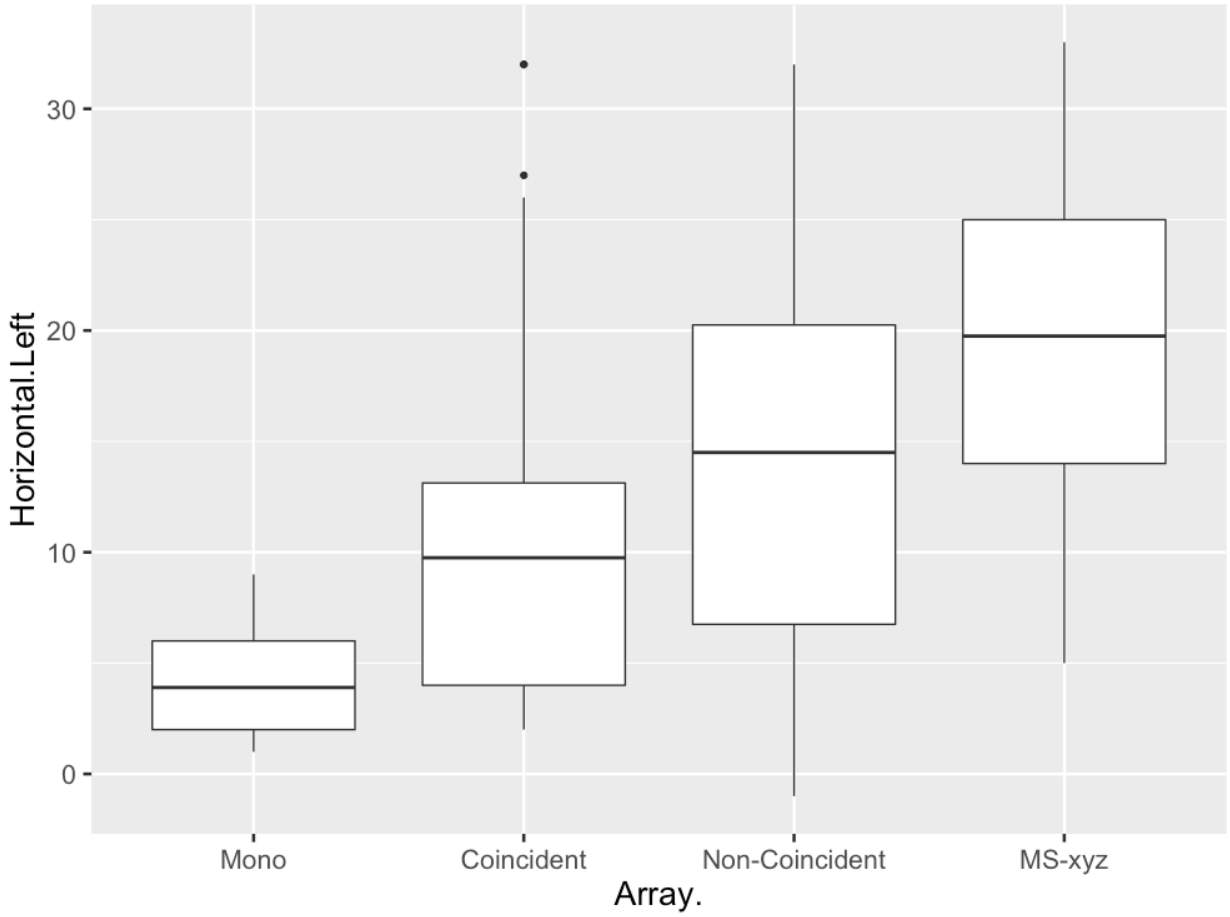


Figure 93: Horizontal Left.

5.9.3.4 Horizontal Right

There was a significant difference in the right boundary between microphone techniques (F (3,144) = 4.83, p <.01) (Figure 94). All microphone techniques were significantly different in their right boundary except the non-coincident versus the MS- XYZ. The mono technique had the narrowest spread to the right at 3.97deg (SD = 2.35). The coincident spread significantly more to the right at 9.07deg (SD = 8.03). The non-coincident technique spread significantly more to the right at 15.13deg (SD = 9.90). The MS- XYZ spread was not statistically different than the non-coincident technique at 15.56deg (SD = 9.26).

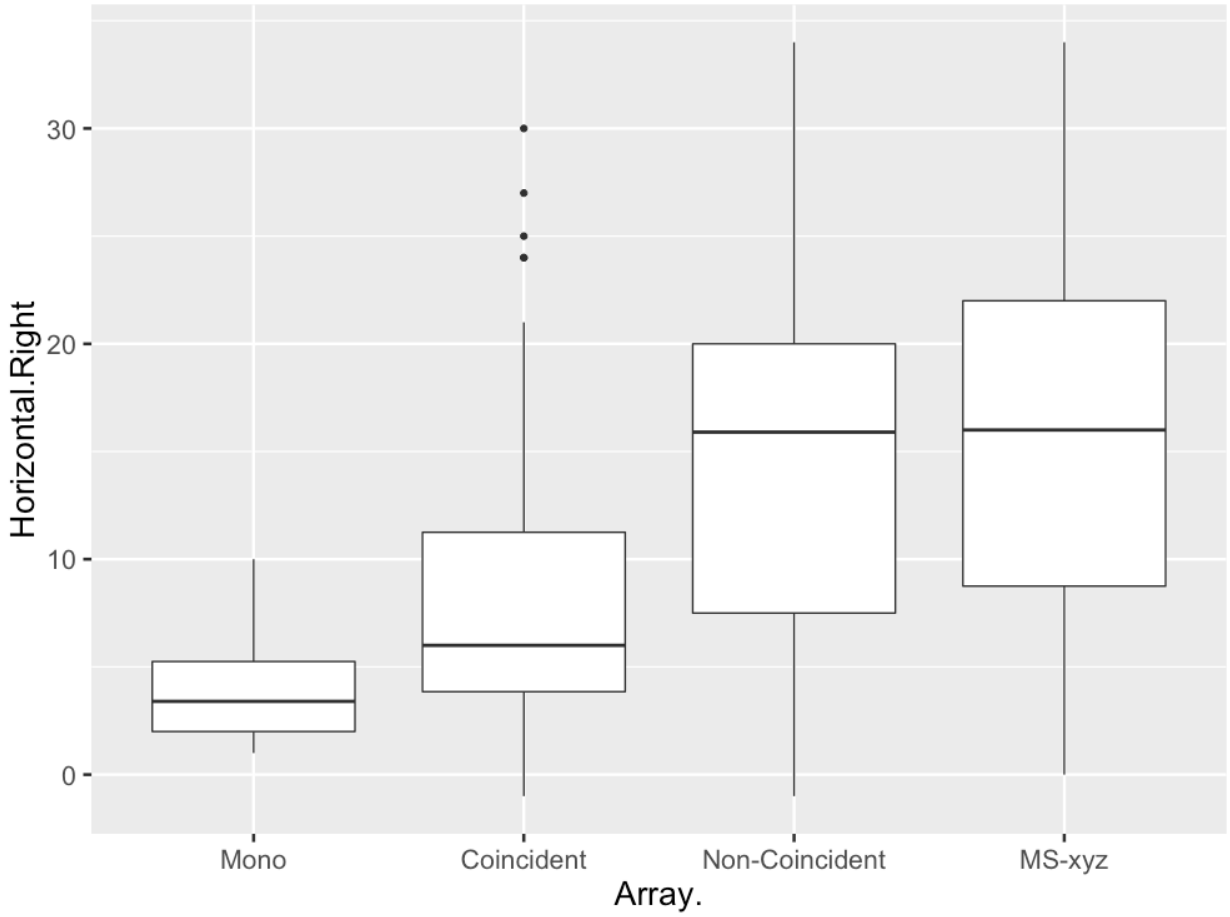


Figure 94: Horizontal Right.

5.9.3.5 Vertical High

There was a significant difference in the upper boundary between microphone techniques ($F(3,144) = 8.07$, $p < .001$) (Figure 95). Differences only existed between the mono technique and the other 3D techniques. The mono technique had the narrowest spread to the top at 7.74deg (SD = 5.82). The coincident (M = 16.85deg, SD = 12.08), non-coincident (M = 17.30deg, SD = 10.63), and MS-XYZ (M = 21.95deg, SD = 12.22) techniques all spread significantly further to the top than the mono technique, but were not different from each other.

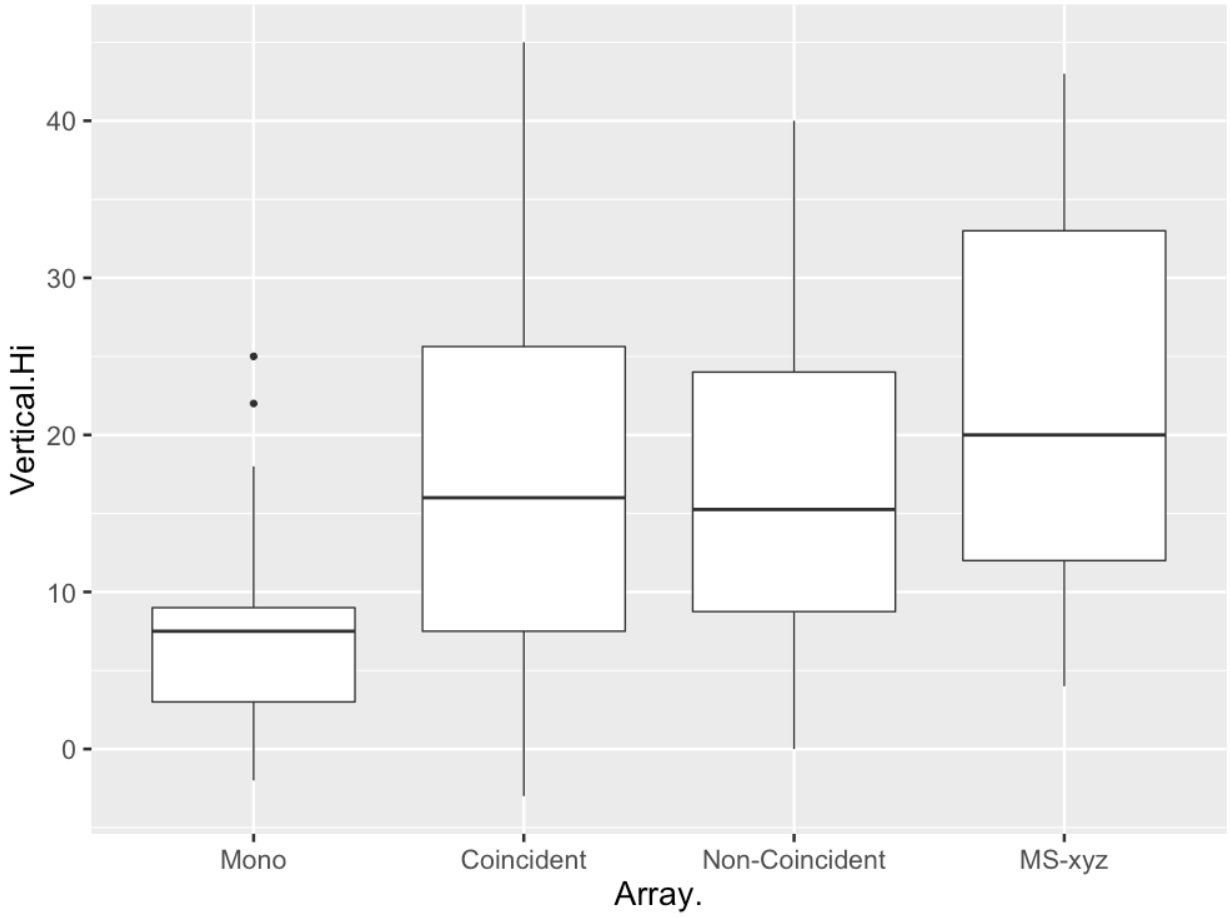


Figure 95: Vertical high.

5.9.3.6 Vertical Low

There was a significant difference in the lower boundary between microphone techniques (F (3,144) = 3.48, p <.05) (Figure 96). Differences only existed between the mono technique and the other 3D techniques. The mono technique had the narrowest spread to the bottom at 5.83deg (SD = 7.81). The coincident (M = 13.92deg, SD = 11.63), non-coincident (M = 12.00deg, SD = 10.06), and MS-XYZ (M = 12.20deg, SD = 11.26) techniques all spread significantly further to the bottom than the mono technique, but were not different from each other.

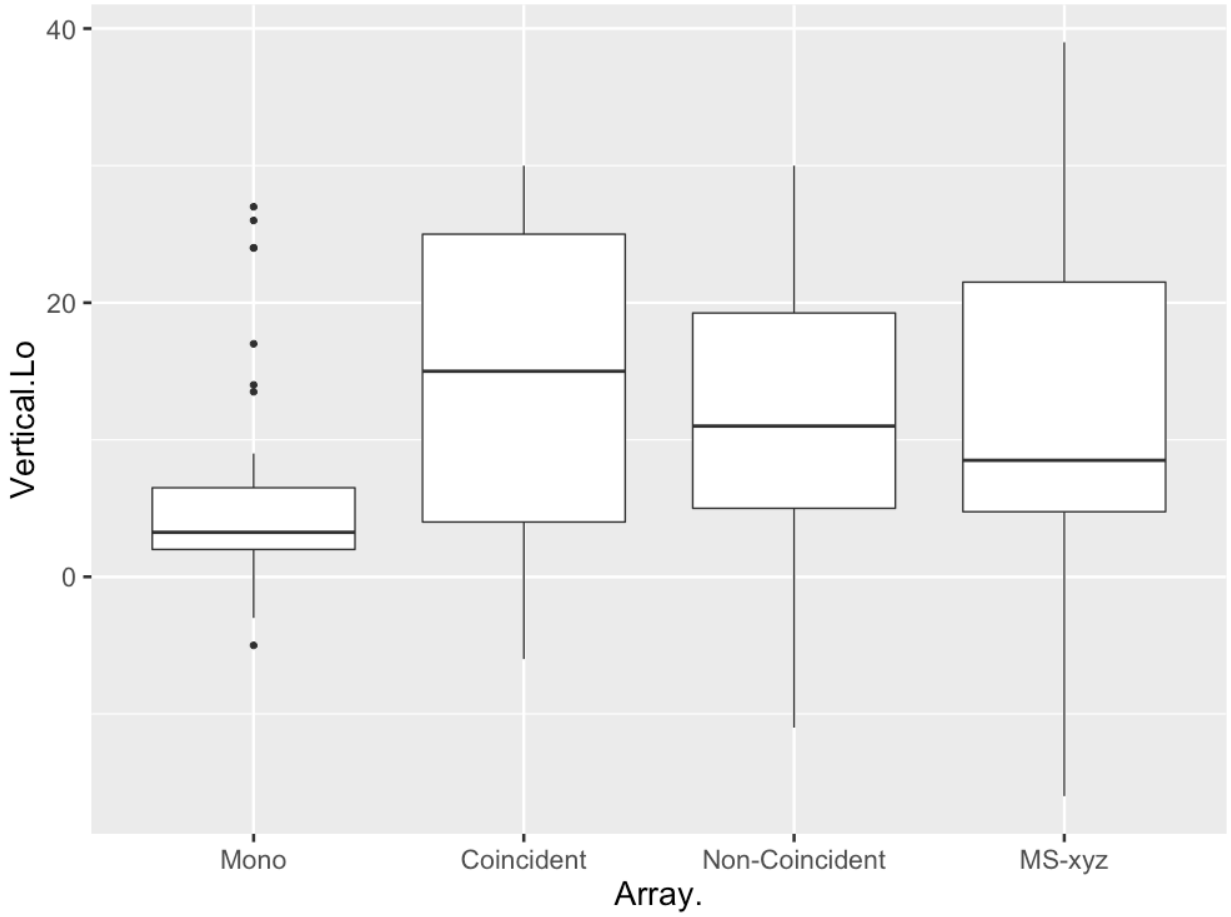


Figure 96: Vertical Low.

5.9.3.7 Depth

There was a significant difference in depth between microphone techniques ($F(3,144) = 2.91$, $p < .05$) (Figure 97). Differences existed between all techniques except for mono vs coincident and coincident vs non-coincident. The mono technique had the lowest depth rating at 1.84 (SD = 1.64), on a scale from 0 to 10. The coincident technique had increased depth at a rating of 2.60 (SD = 1.38). The non-coincident technique had more depth at a rating of 2.78 (SD = 1.33). The MS-XYZ technique had the highest depth rating at 3.67 (SD = 1.91).

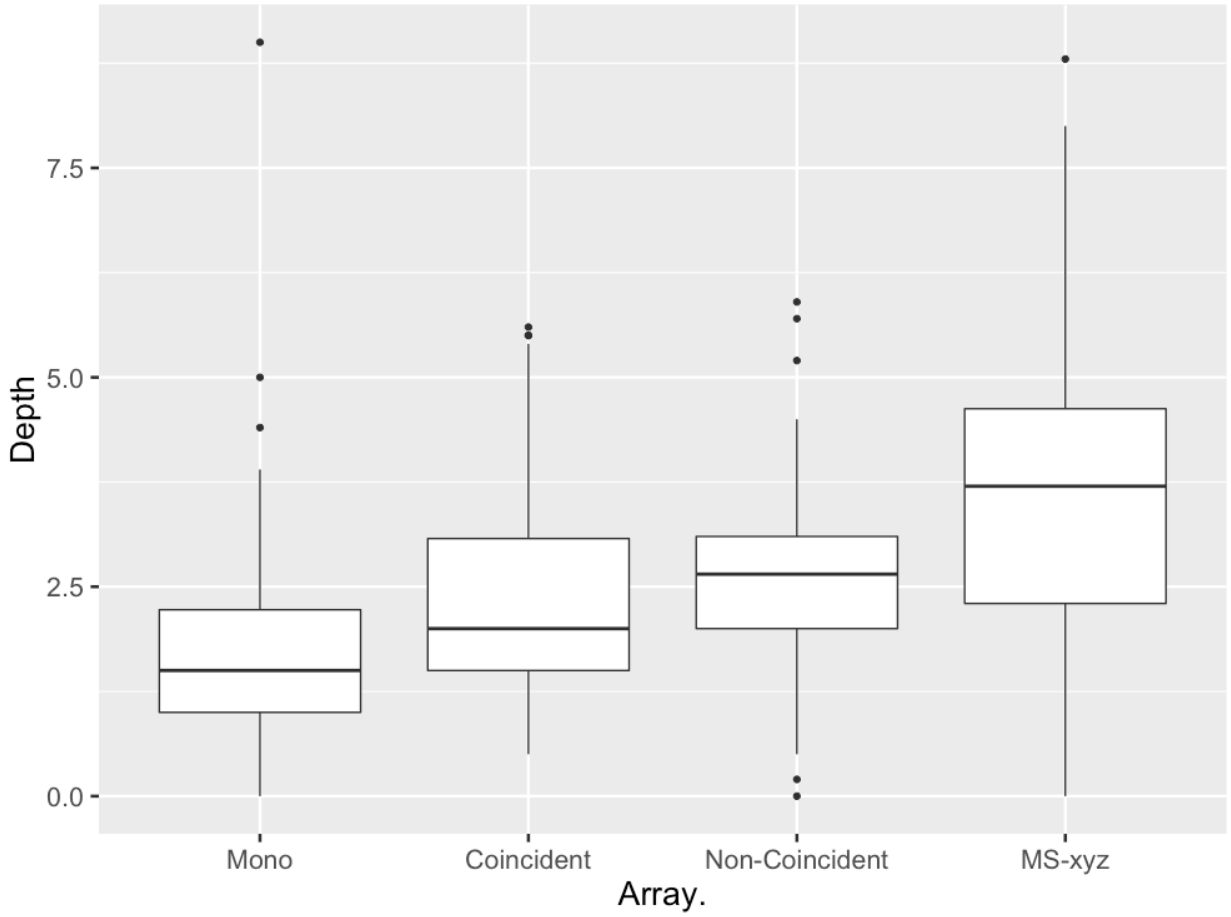


Figure 97: Depth.

5.10 Conclusion

It can be seen in the subjects' representations and the statistical analysis that these techniques clearly capture much more information than a single microphone. The arrays exhibit a greater extent in every dimension—horizontal, vertical, and depth—compared to the monophonic image. In examining the audio images provided by these arrays in the context of immersive/3D mixing and post-production, a case can be made that they will contribute to a more efficient and improved workflow. Five channels are the maximum required for these arrays. The task of up-mixing and

spatializing a mono source can easily include nine extra channels of direct sound and multiple delays and reverb plug-ins [169, 170].

By consulting the array configurations prior to recording it should be possible to determine the correct array to achieve the desired image size. Setup time for these arrays is on par with standard stereo pairs, so there is little impedance to their use in typical recording sessions. Another aspect of the arrays is that they are easily optimized during mixing or post-production. Depending on the chosen array, all axes (horizontal, vertical, and depth) can be made available for manipulation.

With a growing demand to provide 3D and immersive content, it is the hope of the authors that we have provided a small first step into the possibilities and advantages of these techniques.

5.11 Future Work

The understanding of methods and techniques involved in capturing audio in three dimensions is in its infancy. There is much work to be done. The authors' ongoing work in this area include:

1. Qualitative perceptual testing of the above arrays using spatial descriptors to better understand the information offered by each array.
2. Examination of individual subject's representations of each array and mono/point source.
3. Examination of each instrument's specific radiation patterns in relation to its perceived audio image played back through loudspeakers that include height channels.
4. The development of mixing tools for immersive/3D audio.
5. Determining the usable expansion range (size) of the image provided by each array.

5.12 Drawings of Audio Images

5.12.1 Mono Drawings

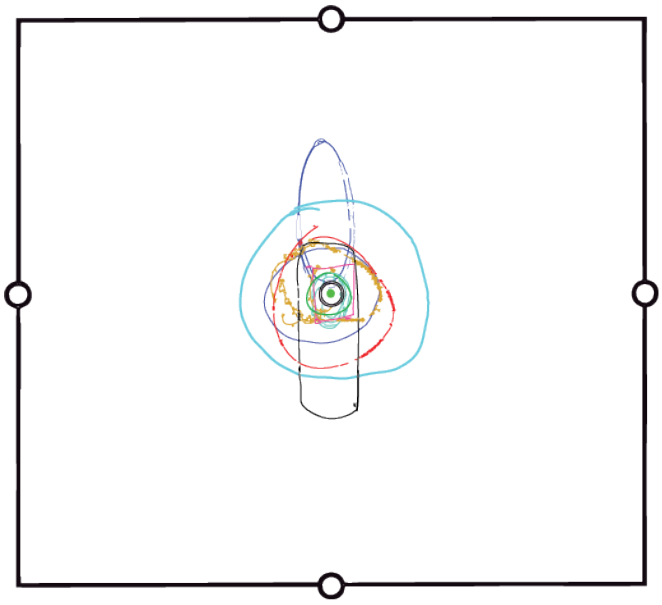


Figure 98: Cello Mono.

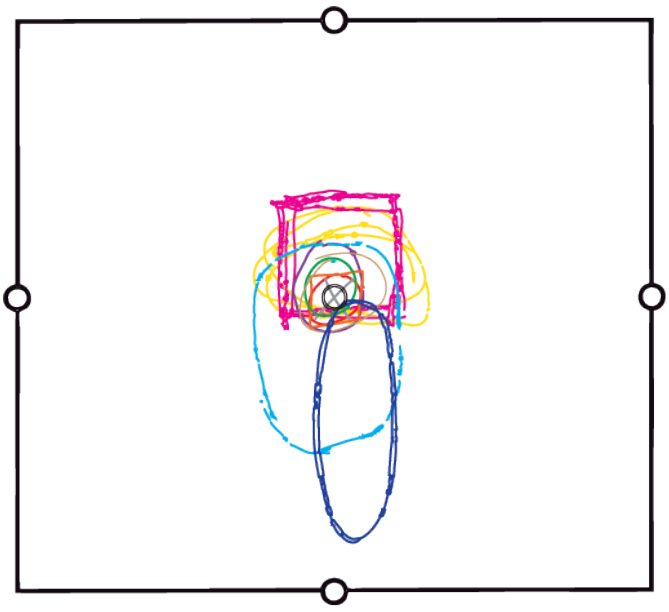


Figure 99: Tenor Saxophone Mono.

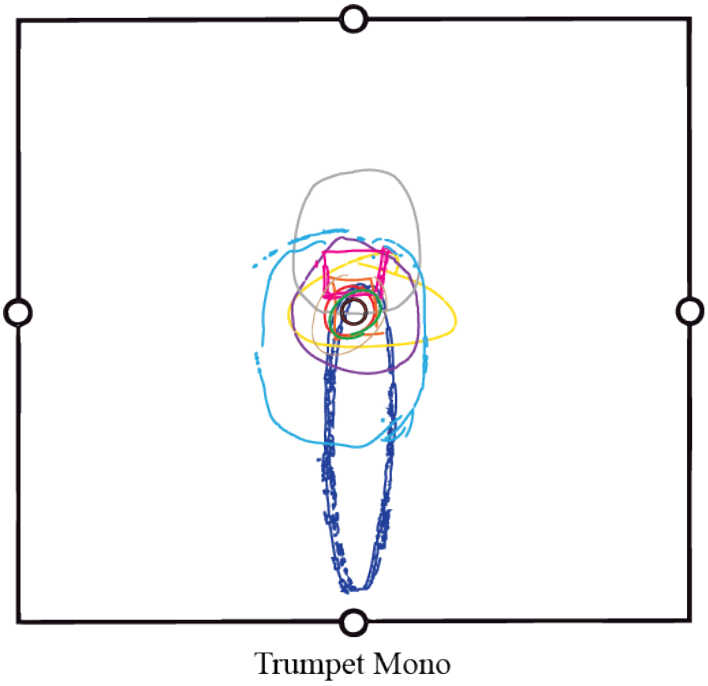


Figure 100: Trumpet Mono.

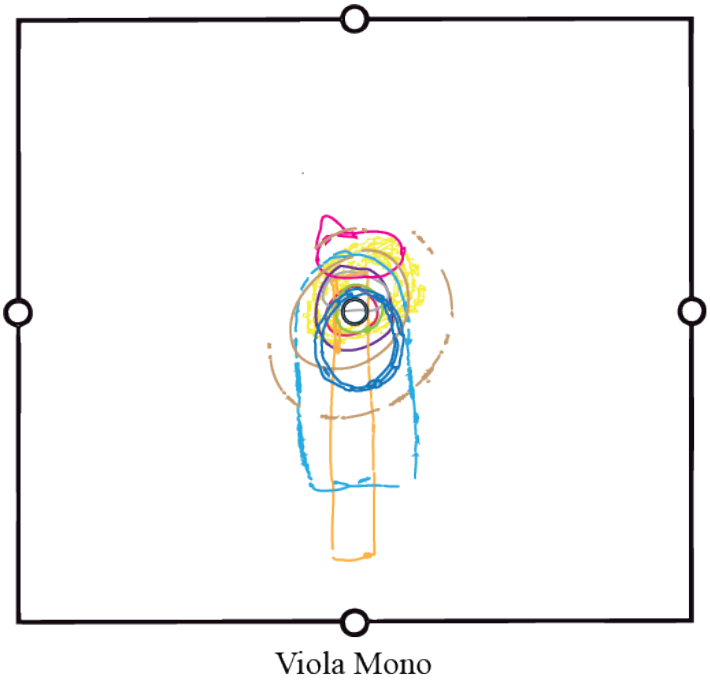


Figure 101: Viola Mono.

5.12.2 Coincident Array Drawings

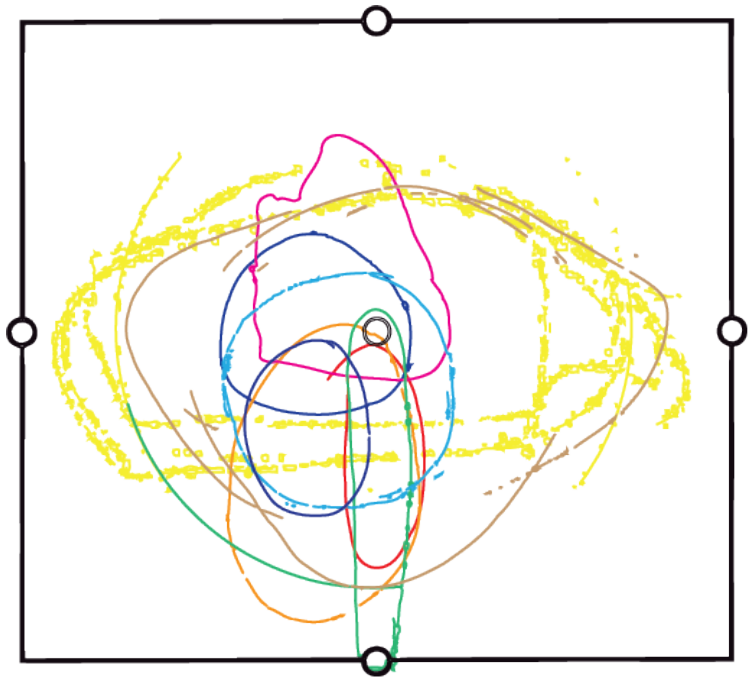


Figure 102: Cello coincident array.

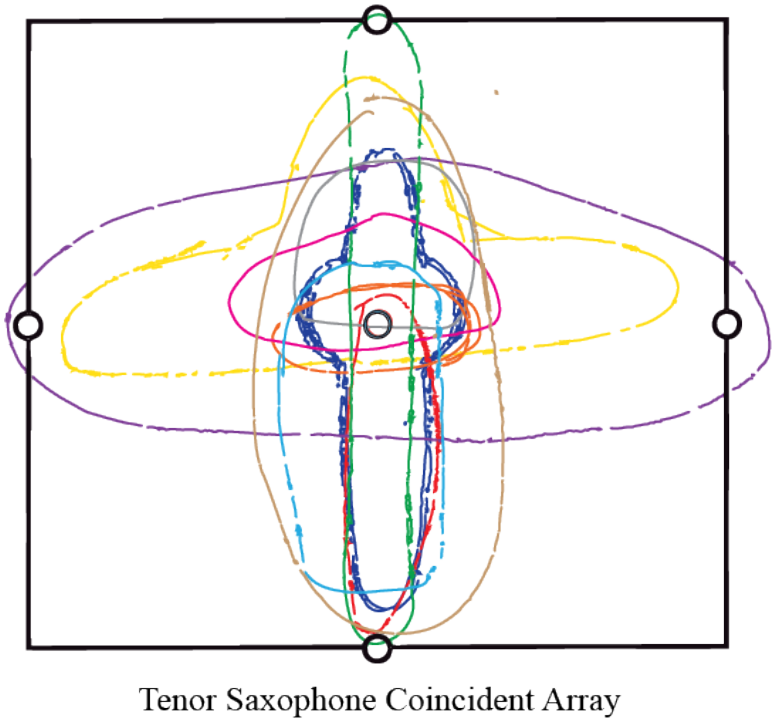


Figure 103: Tenor Saxophone coincident array.

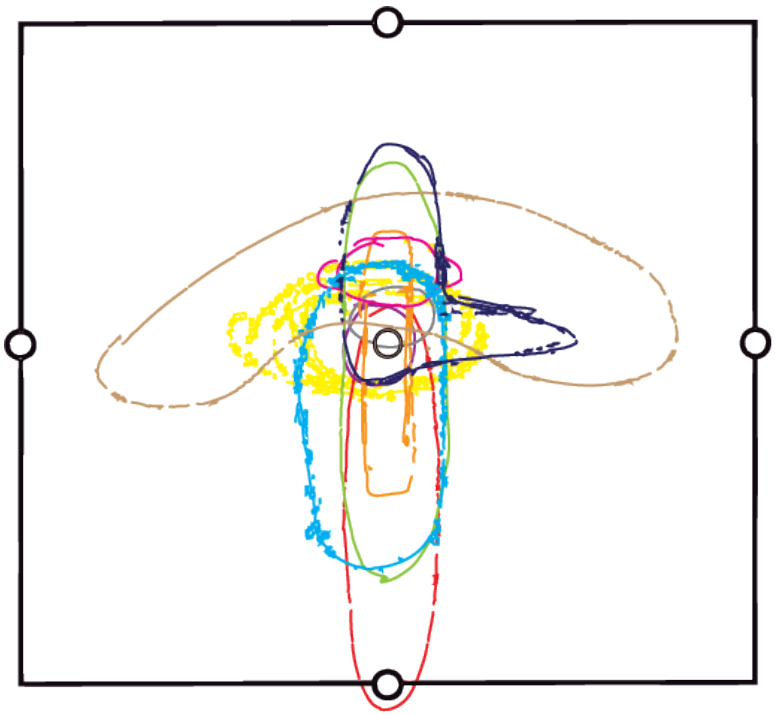


Figure 104: Trumpet coincident array.

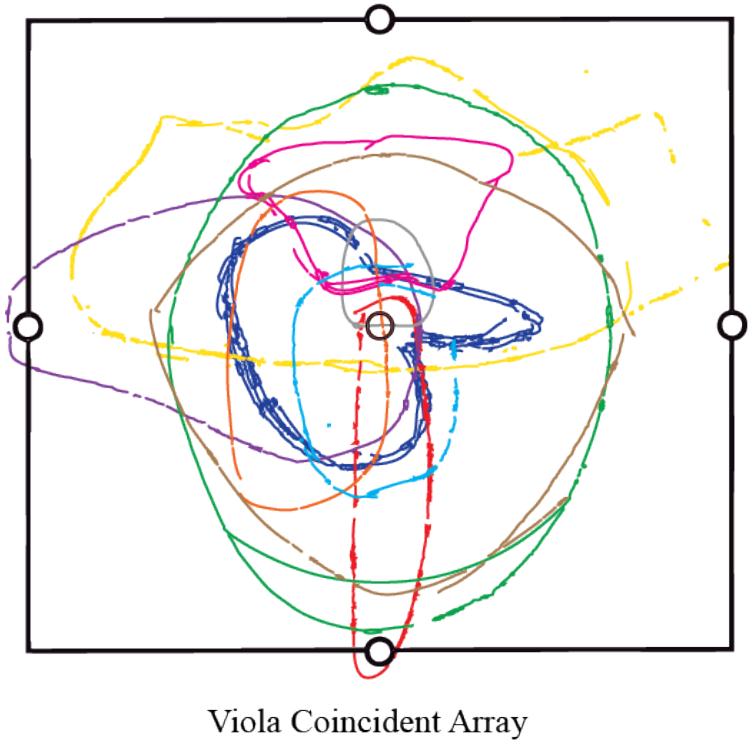


Figure 105: Viola coincident array.

5.12.3 Non-Coincident Array Drawings

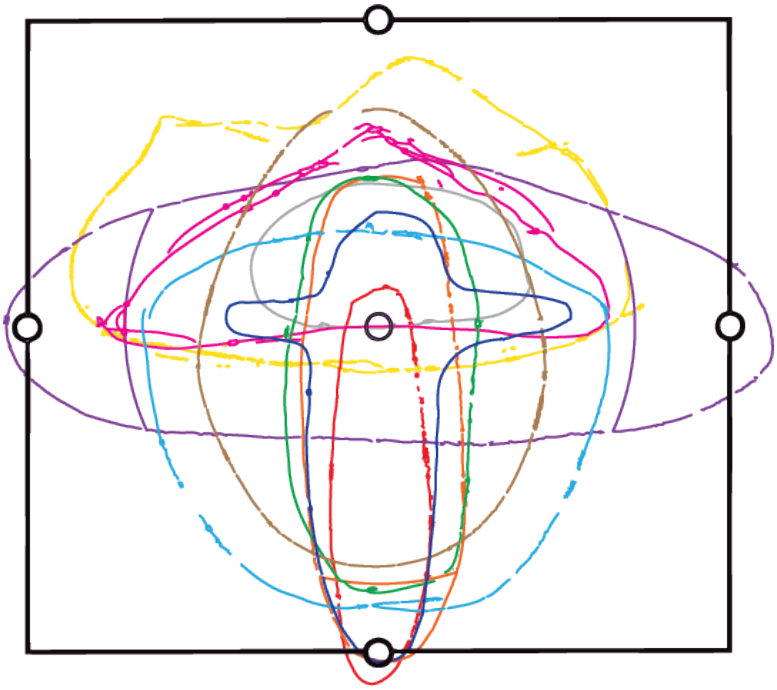


Figure 106: Cello non-coincident array.

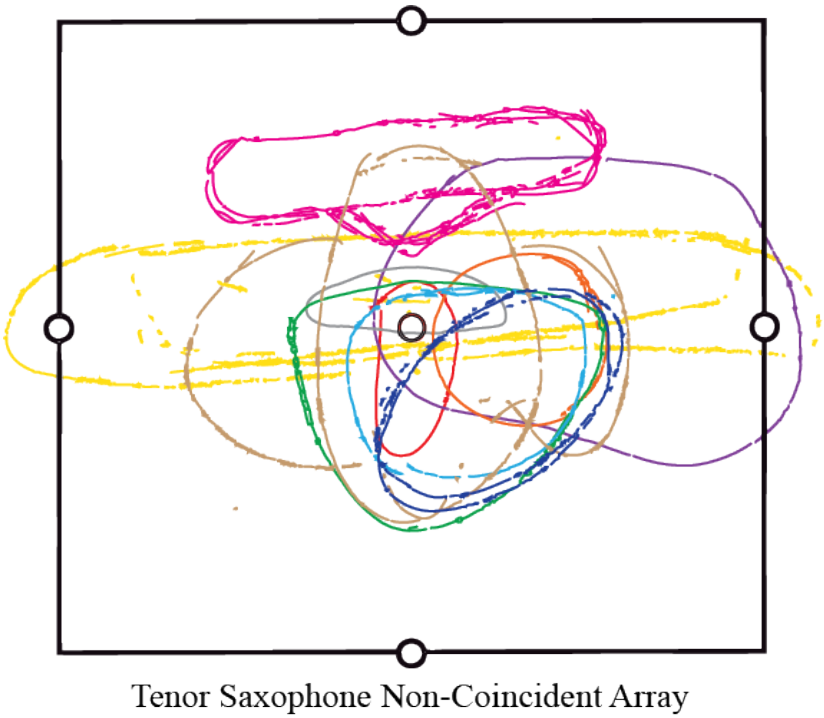


Figure 107: Tenor Saxophone non-coincident array.

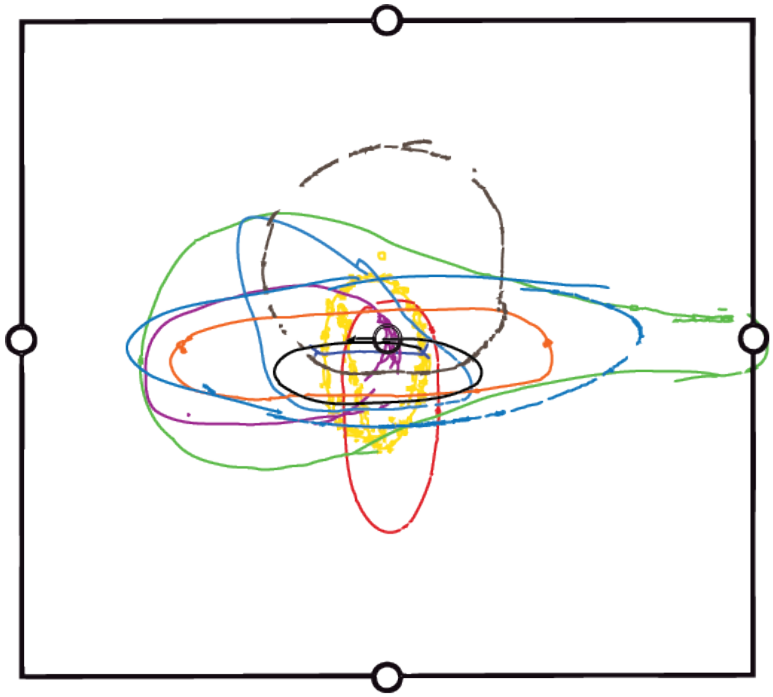


Figure 108: Trumpet non-coincident array.

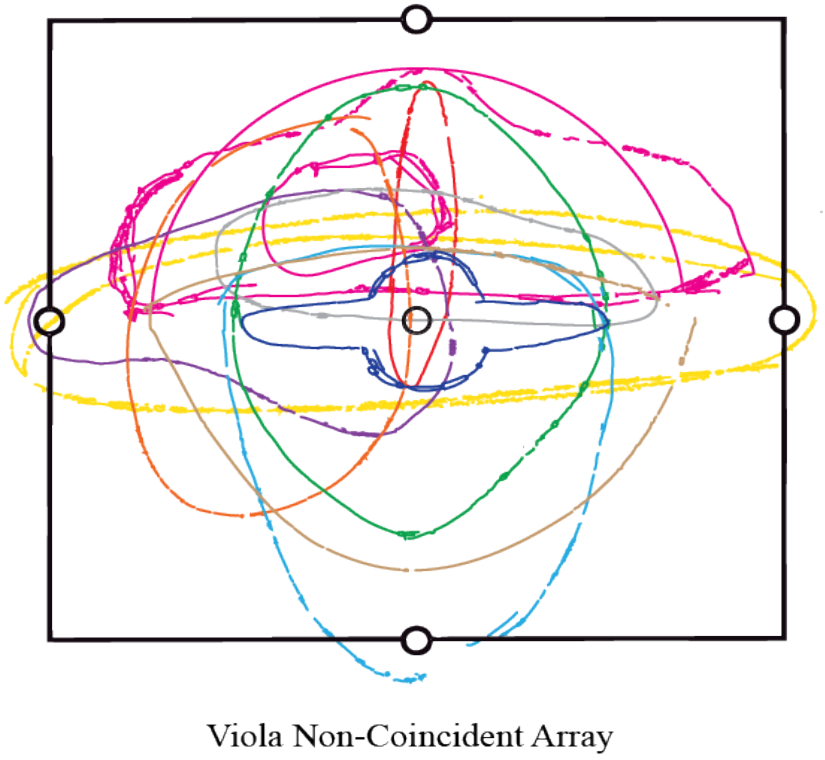


Figure 109: Viola non-coincident array.

5.12.4MS-XYZ Drawings

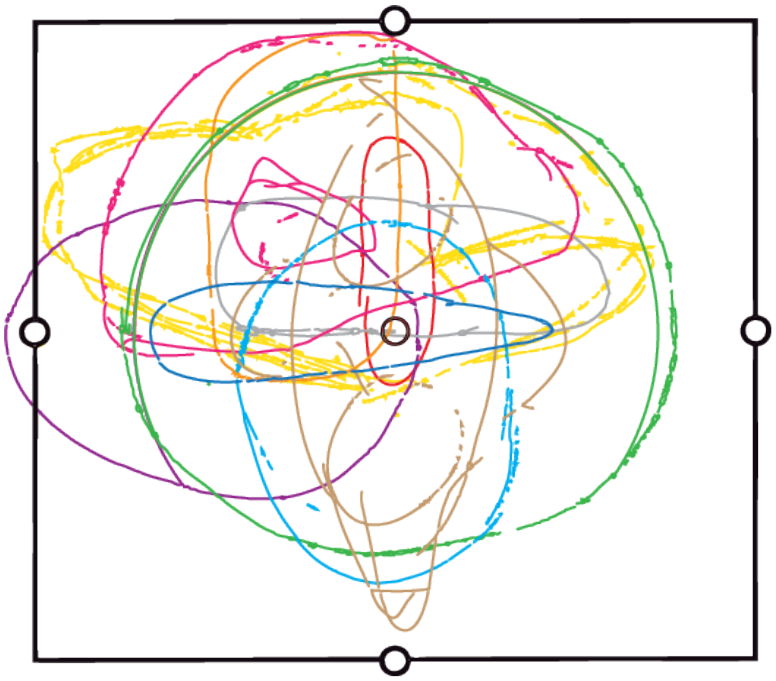


Figure 110: Cello MS-XYZ.

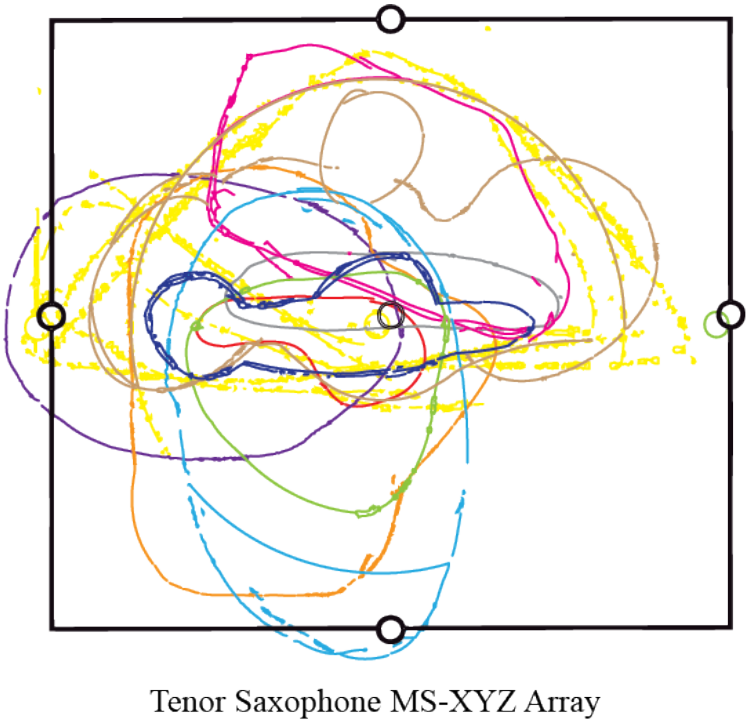


Figure 111: Tenor Saxophone MS-XYZ.

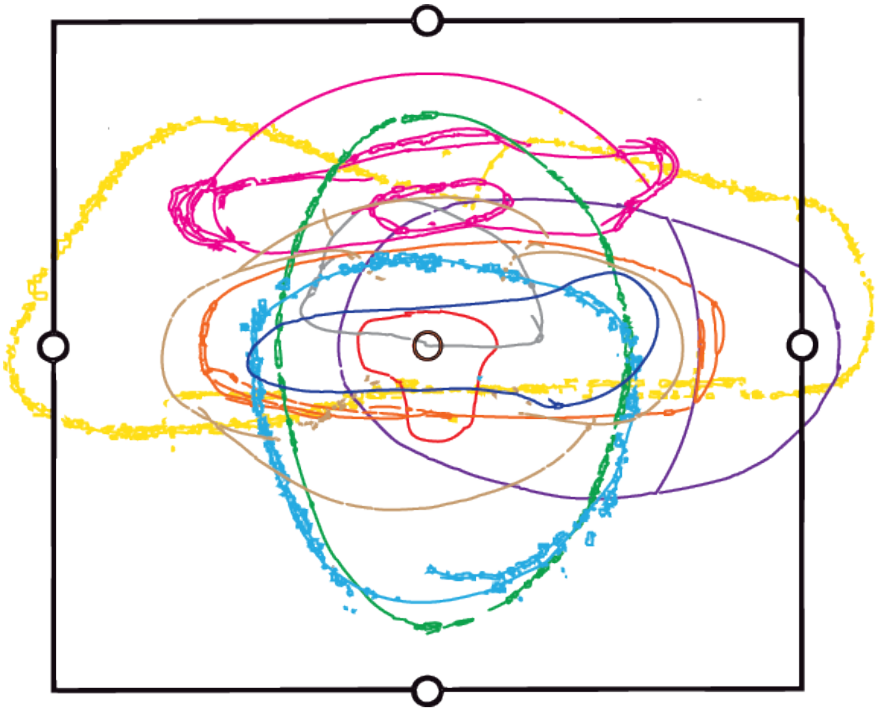


Figure 112: Trumpet MS-XYZ.

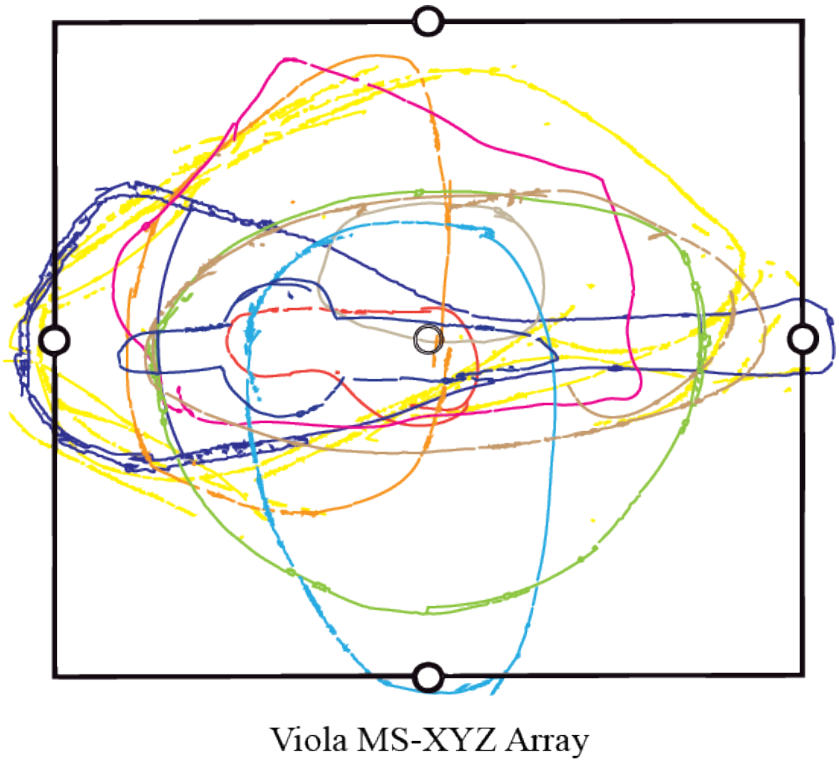


Figure 113: Viola MS-XYZ.

5.12.5 Accumulated Array Drawings

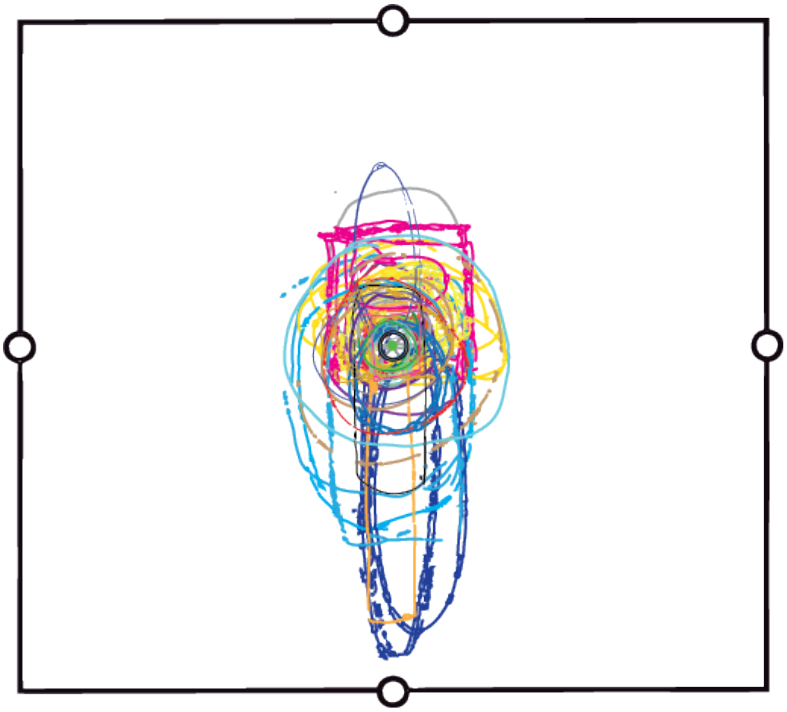


Figure 114: All mono drawings.

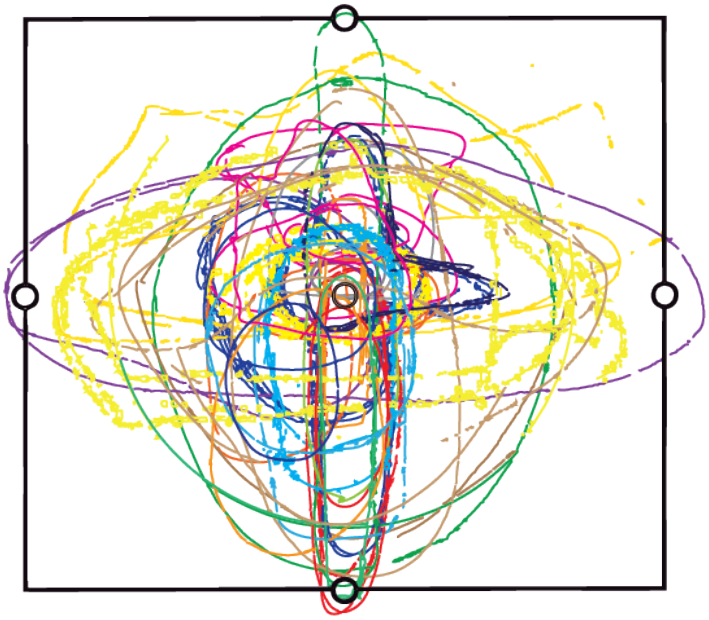


Figure 115: All coincident array drawings.

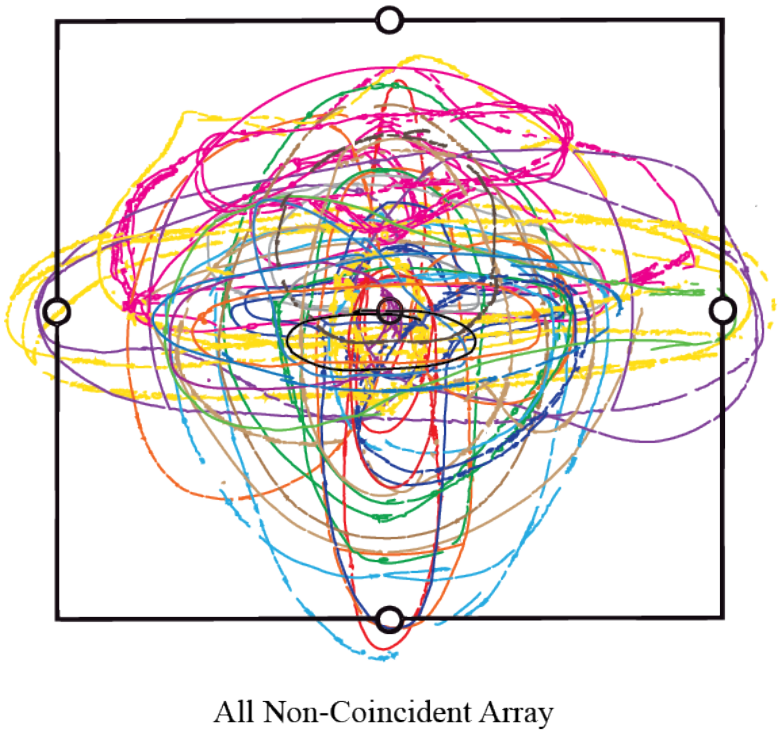


Figure 116: All non-coincident array drawings.

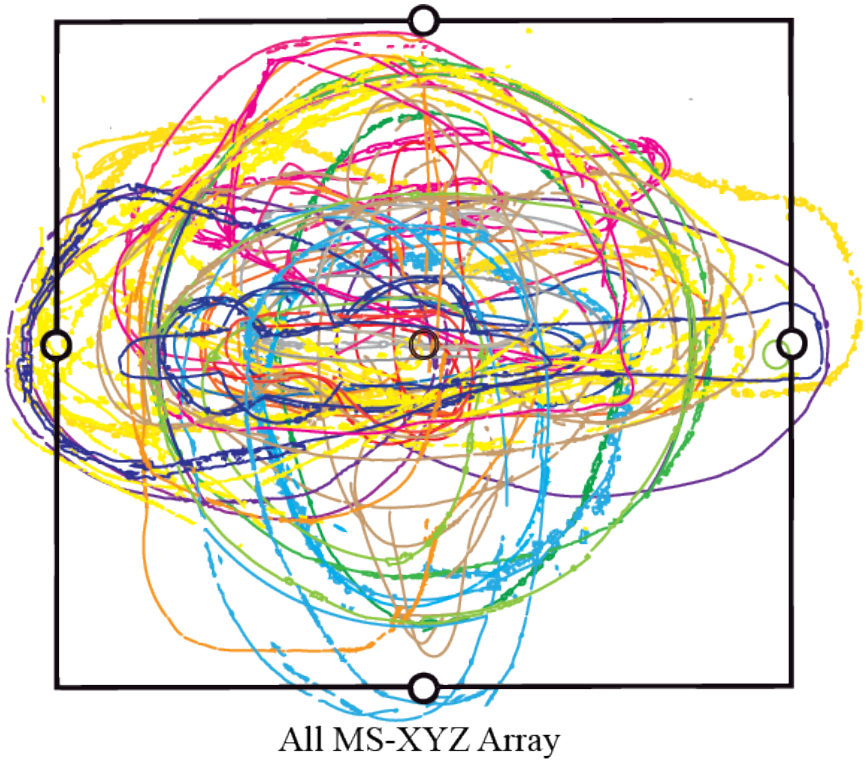


Figure 117: All MS-XYZ drawings.

6 SUBJECTIVE ASSESSMENT OF THE VERSATILITY OF THREE-DIMENSIONAL NEAR-FIELD MICROPHONE ARRAYS FOR VERTICAL AND THREE-DIMENSIONAL IMAGING

6.1 Abstract

This investigation examines the operational size-range of audio images recorded with advanced close-capture microphone arrays for three-dimensional imaging. It employs a 3D panning tool to manipulate audio images. The 3D microphone arrays used in this study were: Coincident-XYZ, M/S-XYZ and Non-coincident-XYZ/five-point. Instruments of the orchestral string, woodwind, and brass sections were recorded. The objective of the test was to determine the point of three-dimensional expansion onset, preferred imaging, and image breakdown point. Subjects were presented with a continuous dial to manipulate the three-dimensional spread of the arrays, allowing them to expand or contract the microphone signals from 0˚ to 90˚ azimuth/elevation. The results showed that the M/S-XYZ array is the perceptually "biggest" of the capture systems under test and displayed the fasted sense of *expansion onset*. The coincident and

non-coincident arrays are much less agreed upon by subjects in terms of *preference* in particular, and also in *expansion onset*.

6.2 Introduction

6.2.1 Immersive Audio Recording and Production

This investigation continues research into multi-channel recording practices which capture three-dimensional (3D) audio images [178, 179]. The goal has been to develop recording techniques—that when played back over loudspeaker configurations that include middle, height and ideally bottom channels—will describe the captured sound source in a coherent three-dimensional representation. During the mixing process, these images will allow for the manipulation of the 3D image in three axes (horizontal, vertical and depth). These arrays are expanded from the standard stereo practices of coincident, spaced pairs, and the M/S technique.

Previous research [26, 28, 180] has reported that music produced employing techniques optimized for three-dimensional audio playback results in program material that delivers an increased impression of presence, reality, depth, envelopment, and naturalness. Most published investigations addressing three-dimensional microphone arrays applied in music recording have focused on the capture of ensembles in live situations. These techniques prioritize the direct sound in the middle loudspeaker layer, ambience in the height layer and employ no bottom layer loudspeakers [112-114, 116, 120, 121]. Notable exceptions are Howie et al. [18] and Hamasaki et al. [172] who have reported on recording techniques for large classical music ensembles optimized for the 22.2 Multi-Channel Sound System [1]. Howie [181] has reported on the discussed close-capture 3D arrays in the production of pop and rock for 22.2 reproduction.

6.2.2 Stereo vs 3D Reproduction

Typical stereo playback is generally restricted to a 60º width of reproduction (-30º left, +30º right), and by comparison the 3D environment is spatially vast, if we exclude the acoustics of the listening room. Martin and King [169] have reported that sound images that provide appropriately sized sonic images for conventional stereo are judged small and unimpactful in an immersive hemispherical acoustic such as the 22.2 Multi Channel Sound System.

Conventional close capture techniques used in studio recording employ one or two microphones per instrument, which is quite adequate for mono or stereo reproduction. But these techniques are not optimal in the reproduction of 3D audio images in the 360º hemispherical acoustic. Mono/stereophonic images provide point-source or one-dimensional information in a sound field that allows for sound to be delivered from all angles.

In the current 3D workflow mixing engineers must expand mono- or stereophonic images to those of appropriate size to be impactful in the 3D environment. This is attempted by the use of track duplication, delay, 3D panning, convolution and algorithmic reverberation, and lately by the use of 3D DSP tools such as SPAT Revolution, DearVR, and Facebook360. The resulting images are generally lacking in stability, cohesion and definition. The microphone arrays investigated in this research seek to solve these issues at the recording stage and provide greater flexibility during the mixing process.

6.2.3 A More Complete Sonic Picture

A complete sonic ‘picture’ is comprised of a sound source and the accompanying acoustics. The fact that the majority of musical instruments do not radiate sound in all directions with equal

intensity [174] coupled with the complexity of the acoustics of the recording space [177] presents a sound image that is both spectrally complex and dynamic in time and timbre.

This reality limits the amount of information that can be obtained by close capture techniques that employ one or two microphones. There is a trade-off between proximity and perspective: the closer the microphone is to the instrument, the less spatial information becomes available. Previous research by Martin et al. [178, 179] on the discussed arrays has shown that microphone arrays designed for 3D recording deliver a more complete sound image and also provide horizontal *and* vertical extent as well as an increase in the perceived depth of the image.

6.3 Test Design and Methods

6.3.1 Testing Environment

See section 3.1.

6.3.2 Microphone Arrays

See section 4.3.1.

6.3.3 Array Level Matching

See section 4.4.

6.3.4 Test Subjects

Twenty-six test subjects participated in the experiment. Their ages ranged from 21 to 60 years, with a mean age of 30.6 years. All subjects were members of the sound recording department

at the Schulich School of Music, McGill University. Undergraduate, Masters, and Ph.D. students, as well as faculty members were represented in the subject pool. All subjects reported having normal hearing.

6.3.5 Test Material

See section 4.6.

6.3.6 Test Procedure

6.3.6.1 Test Design

The listening test was administered using a purpose-built program developed in the Max7 coding environment. The variables 'Mic Array Type' and 'Instrument' were paired for each trial using randomization. The test comprised 36 trials per subject, meaning that each subject was presented with each possible 'Mic Array Type/Instrument pair' three times.

6.3.6.2 Test Objective

The objective of the test was to establish boundaries of acceptability for the horizontal and vertical spread of the various arrays. Subjects were presented with a continuous dial (Griffin Powermate USB multimedia controller)—with no markings or visual or tactile feedback —to manipulate the width of the arrays, allowing them to expand or contract the spread of the microphone signals in relation to 0˚ azimuth/elevation.

Each microphone within a specific array was assigned to a panning axis of discrete loudspeakers (Table 12, Figure 118) within the test environment corresponding to the region of capture during recording. The 'M' microphone from the MS-XYZ and non-coincident array was used for the mono 'C' signal.

Panning was achieved using the SPAT toolkit. Each microphone channel was represented as an object in the SPAT environment (see Table 11: *SPAT Object*), who's azimuth and elevation positions were integrally manipulated using the continuous dial input device. The minimum, original, and maximum extents can be seen in Figure 119, Figure 120 and Figure 121. The microphone channels correspond to the green-circled numbers in the panning Figures. Figure 122 shows a 3D representation of the loudspeaker configuration. VBAP was chosen as the panning technique to move the microphone channels through the loudspeaker array. The elevation distance of the top microphone was scaled from 0˚/30˚ to 0˚/45˚ to ensure that the 'original' positions of the microphone arrays would remain intact.

Arrays:	Mic Position:	Microphone:	SPAT Objects:
Coincident	Left	Schoeps CMC62U / MK 4	**1**
	Right	Schoeps CMC62U / MK 4	**2**
	Up	Schoeps CMC62U / MK 4	**5**
	Down	Schoeps CMC62U / MK 4	**4**
M/S-XYZ	Center	DPA 4011	**3**
	Horizontal Side	Schoeps CCM8	**1, 2**
	Vertical Side	Schoeps CCM8	**4,5**
Non-Coincident	Center	DPA 4011	**3**
	Left	Schoeps CMC62U / MK 4	**1**
	Right	Schoeps CMC62U / MK 4	**2**
	High	Schoeps CMC62U / MK 4	**5**
	Low	Schoeps CMC62U / MK 4	**4**

Table 11: Microphone arrays and corresponding panning object numbers. (See Figures 118-121).

Number	Channel Name	Label	Azimuth	Elevation	Distance
1	Side Left	SiL	90°	0°	2.1
2	Front Left	FL	60°	0°	2.1
3	Front Left centre	FLc	30°	0°	2.1
4	Front Centre	FC	0°	0°	2.05
5	Front Right centre	FRc	-30°	0°	2.1
6	Front Right	FR	-60°	0°	2.1
7	Side Right	SiR	-90°	0°	2.1
8	Bottom Front Centre	BtFC	0°	-20°	2.2
9	Top Front Centre	TpFC	0°	35°	2.6
10	Top Centre	TpC	0	90°	2.1

Table 12: Loudspeaker configuration used for testing.

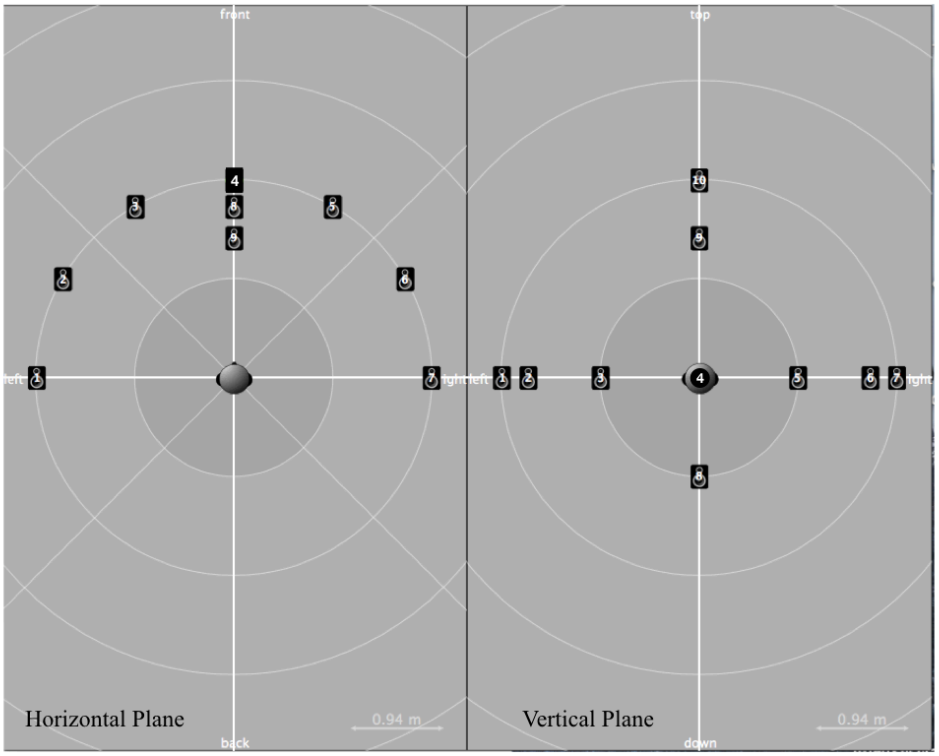


Figure 118: Panning axis of discrete loudspeakers.

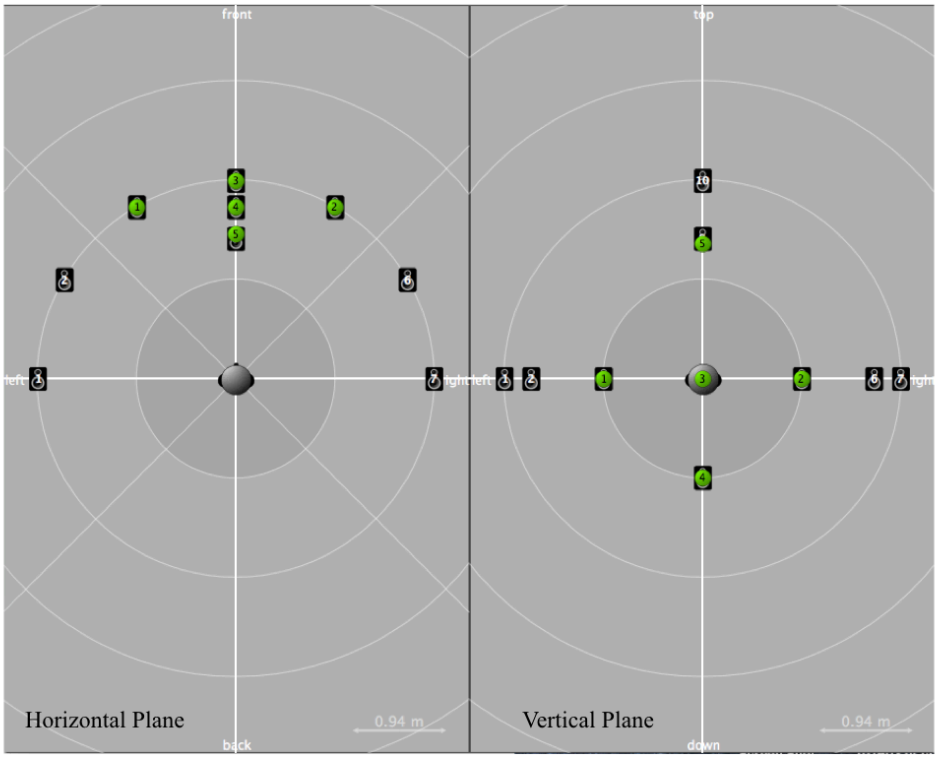


Figure 119: Panning axis of discrete loudspeakers displaying the minimum 0° monophonic image.

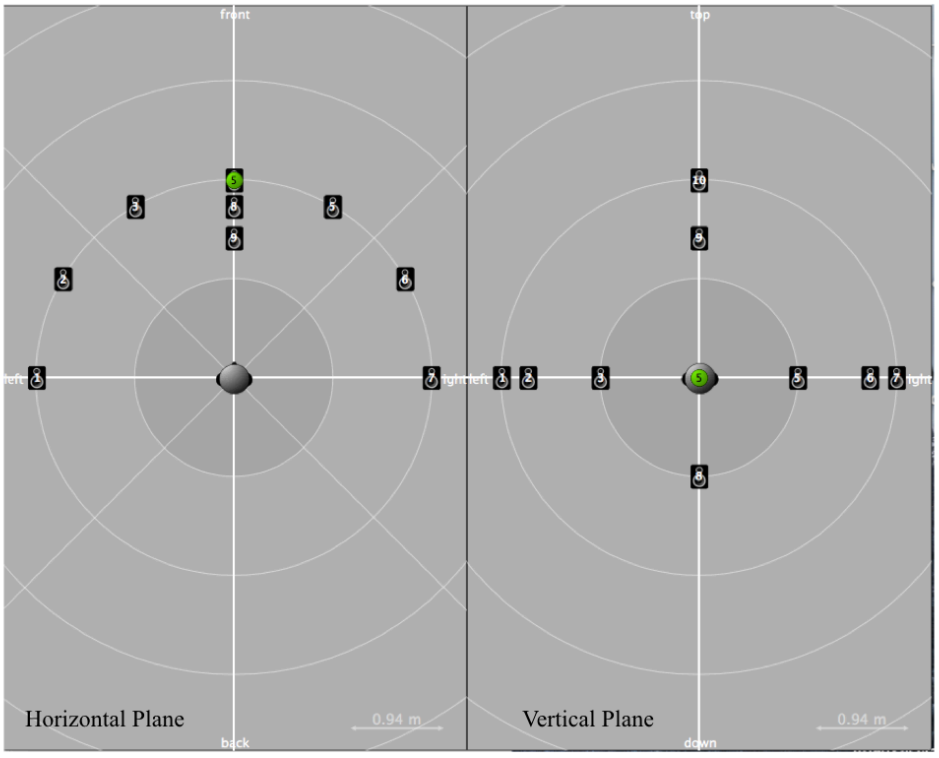


Figure 120: Panning axis of discrete loudspeakers displaying the original 30° image expansion.

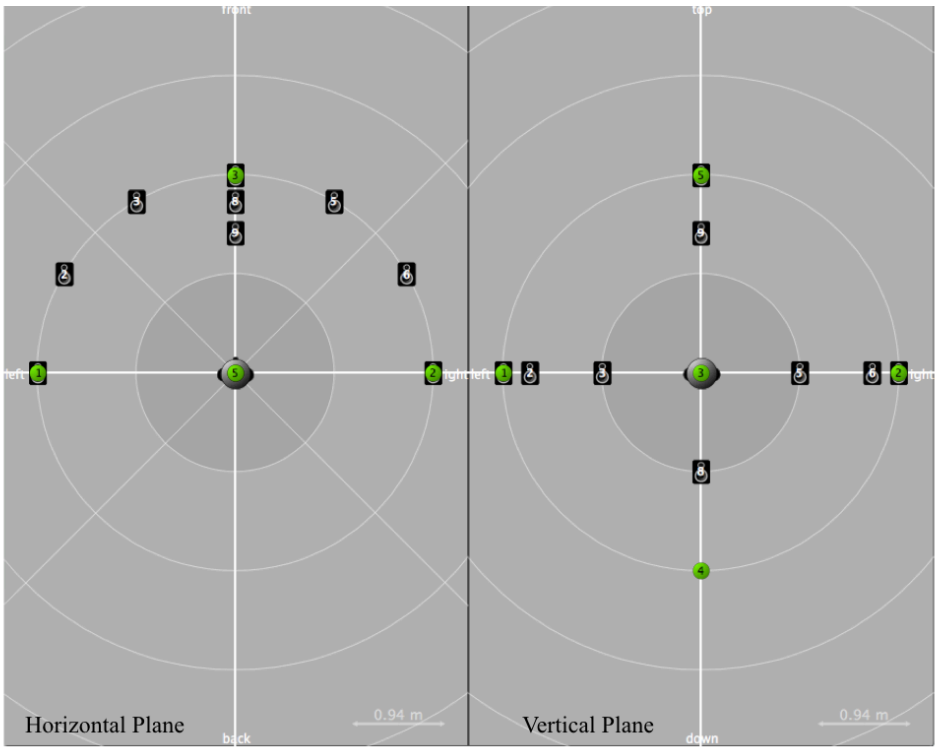


Figure 121: Panning axis of discrete loudspeakers displaying the maximum 90° image expansion.

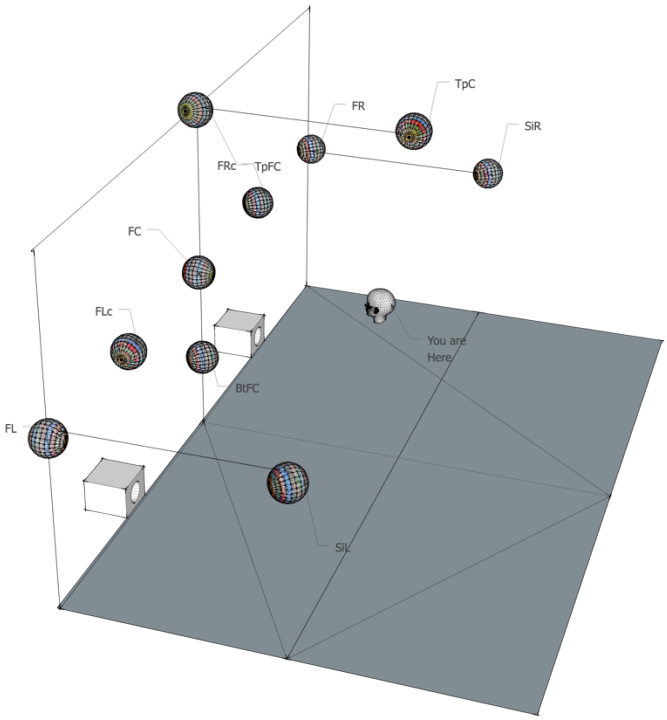


Figure 122: 3D detail of listening configuration.

6.3.7 Training and Familiarization

Prior to the training trials, the subjects were given oral instructions on 1) the use of the control dial, 2) descriptions of the microphone arrays, 3) detail on the instruments used as test stimuli, 4) test procedure, and 5) an explanation of the spatial extent selections (which were the goal of the test and detailed in 6.3.8.1 below).

The training trials consisted of the subjects being provided with one instrument and allowed to toggle through the three microphone arrays under test. This was done to allow the subjects to familiarize themselves with the differences between arrays, and to gain familiarity with the testing interface. Training concluded when the subjects felt they were ready to proceed.

6.3.8 Testing

During the test, the subjects were asked to make three spatial-extent selections on the GUI interface (Figure 123). Selections were made by adjustment of the continuous dial and then registering a choice using the number keys 1, 2 and 3 on a keyboard. Subjects were not required to make these selections in any particular order. Visual feedback was provided to confirm that each selection had been made (Figure 124).

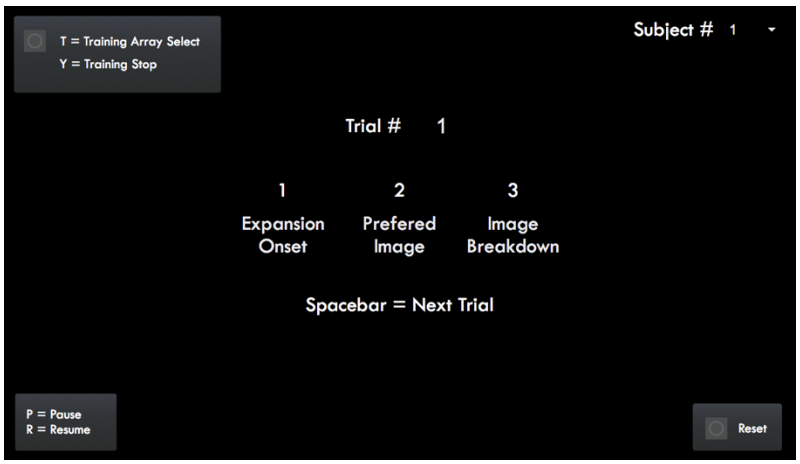


Figure 123: Test GUI.

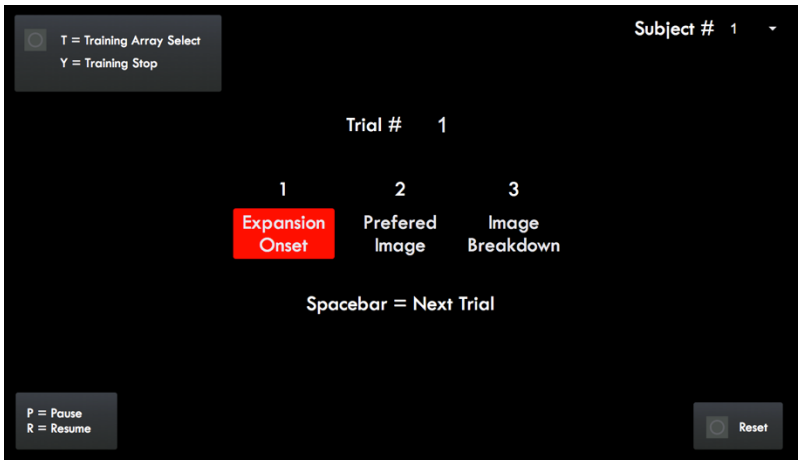


Figure 124: Test GUI with selection.

6.3.8.1 Extent Selections

Expansion Onset: where the audio image begins to expand—horizontally and/or vertically—beyond the monophonic/point source image.

Preferred Image: subjective preference which indicates best 'real-to-life' representation of the instrument.

Image Breakdown: where audio image is no longer a coherent representation of the instrument and is judged to exceed boundary of acceptability or usability.

6.3.9 Post Testing Questionnaire

A short questionnaire was filled by each subject following the testing. The information collected concerned age, music and audio experience, and comments pertaining to the experiment.

6.4 Results

Given the great accumulation of data collected from such a factorial experimental design, the first task was to reduce the two dependent variables set to a single representative dependent variable. Given the inextricable link between elevation and azimuth manipulation in this test's design (i.e. the presented image could not be increased in the horizontal dimension without also increasing in the elevation dimension, and vice versa), a single more generic term of "image size" can be examined using either elevation or azimuth as the enumerated variable. For the purposes of practical application of the data acquired, the generic compound effect of image size will be used primarily throughout the remainder of the analysis, and will range from a minimum of 0, or mono, to a maximum image spread consisting of ±90° in azimuth and +90°/-20° in elevation, the full range available in the reproduction environment. This single dependant variable can then be compared against the myriad of independent variables inherent in such a wide-ranging subjective test.

6.4.1 Nuisance variables

All data were first tested for nuisance/suppressor variables, as well as complex interaction effects that may skew results. As is the case in many ecologically approached studies, the effect of test material was a significant effect. A simple one-way analysis of variance (ANOVA) between the four instruments used in testing revealed their significant influence on image size for *expansion*

onset, *preference*, and *image breakdown*. This effect was most prevalent in *expansion onset*, but was observed across all response questions. While this somewhat complicates analysis by adding statistical noise when viewing the summed results across *instrument*, this suppressor can be accounted for. Unfortunately, the issue was not as simple as a single *instrument* skewing the mean or variance in a given question, but rather slight differences across each *instrument* in each of the three primary responses. A two-way ANOVA revealed that there were not significant interactions between *instrument* and *array* in any of the response questions, rendering further discussion of this interaction moot.

The benefit to using multiple disparate audio samples/sources, in this case four instruments, was borne out in the lack of effect of listener fatigue. There was no change in mean or variance across *trial*. Although *trial* itself was not significant, there was a slight interaction between the *trial* and *array* variables. The manifestation of this interaction seems to be a differing rate of learning (or solidification of opinion) for each array, as manifested by narrowing variances, more slightly and more rapidly for the coincident array (Figure 125). The M/S-XYZ array, while displaying the lowest overall variance, increased in variance with increasing *trials*.

Subjects were also asked to self-diagnose any hearing loss in a post-testing survey. While this would likely be grounds for exclusion of the subject's response data, it was found the subject(s) who self-diagnosed abnormal hearing were as consistent as those reporting normal hearing. Based on a lack of departure from the remaining subject population, these individuals were not excluded from the data set.

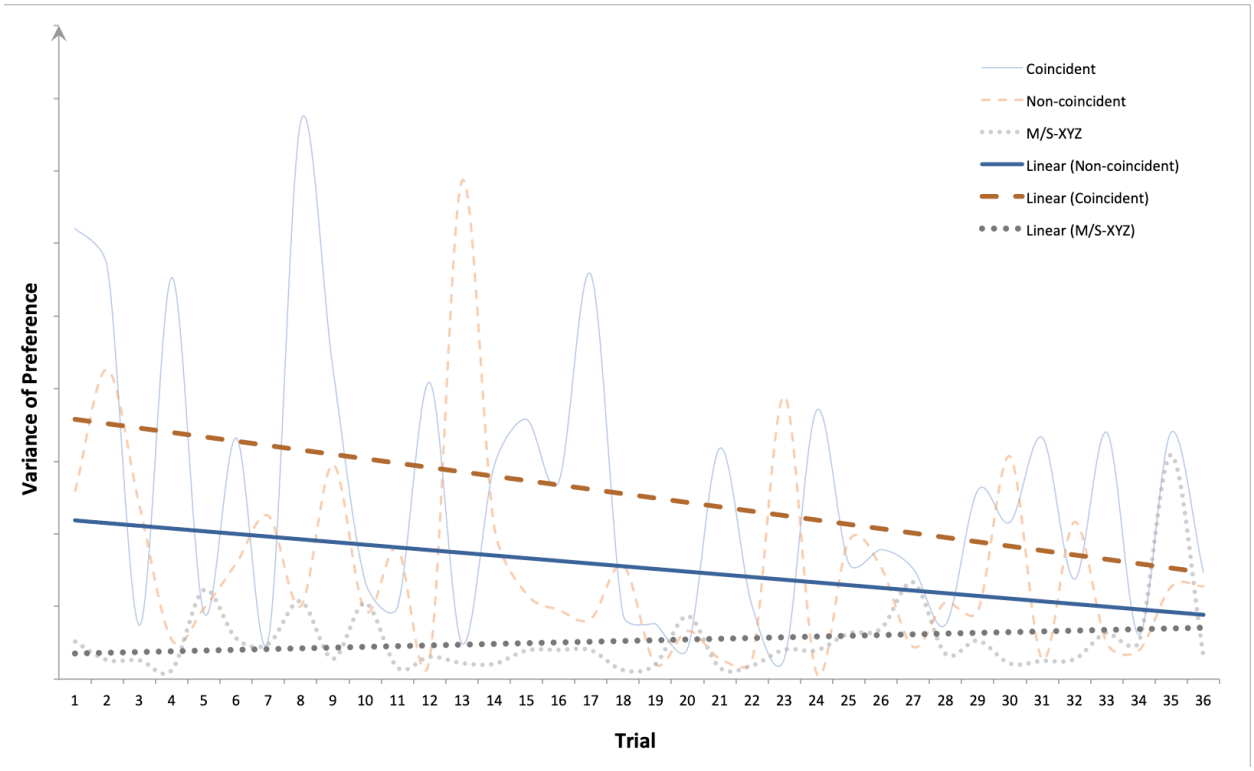


Figure 125: Plot of variance of preference, summed across all subjects, versus trial. The light lines are plots of the raw data, while the heavier lines are trends showing the overall direction of variance using a linear best-fit method. Familiarity with the array seems to manifest itself over time as variance decreases significantly for both the coincident and non-coincident arrays.

6.4.2 The Effect of Array

The primary dependent variable under test, *array*, was found to be significant across *expansion onset*, *preference*, and *image breakdown* at a level of $p = 0.01$. Strong agreement was seen in *expansion onset* within the arrays, particularly the M/S-XYZ array (Figure 126).

The *preference* results were much less consistent for each *array* (Figure 127). The mean preferred spread for each array varied by nearly ±18.5° azimuth, with likewise significantly different levels of variance (Table 13). This again reinforces the trend seen in the *expansion onset* results, with M/S-XYZ being the most consistent across all subjects, and the first to display a realistic image.

Figure 126: Histogram of expansion onset, summed across all subjects, versus array. The M/S-XYZ shows both the earliest expansion onset and the strongest agreement between subjects While the coincident and non-coincident arrays also exhibit a great deal of agreement, the peak frequency of response for each is approximately half of that seen with the M/S-XYZ array.

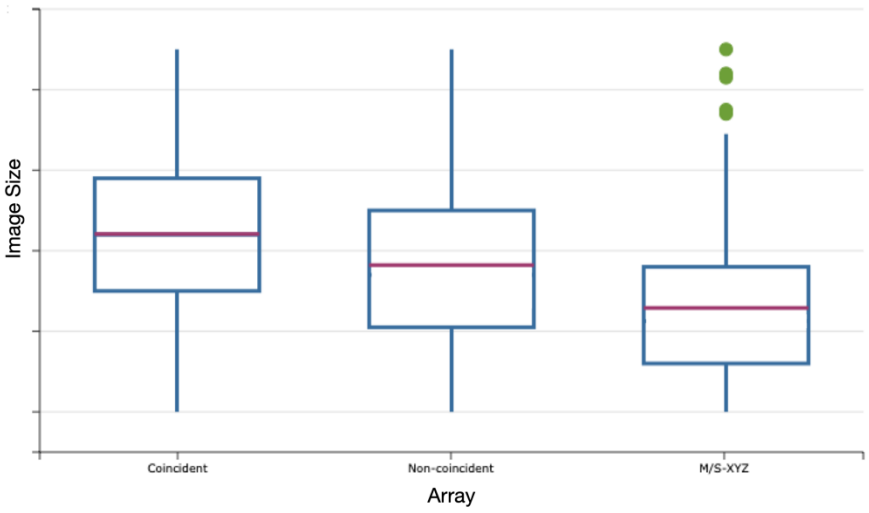


Figure 127: Preference values for all arrays, summed across all subjects.

Preferred Azimuth Spread			
	Coincident	**Non-coincident**	**M/S-XYZ**
Mean	±44.14°	±36.44°	±25.80°
Variance	402.1025	373.49691	324.13028

Table 13: Mean and variance values for preference by array, summed across all subjects.

The *image breakdown* point is the least clear of the three key responses. All three *array* options exhibit an uptick in reports of *image breakdown* at ±70°, with a slight reduction in reports just above ±80°, and then a peak approaching ±90° (Figure 128). Also, in a reversal of the previous two responses, the point of *image breakdown* is least consistent in the M/S-XYZ *array*. The M/S-XYZ array displays what seems to be a clear indicator of multimodality, or at the least a strongly non-parametric population. This is, however, rather hard to judge given the somewhat small sample size.

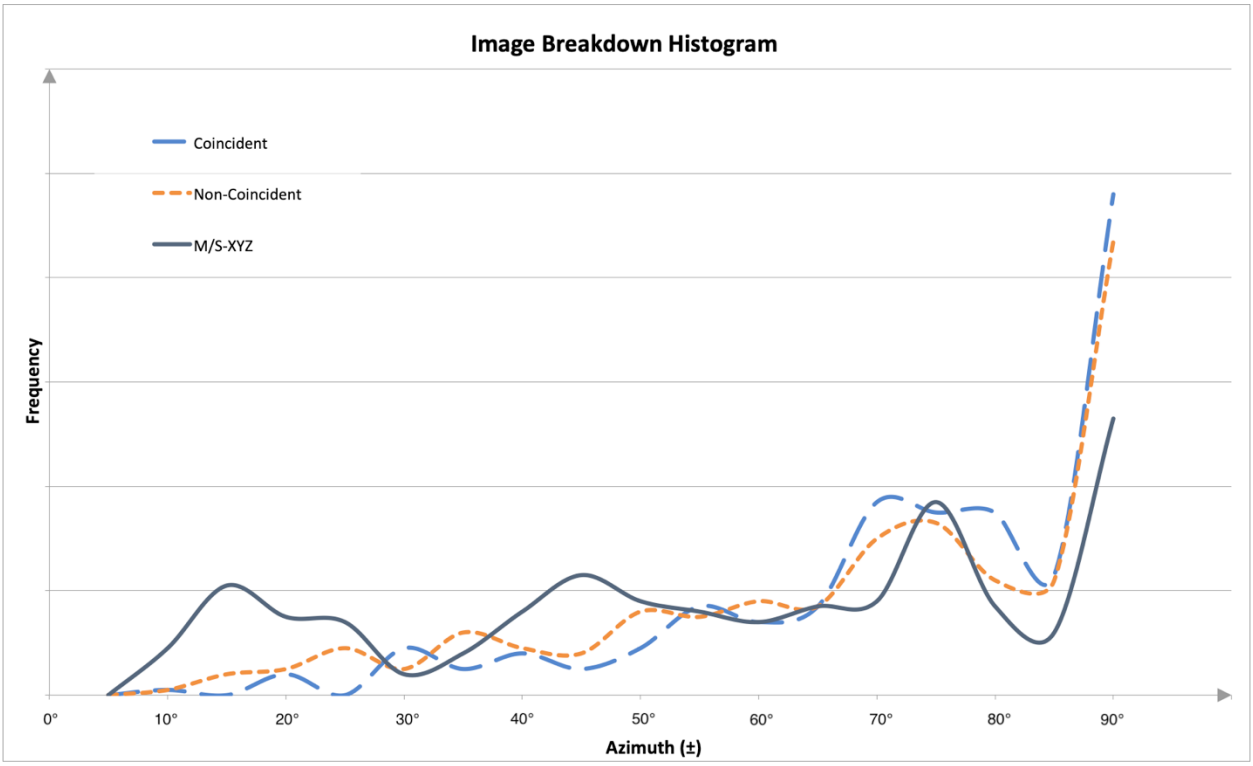


Figure 128: Histogram of image breakdown (in ±° azimuth) by array, summed across all subjects. While all array options exhibit the expected trend towards breakdown at the extreme expansion of the image, the ±15°, ±45°, and ±75° peaks in the M/S-XYZ array are somewhat odd.

6.4.3 Image Range

Table 14 details the expansion onset range, summed across all subjects, versus array. The M/S-XYZ shows both the earliest expansion onset and the strongest agreement between subjects.

While the coincident and non-coincident arrays also exhibit a great deal of agreement, the coincident array exhibits the latest onset.

Expansion Onset

	Azimuth				*Elevation*			
	Mean	**Q1**	**Median**	**Q3**	**Mean**	**Q1**	**Median**	**Q3**
M/S-XYZ	±11.57°	±7°	±10°	±15°	-2.57°/+15.93°	-3.33°/+9°	-2.22°/+13°	-1.56°/+21°
Coincident	±25.72°	±14.25°	±21°	±31	-5.72°/+33.71°	-6.89°/+21°	-4.67°/+31°	-3.17°/+45°
Non-coincident	±20.46°	±11°	±18°	±25°	-4.55°/+27.68°	-5.56°/+16°	-4°/+25°	-2.44°/+37°

Table 14: Array Expansion Onset: azimuth and elevation.

Table 15 details the image breakdown range, summed across all subjects, versus array. All array options exhibit the expected trend towards breakdown at the extreme expansion of the image with the M/S XYZ exhibiting the earliest breakdown followed by the non-coincident and lastly the coincident.

Image Breakdown

	Azimuth				*Elevation*			
	Mean	**Q1**	**Median**	**Q3**	**Mean**	**Q1**	**Median**	**Q3**
M/S-XYZ	±56.11°	±38°	±59.5°	±89°	-12.44°/+61.5°	-16.89°/+49°	-13.11°/+67°	-8.44°/+78°
Coincident	±71.61°	±62°	±75°	±89°	-15.91°/+75.24°	-19.78°/+68°	-16.67°/+77°	-13.78°/+89.75°
Non-coincident	±67.04°	±53°	±71.5°	±76°	-14.91°/+71.5°	-19.78°/+61°	-16°/+75°	-11.84°/+89°

Table 15: Array Image Breakdown: azimuth and elevation.

Figure 129 displays a radar graph of the Image Onset referenced to the mean azimuth and elevation, and summed across all subjects, versus array. Figure 130 displays a radar graph of the Image Breakdown referenced to the mean azimuth and elevation, and summed across all subjects, versus array.

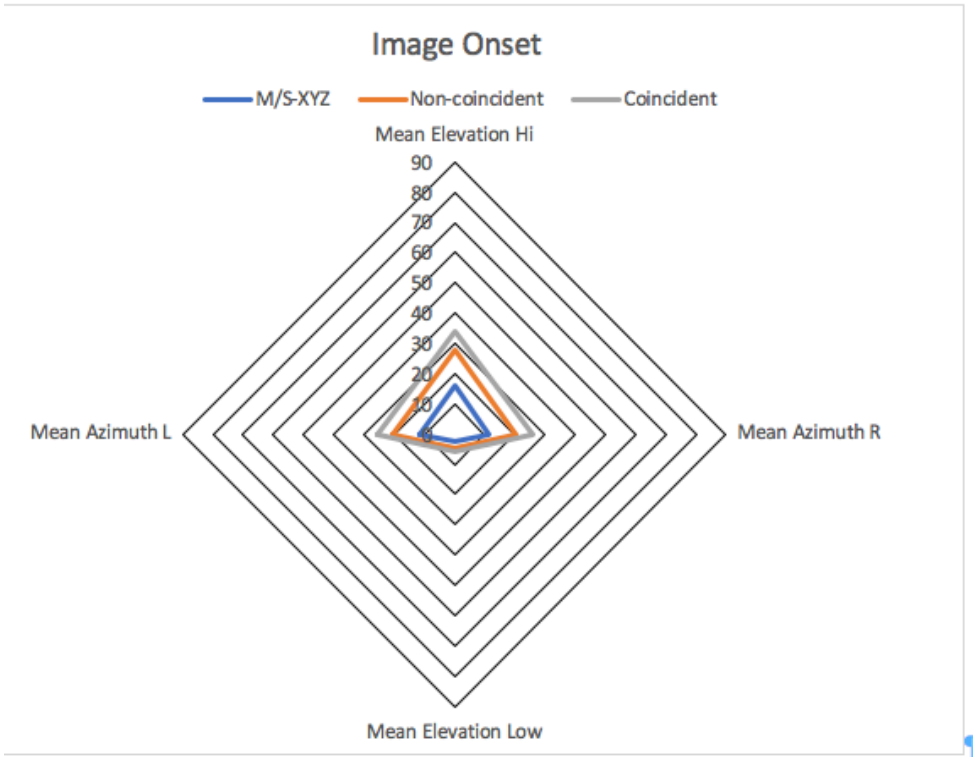


Figure 129: Image onset.

Figure 130: Image breakdown.

6.4.3.1 Range of Arrays

Figure 131, Figure 132 and Figure 133 display the onset-to-breakdown ranges of each array mapped to the Mean values across all subjects.

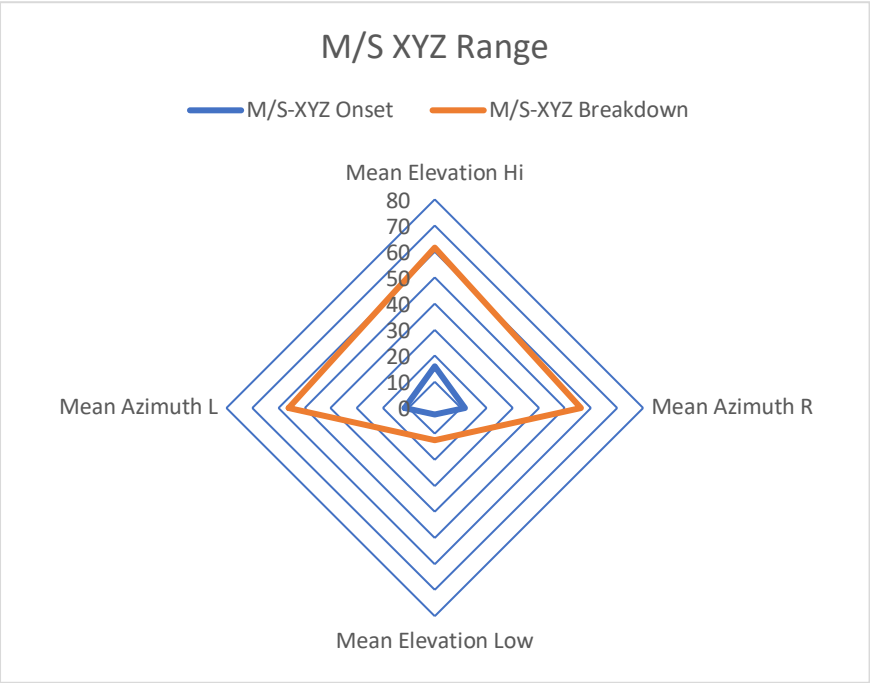


Figure 131: M/S XYZ onset to breakdown range.

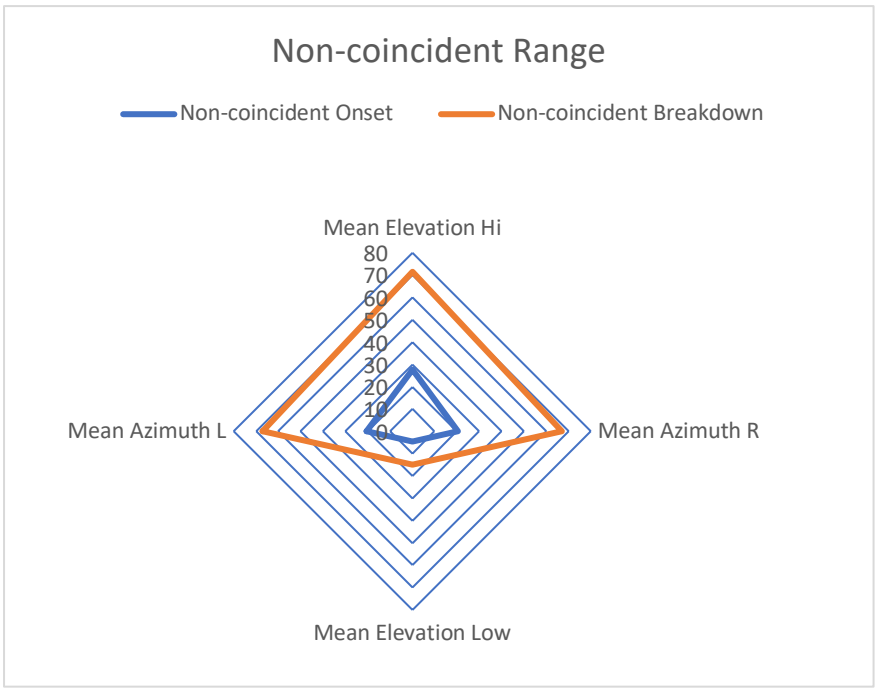


Figure 132: Non-coincident onset to breakdown range.

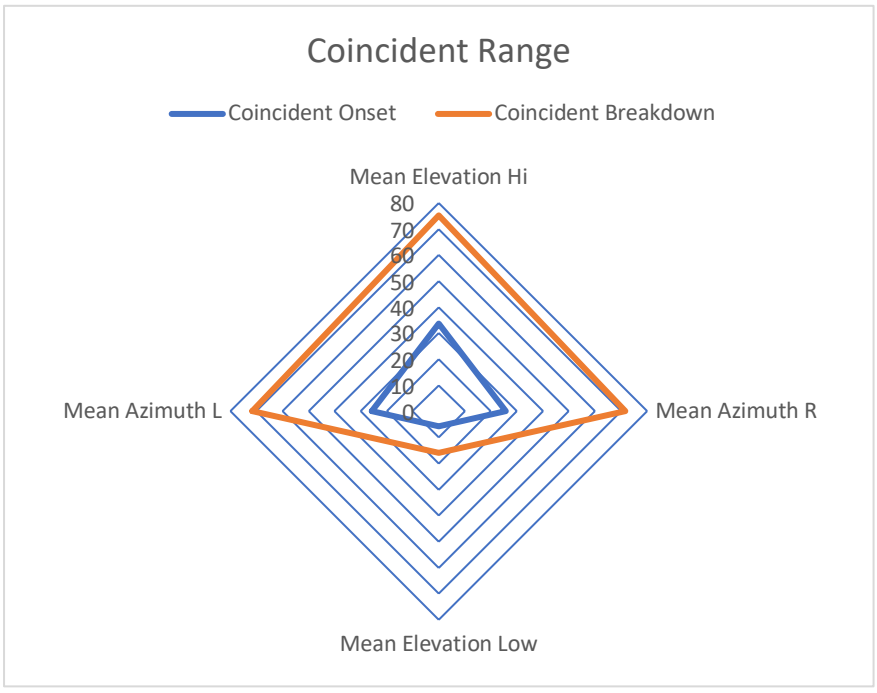


Figure 133: Coincident onset to breakdown range.

6.4.4 The Effect of Experience

Searching for significance using the gathered demographic information also yielded some interesting, if unsurprising, results. The subjects' *age*, years of *musical training*, years of *production experience*, years of *surround sound experience*, and level of *immersive audio experience* all had a significant impact on their responses to the three key questions, particularly *preference*. The way in which these factors influenced *preference* and other responses is, however, less clear. The subjects with the most *production experience*, and the most *immersive audio experience* were neither the most consistent, nor showed any particular trend towards a larger or smaller image size *preference*. In fact, the most consistent subject had 10 years of *production experience,* while the five more experienced subjects displayed a substantially greater variance of *preference* response. Similar trends were seen in *surround sound experience* and *musical training*, with no clear trends, despite their statistical significance.

6.5 Discussion

It seems clear from the data that the M/S-XYZ array is the perceptually "biggest" of the capture systems under test. The M/S-XYZ displayed the earliest sense of *expansion onset* and first to reach a generally-agreed upon *preference*. Likewise, the upper limits of this array's operational range (10º to 60°) are commensurately lower than the other array options. The early onset of the M/S-XYZ could be attributed to the direct pickup of the center microphone dominating the ambient signals from the horizontal and vertical 'side' microphones. It also exhibited the earliest breakdown, which could be attributed to the monophonic center image receiving no direct-signal support from the 'side' microphones as the imaged expands. This may provide some hope that the M/S-XYZ system could provide an ideal archival capture system in many music situations.

The coincident and non-coincident arrays are much less agreed upon by subjects, in terms of *preference* in particular, but in *expansion onset*, as well. One interpretation is that these systems are more subjective than the M/S-XYZ, while another is that there's a wider range of expectably realistic *image size* for these two systems. The median operational range of the non-coincident array (18º to 70º) and coincident array (21º to 75º) seems to indicate their pickup of more correlated information, requiring greater panned distance to create an acceptable image.

The largest *image size* grouping found in testing is held by the coincident *array*. It seems the image can expand to ±35° ±60° to achieve preferential *image size*, while *preference* is dropping off for the others by ±40° to ±50°.

It is also interesting that, while the data seems to indicate, albeit contentiously, that experience is less important that some might think, the trend in decreasing variance for the coincident and non-coincident arrays may indicate that subjects' familiarity with the test *instruments, arrays*, or even the environment or interface may have kept them from peak performance. There would be no way to tell exactly without running a longer test to see if that decrease in variance flattened out at subjects' most comfortable, consistently arrived at point. That being said, there appears to be no subject apathy or fatigue to speak of within this testing.

6.6 Conclusions

It can be seen in the statistical analysis that these techniques clearly offer a wide operational size range within an immersive environment. In the context of immersive/3D mixing and post-production, a case can be made that they will contribute to a more efficient and improved workflow. Five channels are the maximum required for these arrays. The task of up-mixing and spatializing a mono source can easily include nine extra channels of direct sound and multiple

delays and reverb plug-ins [169, 170]. By consulting the array configurations prior to recording it should be possible to determine the correct array according to achieve the desired image size. Setup time for these arrays is on par with standard stereo pairs, so there is little impedance to their use in typical recording sessions.

With a growing demand to provide 3D and immersive content, it is the hope of the authors that we have provided a window into the possibilities and advantages of these techniques.

6.7 Future Work

This testing yielded an incredible range of data that has yet to be fully examined or identified. The sheer number of possible interactions within the variable *could* yield further information about the roles of experience and instrument on the three key responses, as well as more complex relationships between the points of *image expansion*, *preference*/realism, and *image breakdown.*

Likewise, there is always the opinion of more novice listeners to be taken into account. While expert subjects provide a level of consistency, and therefore statistical power, that novice subjects may lack, the untrained listener is the ultimate market for most audio end-products. A simplified comparative test using novice listeners to correlate their preference to the ranges established in this testing may lead to a more universally applicable set of data.

7 CONCLUSIONS

7.1 Specific Conclusions by Chapter

7.1.1 Chapter 3.2

Chapter 3.2 investigates mixing techniques for popular music in multi-channel playback configurations such as the 22.2 sound system. Methods were developed for expanding the size of a sonic image into two- and three-dimensional planes, and best practices for the distribution of the audio spectrum in a hemispherical multi-dimensional acoustic. The investigation details the expansion of the sonic image of a kick drum in the context of a mix for popular music.

The conclusion was that to achieve three-dimensional believability and immersion the source image needed to be distributed in three distinct planes during playback, i.e. sound radiating from the X, Y, and Z axes. In this context, the lack of multi-dimensional and multi-channel tools for the creation of three-dimensional volumetric images, the necessity to employ track multiplying, coupled with the use of multiple delays and other discrete DSP processes (to create each spatial audio image) required a large time investment and tedious workflow. While there has been much advancement in the are of immersive reverberation, there is still a lack of tools for the processing of multi-channel volumetric images.

As of November, 2019 immersive reverberation for the 22.2 format is supported by Exponential Audio Stratus 3D [182] and SPAT Revolution from Flux Audio [183]. In the context of volumetric processing SPAT Revolusion contains a parameter named the *Spread Factor [184]*.

This parameter spreads out the source audio to the different speakers depending on the selected *Room* panning type. At 0%, the source is spread only to the closest speakers. When set at 100%, the source is spread to all the speakers. No information is given on the operation of this tool. Dolby Atmos is an object-based 7.1 surround sound format that originated in cinema sound. It differs from standard 5.1 and 7.1 configuration by the addition of height loudspeaeker channels. It does not support the 22.2 format, not does in employ a volumetric audio tool [185, 186].

7.1.2 Chapter 3.3

The experiment in Chapter 3.3 sought to determine effective playback levels of height-channel information for immersive music presentation. In this study eight discrete channels of height information were presented in conjunction with an eight-channel discrete multi-channel mix of solo acoustic guitar. The ring of height channels copies the number and placement of the first ring of loudspeakers but positioned 1.5 meters above. This configuration comprises the middle and top layers of the 22.2 sound system.

To create the audio material microphones were placed in the recording studio with positioning, spacing and number that mirrored the loudspeakers in each playback “ring” of the control room.

The musical material to be evaluated for immersion consisted of eight discrete channels of height information presented at five different volume levels, in conjunction with an eight-channel discrete multi-channel mix of the solo guitar. These levels were determined to be fairly equal steps between “full immersion” and “very subtle” immersion.

The findings of this study were:

1) There is a minimum level of height information below which subjects could not differentiate added height content. These levels, -16dB, -22dB, provided the same perceived immersion as the mid eight-channel loudspeakers with no additional immersive content (-144dB).

2) The subjects could perceive three distinct content levels during testing: 0dB (immersive), -6dB (immersive), and the "-16, -22, and -144dB" group (little or no-immersive-content).

3) The level of the immersive content needs to be substantially louder to be perceived, ≥ 10dB.

4) The Preference Question results suggest that subjects preferred a more immersive environment than the more subtle levels of immersion, when given the choice.

7.1.3 Chapter 3.4

Chapter 3.4 details an investigation into methods for the creation of compelling high-fidelity mixes of popular music for immersive listening environments such as the 22.2 sound system. The source material was a commercially released track recorded in the early 1980s. The sources were largely recorded via direct injection (DI), and those recorded with microphones contained little or no natural ambience. Commercially available reverberation, delay, and processing plug-ins were used in the creation of the three-dimensional presentation. The goal of the constructed virtual acoustic was to create listener envelopment (LEV), wide audio source width (ASW), and suggest the impression of three-dimensionality in the final musical presentation. The basis for its construction was taken from researchers on listener envelopment. McGill's *Space Builder* multi-channel reverberation was the only dedicated multi-channel tool employed in the

design of the hemispherical spatial acoustic. Listening tests were conducted to gauge: Immersion, 3-Dimensionality, and to determine any expansion of the sweet spot at the mix position.

The subjects' evaluation of *3-Dimensionality* and *Immersion* suggest that it is possible to create believable three-dimensional content using conventional mixing tools from monophonic source material. Regarding the expansion of the 'sweet spot'; the results were inconclusive. Some subjects reported a very wide sweet spot, between 80-100cm of center, but the results were wide ranging and asymmetrical left-to-right.

7.1.4 Chapter 4

The study in Chapter 4 investigates microphone arrays for music recording to capture vertical and three-dimensional images of single instruments. Three separate arrays were investigated: Coincident, M/S-XYZ and Non-coincident/Five-point configurations. Triad/ABX listening tests were conducted to determine if the subjects could: 1) distinguish the three arrays from a mono/point-source and from one another, 2) perceive a vertical image, and 3) If the arrays provided a more immersive experience than the mono signal. The arrays were presented with the discrete microphone channels level-matched for equal-volume and were also level-matched to one another for equal loudness.

The results strongly suggest that a listener can easily identify the arrays from one another when the signals are assigned to a loudspeaker position that roughly approximates its position in the original capture. The majority of the subjects rated the array images having an extended verticality over the mono source, and 90.9% of the subjects rated the arrays as having greater immersion than the mono image.

7.1.5 Chapter 5

The experiment in Chapter 5 employs a simple graphical method to represent the perceived spatial attributes of three microphone arrays and a mono/point-source signal. Three separate arrays were investigated in this study: Coincident, M/S-XYZ and Non-coincident/Five-point capture. The instruments recorded for this investigation were: trumpet, viola, cello and tenor saxophone. Test subjects were asked to represent the spatial attributes of the perceived audio image on a horizontal/vertical grid and a graduated depth grid via a pencil drawing. Direct sound from the individual microphones was assigned to the specific loudspeakers of the 22.2 multi-channel playback system which most closely mirrored their position during capture. The subjects were positioned in the listening/mix position.

For the listening tests the arrays are presented with the discrete microphone channels level-matched for equal-volume. The individual arrays were also level-matched to one another for equal volume, and to the mono/point-source signal. No effort was made to optimize the instrumental presentation of the arrays.

In examining the subjects' drawings and the statistical analysis it can be seen that these techniques clearly capture much more information than a single microphone. The arrays exhibit a greater extent in every dimension—vertical, horizontal and depth—compared to the monophonic image. The statistical trends show that the spatial characteristics of each array are consistent across each dimension. In examining the images provided by these arrays in the context of immersive/3D mixing and post production, a case can be made that they will contribute to a more efficient and improved workflow due to the fact that they are easily optimized during mixing or post-production.

7.1.6 Chapter 6

It can be seen in the statistical analysis that these techniques clearly offer a wide operational size range within an immersive environment. In the context of immersive/3D mixing and post production, a case can be made that they will contribute to a more efficient and improved workflow. Five channels are the maximum required for these arrays. The task of up-mixing and spatializing a mono source can easily include nine extra channels of direct sound and multiple delays and reverb plug-ins [169, 170]. By consulting the array configurations prior to recording it should be possible to determine the correct array according to achieve the desired image size. Setup time for these arrays is on par with standard stereo pairs, so there is little impedance to their use in typical recording sessions.

With a growing demand to provide 3D and immersive content, it is the hope of the authors that we have provided a window into the possibilities and advantages of these techniques.

Immersive, *Spatial* and *3D* are all terms for audio formats that include height channels. This area of audio has become increasingly important over the last decade both for researchers and the greater commercial audio industry. This area of research will impact augmented and virtual reality, gaming, live sound, music presentation, headphone and mobile technology as well as home theatre, cinema and broadcast. Although it is in the early stages—from recording to mixing and delivery—the commercial momentum has arrived to propel this technology into the mainstream.

This research has touched upon the first two areas: mixing and recording. But as in the development of stereo, it will take a wider and more diverse group of researchers, content producers, audio practitioners and technological development to propel this complex artform to its

maximum potential. It is the hope of the author that the ideas, observations and results will serve those who will follow.

8 FUTURE WORK

The mixing and recording of popular music for immersive presentation is in its infancy. Every aspect for this type of capture and presentation will require much work by researchers and practitioners, as well as both hardware and software developers for it to be understood, mastered and integrated into the mainstream.

Practitioners, researchers and industry must collaborate to improve, expand and develop a wider range of recording techniques, advance mixing techniques, multi-channel processing tools for spatial panning, audio object expansion, the creation of immersive environments, as well as conventional multi-channel tools such as compression and equalization. Lastly workflows need to be developed within the various DAWs employed by each industry to streamline and optimize the delivery of spatial content.

Qualitative perceptual research needs to continue to be developed and expanded. But for this to occur, researchers must have access to accurate immersive laboratories. Inaccurate listening and test environments will not yield reliable results, and at the date of this writing many researchers are forced to work in less the optimum conditions. The creation of these facilities will require both ideological and financial commitments from both industry and educational institutions.

8.1 Objective Measure of Vertical Imaging Microphone Arrays

It would be useful to investigate an objective measurement and analysis methodology to examine the signals provided by vertical imaging microphone arrays. These methods could offer

insight into the subjective differences among the various microphone techniques. A beginning method could be the measurement of interaural cross-correlation from a binaural head.

8.2 Research Listening Environments

The monitoring environments in which immersive research is taking place lacks uniformity. Most research by others cited in this dissertation was conducted in 9.1 to 11.1 configurations that fall within some variation of those proposed by Dolby or Auro 3D formats. Little research is taking place in configurations that include both height and bottom channels. The current research has found that three layers of loudspeakers are vital to create realism and impact as well as to optimize the potential of vertical imaging. Howie et al. [187] has reported "*clear perceptual differences between 22.2 and 11.1, 10.2, and 9.1*" in the areas of spatial impression, envelopment, depth and presence. It has also been reported that floor-level bottom layer loudspeakers [18] allow for the creation of sound scenes with both realistic horizontal and vertical extent which is vital to true 3D reproduction.

Future work needs to be undertaken to define a baseline for immersive realism and what minimum loudspeaker configuration would be appropriate.

8.2.1 A Proposed 27.2 Configuration for Maximum Immersion

Over the course of the recordings conducted during this research it was determined that to exploit the potential of vertical and three-dimensional audio images that it would be beneficial for the loudspeaker configuration to possess the capability to reproduce these images at every point across the 360º sound field. The proposed configuration to achieve this aim is a 27.2 multi-channel system (Figure 134) building on the 22.2 system developed my NHK.

The configuration of the 27.2 surround sound system would consist of:

- Middle layer loudspeakers (10 channels): FC 0º, FLc -30º, FL -60º, FRc +30º, FR +60º, SiL -90º, SiR +90º, BL -135º, RL+135º, BC 180º.
- Upper layer loudspeaker (9 channels): TpFC 0º, TpFL -60º, TpFR +60º, TpSiL -90º, TpSiR +90º, TpBL -135º and TpRL +135º, TpBC 180º, TpC 0º overhead.
- Bottom Layer (8+2 channels): BtFC 0º, BtFL -60º, BtFR +60º, BtSiL -90º, BtSiR +90º, BtBL -135º and BtRL +135º, BtBC 180º, LFE1, LFE2.

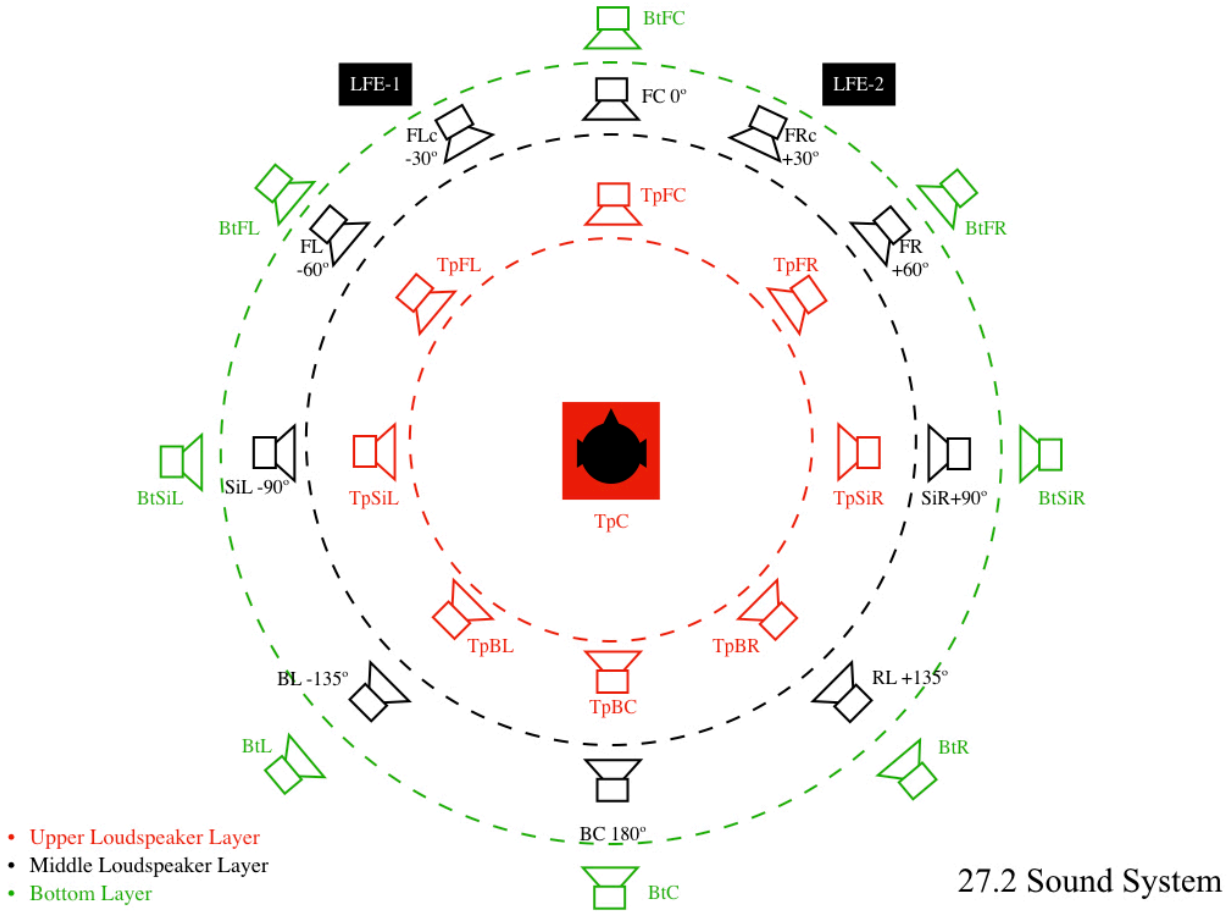


Figure 134: Proposed 27.2 Maximum Immersive Configuration.

It was observed by the researchers that to fully exploit the potential of vertical audio images the LCR bottom layer of the 22.2 system limited the panning of full-height images to the frontal soundscape. Test recordings of solo piano, pop and jazz ensembles were conducted with BtSiL and BtSiR loudspeakers added to the 22.2 system. Informal assessment of the recordings were

undertaken by faculty and students of McGill Sound Recording and the Faculty of Music, as well as audio professionals and musicians from the Montreal area. It was concluded that the addition of the BtSiL and BtSiR loudspeakers greatly improved the immersive experience.

8.3 Investigation Into the Benefits of Bottom Layer Loudspeakers

Much research has investigated height channel impact on the immersive experience (Oode et al. [188], Kim et al. [28], Bowles [120], Lee et al.[189], Kamekawa et al. [27] and King et al. [123]), but little formal work (outside Hamasaki [1]) has been conducted on the impact of bottom layer loudspeakers.

The current research into vertical imaging would benefit greatly from advancement in this area.

8.4 Continued Research in Vertical Audio Imaging

Academic investigation as well as practical applications of vertical audio imaging are in their infancy, and have great potential for the improvement of the immersive experience. Combined with continuing research in8.2 immersive playback environments and a greater understanding of the benefits and application of bottom layer loudspeakers, this field presents a great many unknowns and a vast landscape for investigation.

8.5 More Uniform Test Methods

The concepts of describing spatial perception are quite complex and subjective. While there has been a concerted effort within the research community to categorize and more specifically define the descriptors used in the study of spatial audio [133, 136, 139-142, 150, 151, 173, 190-

192] testing reported in much of the literature is quite diverse, is frequently poorly described and executed, and (as described in section 2.2.7) is performed on a wide variety of playback configurations. This makes the comparison and data and conclusions difficult and misleading.

A concerted effort to standardize test loudspeaker configurations and test methodologies would provide clearer answers and a less confusing path toward understanding the workings of spatial audio. Standardized reference loudspeaker systems could be used to help establish better immersive audio experiences for consumers. Clearly, very few individuals will have such systems for domestic listening, but effective techniques and technologies developed in research systems would inform trickle-down technologies for the consumer.

9 GLOSSARY OF TERMS

A

ABU

Asia-Pacific Broadcasting Union. It has over 280 members in 57 countries and regions covering an area stretching from Turkey in the west to Samoa in the east, and from Mongoliain the north to New Zealand.

ASW

Apparent Source Width is the audible impression of a spatially extended sound source.

ATSC

Advanced Television Systems Committee. It sets standards for digital television transmission over terrestrial, cable, and satellite networks primarily in the United States, Mexico and Canada.

B

BBC

British Broadcasting Corporation.

C

CD

Compact Disc is a digital optical data storage format that was co-developed by Philips and Sony

and released commerically in 1982.

D

DAW

A digital audio workstation (DAW) is an electronic device or application software used for recording editing and producing audio files.

dB

Decibel is a unit of measurement used to express the ratio of one value of a power to another on a logarithmic scale.

DIN

A stereo microphone technique specified by the German national standards organization Deutsches Institut für Normung.

DTS

Digital Theatre Sound is a multi-channel surround technolology.

E

ETRI

Electronics Telecommunications Research Institute in South Korea.

I

IID

Interaural Intensity Differences of sounds arriving at each ear.

IPD

Interaural Phase Differences of sounds arriving at each ear.

ITD

Interaural Time Differences of sounds arriving at each ear.

ITU

International Telecommunications Union is the United Nations agency specialized for information and communication technologies.

L

LFE

Low Frequency Effect channel is intended for deep, low-frequency sounds ranging from 3-150 Hz. It is normally sent to a loudspeaker designed for low-frequency sounds called a subwoofer.

LEV

Listener Envelopment is the sense of being immersed in or surrounded by the sound field.

N

NOS

A stereo microphone technique developed by the Dutch Broadcasting Foundation Nederlandse Omroep Stichting.

O

OCT

Optimized Cardioid Triangle is a multi-channel microphone technique developed by Günther Theile.

ORTF

A stereo microphone technique developed by the French broadcasting organization Office de Radiodiffusion-Television Francaise.

R

RAI

A stereo microphone technique developed by the Italian broadcasting agency Radiotelevisione Italiana.

RIAA

Recording Industry Association of America. It is the trade organization that represents the recording industry in the United States.

RRA

South Korea’s National Radio Research Agency.

S

SAT

Society for Arts and Technology based in Montreal, Quebec, Canada.

SDDS

Sony Dynamic Digital Sound is a multi-channel surround technolology.

SMPTE

The Society of Motion Picture and Television Engineers. It is a global professional association, of engineers, technologists, and executives working in the media and entertainment industry

SPL

Sound Pressure Level usually measured in decibels.

Surround Sound

5.1 surround sound ("five-point one") is the common name for six channel surround sound audio systems. 5.1 is the most commonly used layout in home theatre.

T

THX

High fidelity audio/visual reproduction standards for movie theaters, screening rooms, home theaters, computer speakers, gaming consoles, car audio systems, and video games.

U

UHDTV

Ultra High Definition Television.

V

VOG

Voice Of God is the center-overhead loudspeaker in immersive loudspeaker systems.

10 BIBLIOGRAPHY

1. Hamasaki, K. Hiyama, K.; Okumura, R., The 22.2 Multichannel Sound System and Its Application, in AES Convention:118. 2005, Audio Engineering Society: Barcelona, Spain.
2. Dolby Laboratories Inc., Dolby® AtmosTM Next-Generation Audio for Cinema. 2013, Dolby Laboratories, Inc.
3. Van Baelen, W.; Bert, T.; Claypool, B.; Sinnaeve, T., Auro-3D A new dimension in cinema sound. 2013, Auro Technologies NV.
4. Dolby Laboratories, I. Dolby Atmos for Home Theatre. 2016.
5. Olufsen, B. 2018; Available from: https://www.bang-olufsen.com/en/solutions/automotive.
6. Wilkins, B., 2018; Available from: https://www.bowerswilkins.com/car-audio/maserati.
7. Kardon, H.; Available from: https://www.harmankardon.com/automotive-technologies.html.
8. Sennheiser. 2018; Available from: https://en-us.sennheiser.com/finalstop.
9. 3D Soundlabs. 2018; Available from: http://www.3dsoundlabs.com/ambisonics-binaural-head-tracking/.
10. Storm Audio. 2018; Available from: https://www.stormaudio.com/en/features/sphereaudio/.
11. Bube, S.; Fabris, C.; Hohberger, T.; Köhler, A.; Liebetrau, J.; Sporer, T.; Walther, A., Perceptual Evaluation of Algorithms for Blind Up-mix, in 121 AES Convention. 2006, Audio Engineering Society: San Francisco, CA, USA.

12. Iosono GmbH, I., Anymix 1.5 AAX User Manual. 2013, IOSONO GmbH: Germany.
13. Dolby Laboratories Inc. Dolby AC-4 Audio System. 2016.
14. Jot, J.-M., Fejzo Z., Beyond Surround Sound – Creation, Coding and Reproduction of 3-D Audio Soundtracks, in Audio Engineering Society 131st Convention. 2011, Audio Engineering Society: New York, NY, USA.
15. Martinet, G., SPAT Revolution. 2018, Flux:.
16. DearVR. 2018. https://www.dearvr.com
17. Facebook 360. 2017, Occulus.com.
18. Howie, W.; King, R.; Martin, D., A Three-Dimensional Orchestral Music Recording Technique, Optimized for 22.2 Multichannel Sound, in AEs 141 Convention. 2016, Audio Engineering Society: Los Angeles, CA.
19. Hamasaki, K.; Shinmura, T.; Akita, S.; Hiyama, K., Multichannel Recording Techniques for Reproducing Adequate Spatial Impression, in AES Conference:24th International Conference: Multichannel Audio, The New Reality. 2003, Audio Engineering Society: Banff, Alberta, Canada.
20. Lee, H., Gribben, C. Capturing and Rendering 360º VR Audio Using Cardioid Microphones. in 2016 AES International Conference on Audio for Virtual and Augmented Reality. 2016. Los Angeles, CA: Audio Engineering Society.
21. Hinata, T., Ootakeyama, Y., Sueishi, H., Live Production of 22.2 Multichannel Sound for Sports Programs, in AES 40th International Conference: Spatial Audio: Sense the Sound of Space. 2010, Audio Engineering Society: Tokyo, Japan.
22. ITU-R BS.2159-4 Multichannel sound technology in home and broadcasting applications. 2012, International Telecommunications Union: Geneva, Switzerland.

23. Stenzel, H., Scuda, U., Producing Interactive Immersive Sound for MPEG-H: A Field Test for Sports Broadcasting, in 137 AES Conference. 2014, Audio Engineering Society: Berlin, Germany.
24. Williams, M., The Psychoacoustic Testing of the 3D Multiformat Microphone Array Design, and the Basic Isosceles Triangle Structure of the Array and the Loudspeaker Reproduction Configuration., in Audio Engineering Society 134th Convention. 2013, Audio Engineering Society: Rome, Italy.
25. Rumsey, F., Spatial Audio. 2013, Burlington, MA: Focal Press.
26. Hamasaki, K., Hiyama, K., Nishiguchi, T., Okumura, R., Effectiveness of Height Information for Reproducing the Presence and Reality in Multichannel Audio System, in AES Convention:120. 2006, Audio Engineering Society: Paris, France.
27. Kamekawa, T.; Toru A.; Date, T.; Enatsu, M., Evaluation of Spatial Impression Comparing 2ch Stereo, 5ch Surround, and 7ch Surround with Height Channels for 3D Imagery, in 130 AES Convention. 2011, Audio Engineering Society: London, UK.
28. Kim, S., Ko D., Nagendra, A., Woszczyk, W., Subjective Evaluation of Multichannel Sound with Surround-Height Channels, in Audio Engineering Society 135th Convention. 2013, New York, USA: New York, NY, USA.
29. Wikipedia 5.1 surround sound. 2015.
30. Gerzon, M.A., Periphony: With-Height Sound Reproduction. Journal of Audio Engineering Society, 1973. 22(1): p. 2-10.
31. Ehret, A.; Gröschel A.; Purnhagen, H.; Roedén, JJ.onas, Coding of "2+2+2" Surround Sound Content Using the Mpeg Surround Standard, in AES Convention 122. 2007, Audio Engineering Society: Vienna, Austria.

32. Dawson, S., A Letter to Telarc Concerning Telarc's 'Height' Channel. hifi-writer.com, 2003.

33. Dolby Laboratories Inc. Dolby Atmos Home Theater Installation Guidelines. 2014.

34. Auro Technologies, Auro-3D Authoring Tools Guide. 2015, Auro Technologies NV: Belgium.

35. Ambeo VR Mic. 2018; Available from: https://en-us.sennheiser.com/microphone-3d-audio-ambeo-vr-mic.

36. NT-SF1: Soundfield by Rode. 2018; Available from: https://www.rode.com/ntsf1.

37. Kirn, P., Mics that record in "3D" and ambisonics are the next big thing. 2018; Available from: http://cdm.link/2018/09/3d-ambisonic-microphones/.

38. Godfrey, J., Amos, S., Sound Recording and Reproduction. 1950, London, UK: Iliffe & Sons, Ltd.

39. Handbook for Sound Engineers. Fourth ed. 2008, Oxford, UK: Elsevier, Inc.

40. Clark, H., Dutton, G.; Vanderlyn, P. B., The 'Stereosonic' Recording and Reproducing System: A Two-Channel Systems for Domestic Tape Records. Journal of Audio Engineering Society, 1958. 6(2): p. 102-117.

41. Blumlein, A.D., British Patent Specification 394,325 (Improvements in and relating to Sound-transmission, Sound-recording and Sound-reproducing Systems). Journal of Audio Engineering Society, 1958. 6(2): p. 91-98, 130.

42. Begault, D.R., 3-D Sound for Virtual Reality and Multimedia. 1994, Cambridge, MA, USA: Academic Press.

43. Eargle, J., Music, Sound, and Technology. 2 ed. 1995, New York, NY, USA: Springer Science+Business Media, LLC.

44. Quadraphonic sound. 2018; Available from: https://en.wikipedia.org/wiki/Quadraphonic_sound.
45. Roys, H.E., The Coming of Stereo. Journal of Audio Engineering Society, 1977. 25(10/11): p. 824-827.
46. Davis, C., Frayne, J., The Westrex StereoDisk System. Proceedings of the IRE, 1958. 46(10): p. 1696-1693.
47. RIAA, Disc Phonograph Records for Home Use, Bulletin No. E 4, Also Includes: Standard Recording and Reroducing Characteristics, Bulletin No. E 1. 1978, RIAA.
48. Mowrey, T., Thomas Mowrey Archive. 2015 [cited 2019; Available from: http://quadraphonic.info/Thomas_Mowrey/.
49. Long, R., PIng-Ping-Pong-Pong, in High Fidelity. 1969, High Fidelity: U.S.A.
50. Berkovitz, R., Four-Channel Stereo—the New Surround Sound, in Electronic World. 1970, Electronic World: U.S.A.
51. Guttenberg, S., Whatever happened to 5.1-channel music?, in Stereophile. 2009.
52. Eno, B., Ambient 4: on Land. 1982, EG. p. Music Album.
53. Rumsey, F., Surround Sound, in Immersive Sound: The Art and Science of Binaural and Multi-Channel Audio. 2018, Routledge.
54. Steinberg, J., Snow, W, Auditory perspectives-physical factors. Journal of Audio Engineering Society, 1934.
55. Garity, W., Hawkins, J., Fantasound. SMPTE Motion Imaging Journal, 1941. 37(8): p. 127-146.
56. Hull, J., Surround Sound, in Handbook for Sound Engineers. 2008, Elsevier, Inc.: Oxford, UK.

57. I.T.U., Multichannel Stereophonic Sound System with and without Accompanying Picture, in RECOMMENDATION ITU-R BS.775-1*. 1994.

58. Hanson, J., L., Method and Apparatus for Synchronizing Digital Data Streams, in USPTO, USPTO, Editor. 1992, Imax Corp.: U.S.A.

59. IMAX, The 15/70 Flimmaker's Manual. 1999, IMAX Corporation.

60. International Telecommunications Union, ITU-R BS.775-2: Multichannel stereophonic sound system with and without accompanying picture. 2006, International Telecommunications Union: Geneva, Switzerland.

61. ITU-R BS.2159-7 Multichannel sound technology in home and broadcasting applications. 2015, International Telecommunications Union: Geneva, Switzerland.

62. International Telecommunications Union, Recommendation ITU-R BS.2051-0 Advanced sound system for programme production. 2014, ITU: Geneva, Switzerland.

63. Holman, T., 5.1 Surround Sound, Up and Running. 2007, Oxford, UK: Focal Press.

64. Jae-un, L. 10.2-channel audio becomes int'l standard. Korea.net, 2014.

65. Van Daele, B., Van Baelen, W., PRODUCTIONS IN AURO-3D: Professional workflow and costs. 2012, Auro Technologies NV: Belgium.

66. Kim, S., Height Channels, in Immersive Sound: The Art and Science of Binaural and Multi-Channel Audio. 2018, Routledge.

67. Combined Auro-3D and Dolby Atmos Setup. 2016 [cited 2019; Available from: https://www.stormaudio.com/media/combined_auro3d_and_dolby_atmos_setup_20160219__085700400_1344_13062017.pdf.

68. Theile, G., Wittek, H., Principles in Surround Recordings with Height, in 130 AES Convention. 2011, Audio Engineering Society: London, UK.

69. A.T.S.Committee, A/342 Part 1, Audio Common Elements. 2017, Advanced Television Systems Committee: Washington, DC, U.S.A.

70. Hamasaki, K., 22.2 Multichannel Audio Format Standardization Activity. Broadcast Technology, 2011. 45: p. 14-19.

71. Hamasaki, K., Nakayama, Y.; Nishiguchi, T.; Okumura, R., Wide Listening Area with Exceptional Spatial Sound Quality of a 22.2 Multichannel Sound System, in AES Convention:122. 2007, Audio Engineering Society: Vienna, Austria.

72. S.M.P.T.E., Ultra High Definition Television — Audio Characteristics and Audio Channel Mapping for Program Production. 2008, Society of Motion Picture & Television Engineers: White Plains, NY.

73. Blauert, J., Spatial Hearing: The Psychophysics of Human Sound Localization. 1996, Cambridge, MA MIT Press.

74. Braasch, J., Modelling of Binaural Hearing, in Communication Acoustics. 2005, Springer-Verlag Berlin Heidelberg.

75. Grantham, D.W., Spatial Hearing and Related Phenomena, in Hearing, B.C.J. Moore, Editor. 1995, Academic Press, Inc.: London, UK.

76. Letowski, T. Letowski, S., Localization Error: Accuracy and Precision of Auditory Localization, in Advances in Sound Localization, P. Strumillo, Editor. 2011, IntechOpen.

77. Yost, W.A., Spatial Hearing: The Psychophysics of Human Sound Localization, Revised Edition. Ear and Hearing, 1998. 19(2).

78. Mills, A.W., On the Minimum Audible Angle. J. Acoust. Soc. Am., 1958. 30.

79. Perrott, D., R., Rôle of Signal Onset in Sound Localization. J. Acoust. Soc. Am., 1969. 45: p. 436-445.

80. Kuhn, G.F., Physical acoustics and measurements pertaining to directional hearing, in Directional Hearing. 1987, Springer-Verlag.

81. Zhang, X., P., Psychoacoustics, in Handbook for Sound Engineers. 2008, Elsevier, Inc.: Oxford, UK.

82. Stevens, S., Newman E., The Localization of Actual Sources of Sound. The American Journal of Psychology, 1936. 48(2): p. 297-306.

83. van de Par, S., Kohlrausch, A., A new approach to comparing binaural masking level differences at low and high frequencies. J. Acoust. Soc. Am., 1997. 101: p. 1671-1680.

84. Wallach, H., On Sound Localization. J. Acoust. Soc. Am., 1939. 10: p. 270-274.

85. Steinhauser, A., The theory of binaural audition. A contribution to the theory of sound. Philosophical Magazine, 1879. 7: p. 181-197.

86. Musicant, A.D., Butler, R.A., The influence of pinnae-based spectral cues on sound localization, J. Acoust. Soc. Am., 1984. 75: p. 1195-1200.

87. Lopez-Poveda, E.A., Meddis, R., A physical model of sound diffraction and reflections in the human concha. J. Acoust. Soc. Am., 1996. 100: p. 3248-3259.

88. Hebrank, J., Wright, D., Spectral cues used in the localization of sound sources on the median plane. J. Acoust. Soc. Am., 1974. 56(6): p. 1829-1834.

89. Langendijk, E., Bronkhorst, A., Contribution of spectral cues to human sound localization. J. Acoust. Soc. Am., 2002. 112: p. 1583-1596.

90. Asano, F., Suzuki., Y.; Sone, T., Role of spectral cues in median plane localization. J. Acoust. Soc. Am., 1990. 88: p. 159-168.

91. Bloom, P.J., Determination of monaural sensitivity changes due to the pinna by use of the minimum-audible-field measurements in the lateral vertical plane. J. Acoust. Soc. Am., 1977. 61: p. 820-828.

92. Butler, R.A., Belendiuk, K., Spectral Cues Utilized in the Localization of Sound in the Median Sagittal Plane. J. Acoust. Soc. Am., 1977. 61: p. 1264-1269.

93. Watkins, A.J., Psychoacoustical aspects of synthesized vertical locale cues. J. Acoust. Soc. Am., 1978. 63: p. 1152-1165.

94. Roffler, S., Butler, R., Factors That Influence the Localization of Sound in the Vertical Plane. Journal of the Acoustical Society of America, 1967. 43(6): p. 1255-1259.

95. Algazi, V. R., Avendano, C.; Duda, R. O., Elevation localization and head-related transfer function analysis at low frequencies. J. Acoust. Soc. Am., 2001. 109(3): p. 1110-1122.

96. Lee, H., Phantom Image Elevation Explained, in AES Convention 141. 2016, Audio Engineering Society: Los Angeles.

97. Damaske, P., Wagener, B., Subjective investigations of sound fields. Acustica, 1967. 19(4): p. 198-213.

98. Wallis, R., Lee, H., Directional Bands Revisited, in AES Convention: 138. 2015, Audio Engineering Society: Warsaw, Poland.

99. Wenzel, E.M. et. al, Perception of Spatial Sound, in Immersive Sound: The Art and Science of Binaural and Multi-Channel Audio. 2018, Routledge.

100. Streicher, R., Everest, F., The New Stereo Sound Book. 1988, Pasadena, CA, USA: Audio Engineering Associates.

101. Williams, M., Unified theory of microphone systems for stereophonic sound recording, in Journal of Audio Engineering Society. 1987, Audio Engineering Society.

102. Franssen, N.V., Stereophony. Philips Technical Library. 1962: N.V. Philips.

103. Eargle, J., Basic Stereosonic Recording Techniques, in The Microphone Book. 2005, Focal Press: Burlington MA.

104. Snow, W., B., Basic Principles of Stereophonic Sound, in An anthology of reprinted articles on stereophonic techniques. 1953, Audio Engineering Society: New York, NY, USA. p. 9-31.

105. McGinn, R.E., Stokowski and the Bell Telephone Laboratories: Collaboration in the Development of High- Fidelity Sound Reproduction. Technology and Culture, 1983. 24(1).

106. Hugonnet, C., Walder, P., Stereophonic Sound Recording: Theory and Practice. 1995, West Sussex, UK: John Wiley & Sons, Ltd.

107. Eargle, J., The Microphone Book. 2005, Focal Press: Burlington MA.

108. Rumsey, F., McCormick, T., Sound Recording Applications the Theory. 2014, Focal Press: Oxon, UK.

109. King, R., Main Microphone Systems—How to Record It, in Recording Orchestra and Other Classical Music Ensembles. 2017, Routledge: Abingdon, Oxon, U.K.

110. Howie, W., Martin, D.; Benson, D.; Kelly, J.; King, R., Subjective and objective evaluation of 9ch three-dimensional acoustic music recording techniques, in International Conference on Spatial Reproduction – Aesthetics and Science. 2018, Audio Engineering Society: Tokyo, Japan.

111. Howie, W., Capturing Orchestral Music For Three-Dimensional Audio Playback, in Sound Recording. 2018, McGill: Montreal, QC.

112. Williams, M., Microphone Array Design for localisation with elevation cues, in Audio Engineering Society 132nd Convention. 2012, Audio Engineering Society: Budapest, Hungary.

113. Theile, G., Natural 5.1 Music Recording Based on Psychoacoustic Principals, in 19th International Conference: Surround Sound - Techniques, Technology, and Perception 2001, Audio Engineering Society: Schloss Elmau, Germany.

114. LIndberg, M. 3D Recording with the “2L-cube”.

115. Sennheiser MKH 800 Twin. [cited 2019; Available from: https://en-us.sennheiser.com/studio-condenser-microphone-stereo-surround-mkh-800-twin-ni.

116. Eskow, G., Recording the Orchestra in 9.1: Tonmeister Gregor Zielinsky Teams Up With Sennheiser. Mix, 2016.

117. Wittek, H. Development and application of a stereophonic multichannel recording technique for 3D Audio & VR. 2016.

118. Wittek, H. Systems and Techniques for 3D Recording.

119. Wittek, H., Theile, G., Development and Application of a Stereophonic Multichannel Recording Technique for 3D Audio and VR, in 143 AEs Convention. 2017, Audio Engineering Society: New York, NY.

120. Bowles, D., A Microphone Array for Recording Music in Surround-Sound with Height Channels, in 139 AEs Convention. 2015, Audio Engineering Society: New York, NY.

121. Geluso, P., Capturing Height: The Addition of Z Microphones to Stereo and Surround Microphone Arrays, in Audio Engineering Society 132nd Convention. 2012, Audio Engineering Society: Budapest, Hungary.

122. Ono, K.; Nishiguchi, T.; Matsui, K.; Hamasaki, K., Portable Spherical Microphone for Super Hi-Vision 22.2 Multichannel Audio, in AES Convention:135. 2013, Audio Engineering Society: New York, NY, USA.

123. King, R., Howie, W; Kelly, J., A Survey of Suggested Techniques for Height Channel Capture in Multi-channel Recording, in 140 AES Conference. 2016, Audio Engineering Society: Paris France.

124. Woszczyk, W., Geluso, P., Streamlined 3D Sound Design: The Capture and Composition of a Sound Field, in 145 AES Convention. 2018, Audio Engineering Society: New York, NY.

125. Toole, F.E., Listening Tests-Turning Opinion into Fact. J. Audio Eng. Soc, 1982. 30(6).

126. Toole, F.E., Subjective Evaluation: Identifying and Controlling the Variables in AES Conference:8th International Conference: The Sound of Audio. 1990, Audio Engineering Society: Washington D.C.

127. Toole, F.E., Olive S.., Hearing is Believing vs. Believing is Hearing: Blind vs. Sighted Listening Tests, and Other Interesting Things, in AES Convention:97. 1994, Audio Engineering Society: San Francisco, California.

128. Olive, S., A New Reference Listening Room for Consumer, Professional, and Automotive Audio Research, in AES Convention:126. 2009, Audio Engineering Society: Munich, Germany.

129. Olive, S., A Method for Training Listeners and Selecting Program Material for Listening Tests, in AES 97 Convention. 1994, Audio Engineering Society: San Francisco, CA, USA.

130. Bech, S., Zacharov, N., Perceptual Audio Evaluation, Theory, Method and Application. 2006, West Sussex: John Wiley & Sons Ltd.

131. I.T.U., Recommendation ITU-R BS.1116-1*, Methods for the Subjective Assessment of Small Impairments in Audio Systems Including Multichannel Sound Systems, in Rec. ITU-R BS.1116-1 1. 1997, International Telecommunications Union (ITU).

132. Zacharov, N., Pedersen, T., Spatial Sound Attributes—Development of a Common Lexicon, in 139 AES Covention. 2015, Audio Engineering Society: New York, NY.

133. Le Bagousse, S., Paquier, M., Colomes, C., Categorization of Sound Attributes for Audio Quality Assessment—A Lexical Study. Journal of the Audio Engineering Society, 2014. 62(11): p. 736-747.

134. Rumsey, F., Spatial Quality Evaluation for Reproduced Sound: Terminology, Meaning, and a Scene-Based Paradigm. Journal of Audio Engineering Society, 2002. 50(9): p. 651-666.

135. Rumsey, F., Subjective Assessment of the Spatial Attributes of Reproduced Sound, in AES 15th International Conference. 1998, Audio Engineering Society: Copenhagen.

136. Berg, J., Rumsey, F., Verification and correlation of attributes used for describing the spatial quality of reproduced sound, in 19th International Conference: Surround Sound - Techniques, Technology, and Perception. 2001, Audio Engineering Society.

137. Berg, J., Rumsey, F., Identification of perceived spatial attributes of recordings by repertory grid technique and other methods., in AES Convention:106. 1999, Audio Engineering Society: Munich, Germany.

138. Berg, J., Rumsey, F., Systematic Evaluation of PErceived Spatial Quality, in 24th International Conference: Multichannel Audio, The New Reality. 2003, Audio Engineering Society.

139. Le Bagousse, S., Paquier, M.; Colomes, C., Families of Sound Attributes for Assessment of Spatial Audio, in AES Convention: 129. 2010, Audio Engineering Society: San Francisco, CA, USA.

140. Choisel, S., Wickelmaier, F., Relating Auditory Attributes of Multichannel Reproduced Sound to Preference and to Physical Parameters, in AES Convention: 120. 2006, Audio Engineering Society: Paris, France.

141. Choisel, S. Wickelmaier, F. Extraction of Auditory Features and Elicitation of Attributes for the Assessment of Multichannel Reproduced Sound. Journal of the Audio Engineering Society, 2006. 54(9): p. 815-826.

142. Choisel, S. Wickelmaier, F., Evaluation of multichannel reproduced sound: Scaling auditory attributes underlying listener preference. Journal of the Acoutical Society of America, 2006.

143. Liebetrau, J.; Sporer, T.; Korn, T.; Kunze, K.; Mank, C.; Marquard, D.; Matheja, T.; Mauer, S.; Mayenfels, T.; Möller, R.; Schnabel, M.; Slobbe, B.; Überschär, A., Localization in Spatial Audio—From Wave Field Synthesis to 22.2, in 123 AES Convention. 2007, Audio Engineering Society: New York, NY, USA.

144. Darcy, D.; Terry, K.; Davidson, G.; Graff, R.; Brandmeyer, A.; Crum, P., Methodologies for High-dimensional Objective Assessment of Spatial Audio Quality, in 140 AES Convention. 2016, Audio Engineering Society: Paris, France.

145. Zacharov, N., Pedersen, T., Spatial Sound Attributes—Development of a Common Lexicon, in AES Convention 139. 2015, Audio Engineering Society.

146. Zacharov, N., Koivuniemi, K. Unraveling the perception of spatial sound reproduction: Techniques and experimental design, in 19th International Conference: Surround Sound -

Techniques, Technology, and Perception. 2001, Audio Engineering Society: Schloss Elmau, Germany.

147. Letowski, T., Sound Quality Assessment: Concepts and Criteria, in AES Convention 87. 1989, Audio Engineering Society: New York, NY.

148. Lokki, T., Kajastila, R.; Takala, T., Virtual Acoustic Spaces With Multiple Reverberation Enhancement Systems, in AES Conference: 30th International Conference: Intelligent Audio Environments 2007.

149. Mason, R., Ford, N.; Rumsey, F.; De Bruyn, B., Verbal and Nonverbal Elicitation Techniques in the Subjective Assessment of Spatial Sound Reproduction. Journal of the Acoutical Society of America, 2001. 49(5): p. 366-384.

150. Ford, N., Rumsey, F., de Bruyn, B., Graphical Elicitation Techniques for Subjective Assessment of the Spatial Attributes of Loudspeaker Reproduction – A Pilot Investigation, in Audio Engineering Society 110th International Convention. 2001: Amsterdam, The Netherlands.

151. Ford, N., Rumsey, F., Nind, T., Evaluating Spatial Attributes of Reproduced Audio Events Using a Graphical Assessment Language - Understanding Differences in Listener Depictions, in AES Conference:24th International Conference: Multichannel Audio, The New Reality. 2003, Audio Engineering Society.

152. Usher, J., Woszczyk, W., Visualizing Auditory Spatial Imagery of Multi-channel Audio, in AES Convention: 116. 2004, Audio Engineering Society: Berlin, Germany.

153. Usher, J., Woszczyk, W., Design and Testing of a Graphical Mapping Tool for Analyzing Spatial Audio Scenes, in AES Conference:24th International Conference: Multichannel Audio, The New Reality. 2003, Audio Engineering Society.

154. ITU, Methods for the Subjective Assessment of Small Impairments in Audio Systems Including Multichannel Sound Systems. . 1997, International Telecommunications Union: Geneva, Switzerland.

155. Woszczyk, W., Hong, J., MPEG_McGill_TestSite_Specifications_Studio22_2. 2015, McGill University, Schulich School of Music, Department of Sound Recording: Montreal, QC.

156. ME25. May 2019]; Available from: https://www.me-geithain.de/en/me-25.html.

157. Auro-3D Home theater Setup Installation Guidelines. 2015, Auro Technologies NV: Belgium.

158. The Bang & Olufsen 3D Advanced Sound System. 2019 [cited March 2019; Available from: https://www.bang-olufsen.com/en/solutions/automotive/audi.

159. Continental and Auro Technologies Bring The New True 3D Immersive Sound-Esperience into the Car. 2014 March 2019]; Available from: https://www.auro-3d.com/press/tag/automotive/.

160. Fraunhofer Cingo. 2019 [cited 2019 March]; Available from: https://www.iis.fraunhofer.de/en/ff/amm/prod/audiocodec/audiocodecs/cingo.html.

161. Dirac 3D Audio. 2019; Available from: https://www.dirac.com/3d-audio.

162. Herre, J., Hilpert, J., Kuntz, A., Plogsties, J., MPEG-H 3D Audio—The New Standard for Coding of Immersive Spatial Audio. IEEE JOURNAL OF SELECTED TOPICS IN SIGNAL PROCESSING, 2015. 9(5).

163. Williams, M., The Psychoacoustic Testing of the 3-D Multiformat Microphone Array Design and the Basic Isosceles Triangle Structure of the Array and the Loudspeaker

Reproduction Configuration, in 134 AES Convention. 2013, Audio Engineering Society: Berlin, Germany.

164. Ambeo VR Mic. 2019 [cited 2019; Available from: https://en-ca.sennheiser.com/microphone-3d-audio-ambeo-vr-mic.

165. Nyberg, D., Berg, J., Listener Envelopment – What has been done and what future research is needed?, in AES 124 Convention. 2008, Audio Engineering Society: Amsterdam, Netherlands.

166. Beranek, L.L., Concert and opera halls: how they sound. 1996, Woodbury, NY: Acoustical Society of America through the American Institute of Physics.

167. Bradley, J.; Soulodre, G., Objective measures of listener envelopment. Journal of the Acoutical Society of America, 1995. 98(5): p. 2590-7.

168. Morimoto, M., Iida, K., Furue, Y., Relation between Auditory Source Width in Various Sound Fields and Degree of Interaural Cross-Correlation. Applied Acoustics, 1993. 38: p. 291-301.

169. Martin, B., Mixing Popular Music in Three Dimensions, in Innovation in Music 2015. 2015, In Music: Cambridge, UK.

170. Martin, B., King, R., Three Dimensional Spatial Techniques in 22.2 Multichannel Surround Sound for Popular Music Mixing, in 139 AES Convention. 2015, Audio Engineering Society: New York, NY, USA.

171. Martin, B., King, R.; Leonard, B.; Benson, D.; Howie, W., Immersive Content in Three Dimensional Recording Techniques for Single Instruments in Popular Music, in 138 AES Convention. 2015, Audio Engineering Society: Warsaw, Poland.

172. Hamasaki, K., Nishiguchi, T.; Hiyama, K.; Ono, K., Advanced Multichannel Audio Systems with Superior Impression of Presence and Reality, in AES Convention:116. 2004, Audio Engineering Society: Berlin, Germany.

173. Le Bagousse, S., Colomes, C.; Paquier, M., State of the Art on Subjective Assessment of Spatial Sound Quality, in AES Conference:38th International Conference: Sound Quality Evaluation. 2010, Audio Engineering Society.

174. Meyer, J., Hansen, U., Acoustics and the performance of music : manual for acousticians, audio engineers, musicians, architects and musical instruments makers. 2009, New York, NY, USA: Springer.

175. Pätynen, J., Lokki, T., Directivities of Symphony Orchestra Instruments. Acta Acustica united with Acustica, 2010. 96: p. 138-167.

176. Bech, S., Selection and training of subjects for listening tests on sound-reproducing equipment. Journal of the Audio Engineering Society, 1992. 40(7/8): p. 590-610.

177. Beranek, L.L., Music, Acoustics & Architecture. 1962, New York, London: John Wiley & Sons, Inc.

178. Martin, B., King, R., Woszczyk, W., Subjective Graphical Representation of Microphone Arrays for Vertical Imaging and Three-Dimensional Capture of Acoustic Instruments, Part I, in AES 141 Convention. 2016, Audio Engineering Society: Los Angeles, CA.

179. Martin, B., King R., Woszczyk, W., Microphone Arrays for Three-Dimensional Capture of Acoustic Instruments, in 2016 AES International Conference on Sound Field Control. 2016, Audio Engineering Society: Guilford, UK.

180. Howie, W.; King, R.; Martin, D.; Grond, F., Subjective Evaluation of Orchestral Music Recording Techniques for Three-Dimensional Audio, in AES Convention: 142. 2017, Audio Engineering Society: Montreal, QC.

181. Howie, W., Pop and Rock music audio production for 22.2 Multichannel Sound: A Case Study, in AES Convention 146. 2019, Audio Engineering Society: Dublin, Ireland.

182. Exponential Audio Stratus 3D. 2019 [cited 2019 November 14]; Available from: https://www.timespace.com/products/exponential-audio-stratus-3d.

183. SPAT Revolution. 2019 [cited 2019 November 14]; Available from: https://www.flux.audio/project/spat-revolution/.

184. SPAT Revolution Parameter Guide. 2019, Flux Audio.

185. Dolby Music, Create in Dolby Atmos. 2019 [cited 2019 November 15]; Available from: https://music.dolby.com/dolby-atmos-for-creators/.

186. Dolby Laboratories, I., Authoring for Dolby® AtmosTM Cinema Sound Manual. 2013, Dolby Laboratories, Inc.: San Francisco, CA.

187. Howie, W., King, R.; Martin, D., Listener Discrimination Between Common Speaker-Based 3D Audio Reproduction Formats. Journal of Audio Engineering Society, 2017. 65(10): p. 796-805.

188. Oode, S.; Sawaya, I.; Ono, K.; Ozawa, K., Three-Dimensional Loudspeaker Arrangement for Creating Sound Envelopment, in IEICE Technical Report. 2012.

189. Lee, H., Gribben, C., On the optimum microphone array configuration for height channels, in 134 AES Convention. 2013, Audio Engineering Society: Rome, Italy.

190. Berg, J., Rumsey, F., Systematic Evaluation of Perceived Spatial Quality, in AES 24th International Conference. 2003, Audio Engineering Society: Banff, Alberta.

191. Ford, N., Rumsey F.; Nind, T., Communicating Listeners' Auditory Spatial Experiences: A Method for Developing a Descriptive Language., in AES Convention: 118. 2005, Audio Engineering Society: Barcelona, Spaine.

192. Griesinger, D., Objective Measures of Spaciousness and Envelopment, in AES Conference:16th International Conference: Spatial Sound Reproduction. 1999, Audio Engineering Society: Rovaniemi, Finland.