

**LOCALIZATION PERFORMANCE WITH LOW-ORDER
AMBISONICS AURALIZATION**

By

Ioana Nicola Pieleanu

A Thesis Submitted to the Graduate
Faculty of Rensselaer Polytechnic Institute
in Partial Fulfillment of the
Requirements for the Degree of
MASTER OF SCIENCE

Major Subject: Building Science, Concentration in Architectural Acoustics

Approved by the
Examining Committee:

Rendell Torres, Thesis Adviser

Mendel Kleiner, Member

Ning Xiang, Member

Rensselaer Polytechnic Institute
Troy, New York

August 2004
(For Graduation August 2004)

LOCALIZATION PERFORMANCE WITH LOW-ORDER
AMBISONICS AURALIZATION

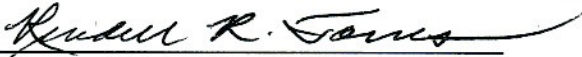
By

Ioana Nicola Pieleanu

A Thesis Submitted to the Graduate
Faculty of Rensselaer Polytechnic Institute
in Partial Fulfillment of the
Requirements for the Degree of
MASTER OF SCIENCE

Major Subject: Building Science, Concentration in Architectural Acoustics

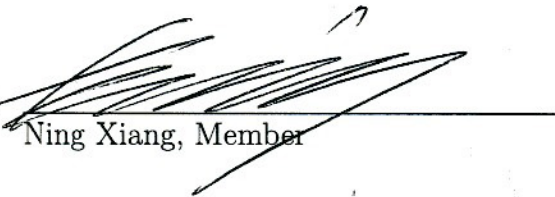
Approved by the
Examining Committee:



Rendell Torres, Thesis Adviser



Mendel Kleiner, Member



Ning Xiang, Member

Rensselaer Polytechnic Institute
Troy, New York

August 2004
(For Graduation August 2004)

© Copyright 2004
by
Ioana Nicola Pieleanu
All Rights Reserved

CONTENTS

LIST OF TABLES	vi
LIST OF FIGURES	viii
ACKNOWLEDGMENT	xii
ABSTRACT	xiii
1. INTRODUCTION	1
1.1 Motivation and Goals	1
1.2 The Concept of Virtual Environments	2
1.3 A Historical Overview on Sound Reproduction	3
1.4 The Two-Channel Stereophonic System: First Spatial Sound Reproduction Technique	6
1.5 The Beginnings of Surround Sound	8
2. SOUND LOCALIZATION: CONCEPTUAL OVERVIEW AND PRIOR WORK	10
2.1 Localization and Localization Blur	10
2.2 A History of Sound Localization Research	10
2.3 The Concepts of Sound Event vs. Auditory Event	11
2.4 Localization in the Horizontal Plane	12
2.5 Localization in the Median Plane	13
2.6 Baseline localization test: Investigation of the Auditory System Localization Blur in the Horizontal Plane	13
2.6.1 Experimental Design	13
2.6.1.1 Test environment	14
2.6.1.2 Test Sound Sources	14
2.6.1.3 Subject Population	14
2.6.1.4 Test Stimuli	14
2.6.1.5 Test Procedure and Data Acquisition	15
2.6.2 Test Results	16

3.	AMBISONICS: THEORY	17
3.1	Overview	17
3.2	The Psychoacoustics Behind the Ambisonics Concept	18
3.3	The Encoding Process	21
4.	PERCEPTUAL INVESTIGATION OF THE AMBISONICS RENDERING SYSTEM	25
4.1	Goals for the Perceptual Experiments	25
4.2	Experimental Design	25
4.2.1	Test Environment	25
4.2.2	Test Virtual Sound Sources	26
4.2.3	Test Stimuli	29
4.2.4	Subject Population	30
4.2.5	Test Cases and Task Distribution	31
4.2.6	The User Interface	31
4.2.7	Test's Design Procedure and Methodologies	32
4.3	Results of the Perceptual Localization Experiment	38
4.4	Perceptual Investigation Results with Mean Signed Error Values	39
4.4.1	Mean Signed Localization Error Results in the Horizontal Plane, for a First-Order Ambisonics Sound Rendering System	39
4.4.2	Mean Signed Localization Error Results in the Horizontal Plane, for a Second-Order Ambisonics Sound Rendering System	41
4.4.3	Comparison of the localization accuracy in the horizontal plane between first and second order Ambisonics rendering	42
4.4.4	Mean Signed Localization Error Results in the Median Plane, for a First-Order Ambisonics Sound Rendering System	44
4.4.5	Mean Signed Localization Error Results in the Median Plane, for a Second-Order Ambisonics Sound Rendering System	46
4.4.6	Comparison of the localization accuracy in the median plane between 1st and 2nd order Ambisonics rendering	47
4.5	Perceptual Investigation Results with Absolute Error Values	52
4.6	Correlation between stimuli location and perceived location	55
5.	AMBISONICS SYSTEMS PERFORMANCE INVESTIGATION	58
5.1	Goals for the Performance Simulation Investigation	58
5.2	Simulation Procedure and Design	58

5.3	Simulation Results	61
5.3.1	Ideal simulation of the wavefront generated by a single sound source	61
5.3.2	Simulation of the wavefront using the ideal dodecahedron loudspeakers configuration	62
5.3.3	Simulation of the wavefront using the warped dodecahedron setup	66
6.	CONCLUSIONS	86
6.1	Research Results	86
6.1.1	System Implementation	86
6.1.2	Perceptual Investigation	87
6.1.2.1	Localization in the Horizontal Plane	87
6.1.2.2	Localization in the Median Plane	88
6.1.3	System Performance Investigation	89
6.2	Final Conclusion and Future Work	91
	LITERATURE CITED	92

LIST OF TABLES

2.1	Localization in the median plane as a function of signal content [1] . . .	13
4.1	Localization Test Cases	32
4.2	Ambisonics Encoding Coefficients, up to the Second Order	34
4.3	Cartesian Coordinates of the Loudspeakers in the Ambisonics Rendering System	34
4.4	Ambisonics Decoding Coefficients, up to the Second Order, for a Tilted Dodecahedron Loudspeaker Configuration	36
4.5	Cartesian Coordinates of the Loudspeakers in the Modified Decoding Configuration	37
4.6	Varying Components of the Experiment	37
4.7	Mean and Standard Deviation Values of the Azimuth Localization Results, with First Order Ambisonics	40
4.8	Best-case Mean and Standard Deviation Values of the Azimuth Localization Results, with First Order Ambisonics	40
4.9	Mean and Standard Deviation Values of the Azimuth Localization Results, with Second Order Ambisonics	42
4.10	Best-case Mean and Standard Deviation Values of the Azimuth Localization Results, with Second-Order Ambisonics	42
4.11	Azimuth Localization Comparison for First and Second Order Ambisonics Renderings, over the entire subject population	43
4.12	Azimuth Localization Comparison for First and Second Order Ambisonics Renderings, for the subject with best-performance	44
4.13	Mean and Standard Deviation Values of the Elevation Localization Results, with First-Order Ambisonics	45
4.14	Best-case Mean and Standard Deviation Values of the Elevation Localization Results, with First-Order Ambisonics	45
4.15	Mean and Standard Deviation Values of the Elevation Localization Results, with Second Order Ambisonics	46

4.16	Best-case Mean and Standard Deviation Values of the Elevation Localization Results, with Second-Order Ambisonics	47
4.17	Median Plane Localization Comparison for First and Second Order Ambisonics Renderings	52
4.18	Median Plane Localization Comparison for First and Second Order Ambisonics Renderings, for the subject with best-performance	53
4.19	Absolute Error Analysis of Localization Results	54
4.20	Absolute Error Analysis of Localization Results, for subject with best performance	55
4.21	Correlation Coefficients of Localization Results	57
4.22	Correlation Coefficients of Localization Results	57

LIST OF FIGURES

2.1	Baseline localization test set-up	15
2.2	Area of consistency of auditory localization in the horizontal plane . . .	16
3.1	Ambisonics Coordinate System	22
4.1	Ideal dodecahedron configuration	26
4.2	Warped dodecahedron configuration	27
4.3	Virtual Sources Location in the Horizontal Plane	28
4.4	Virtual Sources Location in the Vertical Plane	29
4.5	Photograph of the testing environment	31
4.6	Projection Screen, for localization in the Horizontal Plane	33
4.7	Projection Screen, for localization in the Median Plane	33
4.8	Shelf Filter applied to the velocity components	35
4.9	Shelf Filter applied to the energy components	35
4.10	Localization Results in the Horizontal Plane for First Order Ambisonics: a & b: free field simulations; c & d: reverberant field simulations	48
4.11	Best-case localization results in the Horizontal Plane for first-order Am- bisonics: a - free field simulations; b - reverberant field simulations . . .	48
4.12	Localization Results in the Horizontal Plane for Second Order Ambison- ics: a & b: free field simulations; c & d: reverberant field simulations . .	49
4.13	Best-case localization results in the Horizontal Plane for second-order Ambisonics: a - free field simulations; b - reverberant field simulations .	49
4.14	Localization Results in the Median Plane for First Order Ambisonics: a & b: free field simulations; c & d: reverberant field simulations	50
4.15	Best-case localization results in the Median Plane for first-order Am- bisonics: a - free field simulations; b - reverberant field simulations . . .	50
4.16	Localization Results in the Median Plane for Second Order Ambisonics: a & b: free field simulations; c & d: reverberant field simulations	51

4.17	Best-case localization results in the Median Plane for second-order Ambisonics: a - free field simulations; b - reverberant field simulations . . .	51
5.1	Source 1 angular direction	60
5.2	Source 2 angular direction	61
5.3	Source 3 angular direction	62
5.4	Simulation of an ideal wavefront for <i>Source 1</i> : (a), (b) and (c) 250Hz, 500 Hz and 1 kHz in the horizontal plane; (d), (e) and (f) 250Hz, 500 Hz and 1 kHz in the vertical plane.	63
5.5	Simulation of an ideal wavefront for <i>Source 2</i> : (a), (b) and (c) 250Hz, 500 Hz and 1 kHz in the horizontal plane; (d), (e) and (f) 250Hz, 500 Hz and 1 kHz in the vertical plane.	64
5.6	Simulation of an ideal wavefront for <i>Source 3</i> : (a), (b) and (c) 250Hz, 500 Hz and 1 kHz in the horizontal plane; (d), (e) and (f) 250Hz, 500 Hz and 1 kHz in the vertical plane.	65
5.7	Ideal Ambisonics simulation of the wavefront in the horizontal plane for <i>Source 1</i> : (a), (b) and (c) 250Hz, 500 Hz and 1 kHz with <i>1st order Ambisonics</i> ; (d), (e) and (f) 250Hz, 500 Hz and 1 kHz with <i>2nd order Ambisonics</i>	68
5.8	Ideal Ambisonics simulation of the wavefront in the horizontal plane for <i>Source 2</i> : (a), (b) and (c) 250Hz, 500 Hz and 1 kHz with <i>1st order Ambisonics</i> ; (d), (e) and (f) 250Hz, 500 Hz and 1 kHz with <i>2nd order Ambisonics</i>	69
5.9	Ideal Ambisonics simulation of the wavefront in the horizontal plane for <i>Source 3</i> : (a), (b) and (c) 250Hz, 500 Hz and 1 kHz with <i>1st order Ambisonics</i> ; (d), (e) and (f) 250Hz, 500 Hz and 1 kHz with <i>2nd order Ambisonics</i>	70
5.10	Ideal Ambisonics simulation of the wavefront in the vertical plane for <i>Source 1</i> : (a), (b) and (c) 250Hz, 500 Hz and 1 kHz with <i>1st order Ambisonics</i> ; (d), (e) and (f) 250Hz, 500 Hz and 1 kHz with <i>2nd order Ambisonics</i>	71
5.11	Ideal Ambisonics simulation of the wavefront in the vertical plane for <i>Source 2</i> : (a), (b) and (c) 250Hz, 500 Hz and 1 kHz with <i>1st order Ambisonics</i> ; (d), (e) and (f) 250Hz, 500 Hz and 1 kHz with <i>2nd order Ambisonics</i>	72

5.12	Ideal Ambisonics simulation of the wavefront in the vertical plane for <i>Source 3</i> : (a), (b) and (c) 250Hz, 500 Hz and 1 kHz with <i>1st order Ambisonics</i> ; (d), (e) and (f) 250Hz, 500 Hz and 1 kHz with <i>2nd order Ambisonics</i>	73
5.13	Warped Ambisonics simulation of the wavefront in the horizontal plane for <i>Source 1</i> : (a), (b) and (c) 250Hz, 500 Hz and 1 kHz with <i>1st order Ambisonics</i> ; (d), (e) and (f) 250Hz, 500 Hz and 1 kHz with <i>2nd order Ambisonics</i>	74
5.14	Warped Ambisonics simulation of the wavefront in the horizontal plane for <i>Source 2</i> : (a), (b) and (c) 250Hz, 500 Hz and 1 kHz with <i>1st order Ambisonics</i> ; (d), (e) and (f) 250Hz, 500 Hz and 1 kHz with <i>2nd order Ambisonics</i>	75
5.15	Warped Ambisonics simulation of the wavefront in the horizontal plane for <i>Source 3</i> : (a), (b) and (c) 250Hz, 500 Hz and 1 kHz with <i>1st order Ambisonics</i> ; (d), (e) and (f) 250Hz, 500 Hz and 1 kHz with <i>2nd order Ambisonics</i>	76
5.16	Warped Ambisonics simulation of the wavefront in the vertical plane for <i>Source 1</i> : (a), (b) and (c) 250Hz, 500 Hz and 1 kHz with <i>1st order Ambisonics</i> ; (d), (e) and (f) 250Hz, 500 Hz and 1 kHz <i>2nd order Ambisonics</i>	77
5.17	Warped Ambisonics simulation of the wavefront in the vertical plane for <i>Source 2</i> : (a), (b) and (c) 250Hz, 500 Hz and 1 kHz with <i>1st order Ambisonics</i> ; (d), (e) and (f) 250Hz, 500 Hz and 1 kHz with <i>2nd order Ambisonics</i>	78
5.18	Warped Ambisonics simulation of the wavefront in the vertical plane for <i>Source 3</i> : (a), (b) and (c) 250Hz, 500 Hz and 1 kHz with <i>1st order Ambisonics</i> ; (d), (e) and (f) 250Hz, 500 Hz and 1 kHz with <i>2nd order Ambisonics</i>	79
5.19	First-order Ambisonics, magnified simulation comparison for <i>Source 1</i> : (a), (b) and (c) 250Hz, 500 Hz and 1 kHz with an ideal configuration; (d), (e) and (f) 250Hz, 500 Hz and 1 kHz with a warped configuration.	80
5.20	Second-order Ambisonics, magnified simulation comparison for <i>Source 1</i> : (a), (b) and (c) 250Hz, 500 Hz and 1 kHz with an ideal configuration; (d), (e) and (f) 250Hz, 500 Hz and 1 kHz with a warped configuration.	81
5.21	First-order Ambisonics, magnified simulation comparison for <i>Source 2</i> : (a), (b) and (c) 250Hz, 500 Hz and 1 kHz with an ideal configuration; (d), (e) and (f) 250Hz, 500 Hz and 1 kHz with a warped configuration.	82

- 5.22 Second-order Ambisonics, magnified simulation comparison for *Source 2*: (a), (b) and (c) 250Hz, 500 Hz and 1 kHz with an ideal configuration; (d), (e) and (f) 250Hz, 500 Hz and 1 kHz with a warped configuration. . 83
- 5.23 First-order Ambisonics, magnified simulation comparison for *Source 3*: (a), (b) and (c) 250Hz, 500 Hz and 1 kHz with an ideal configuration; (d), (e) and (f) 250Hz, 500 Hz and 1 kHz with a warped configuration. . 84
- 5.24 Second-order Ambisonics, magnified simulation comparison for *Source 3*: (a), (b) and (c) 250Hz, 500 Hz and 1 kHz with an ideal configuration; (d), (e) and (f) 250Hz, 500 Hz and 1 kHz with a warped configuration. . 85

ACKNOWLEDGMENT

The author would like to thank Dr. Rendell R. Torres for being the advisor of this thesis and to Mr. Paul Henderson and Ms. Evelyn Way for their substantial support and contribution to the accomplishment of this research. Many thanks to the Department of Acoustics at RPI for the opportunity given to be part of its Master program, and in particular to Dr. Mendel Kleiner and Dr. Ning Xiang for having me as a student and for the knowledge they had the patience to share with me. And thank you to my parents for a lifetime of love and support, and to Jorge Castellanos for his perspective on life that he tirelessly shares with me every day.

ABSTRACT

This research investigates the performance of an Ambisonics reproduction system within the context of virtual telepresence environments, e.g., for high-resolution teleconferencing or for transmitting musical performances to remote audiences. To this end, the work examines the perceptual and physical characteristics of first- and second-order Ambisonics reproduction techniques, and whether these techniques provide both accurate sound localization and physical reproduction of a sound field over a reasonable listening area. Results from perceptual experiments on the localization error of the system show values as low as 3° for free field simulations and as low as 6° for reverberant field simulations in the horizontal plane, when using both first- and second-order Ambisonics, for a listening position located at the center of the system. Further work should test for listening locations off center, and determine to what degree the localization error will vary according to the Ambisonics order used. Computed sound field simulations complement the perceptual investigation and show that accurate reconstruction of a wavefront can be obtained with Ambisonics rendering. However, this can be achieved only over a limited listening area, depending on the rendering conditions. Further work should investigate the effective size of the effective listening area as a function of frequency, Ambisonics order, and other system parameters.

CHAPTER 1

INTRODUCTION

Virtual Environments are a large part of today's entertainment industry and a medium of remote communication. To be successful in simulating real life situations, computer based virtual environments aim to reproducing two kinds of human perceptual sensations: feelings of presence and of immersion. As one of this field's development pioneers declared in the early days of this technology [2]:

The ultimate display would, of course, be a room within which the computer can control the existence of matter. A chair displayed in such a room would be good enough to sit in. Handcuffs displayed in such a room would be confining, and a bullet displayed in such a room would be fatal. With appropriate programming such a display could literally be the Wonderland into which Alice walked.

In reality one makes use of five senses (visual, aural, haptic, olfactory and saporific) to acknowledge and get immersed in the world around us; instead, simulations of the visual and aural perceptions have been mainly developed to provide one with a feeling of presence and immersion when creating virtual realities.

1.1 Motivation and Goals

The rendering of virtual environments can be approached from two perspectives:

1. The setup is addressed to a single subject, or to a very small number of subjects that can be physically isolated from each other. In this case, the visual representation can be done via stereoscopic shutter glasses, while aurally, binaural sound reproduction may be achieved using headphones or cross-talk cancellation loudspeakers, to give only a few examples. By using a computer mouse or a joystick, the subject could even navigate through the imaginary space.

2. The setup is addressed to a large group of subjects who can physically interact among themselves and feel they participate to a common event. To achieve this, one needs methods of visual and aural representation that are not individual for each subject in the group (like a pair of headphones or a head-mounted display). Instead, a single virtual source must render the aural/visual information to as many people as possible, independent of their location in the room in relation to the source.

From an aural perspective, an ideal representation of the second hypothesis is still to be found. The present research investigates to what extent the Ambisonics system can be successfully implemented into this scenario. Both physical measurements as well as perceptual tests are performed in order to determine if an Ambisonics rendering system is an appropriate solution.

1.2 The Concept of Virtual Environments

In 1965 Ivan Sutherland, a recent PhD. graduate from Massachusetts Institute of Technology, raised for the first time the concept of virtual realities [2]. He expressed the belief that by making use of computer technologies we could create an ideal virtual world, an “ultimate display” that would “be a room within which the computer can control the existence of matter”.

Ever since, researchers and people involved in the computer sciences as well as in the media arts have been trying to develop methods and tools leading to artificially originated environments, to create simulated realities. This type of environment is produced using technological tools, mainly computer hardware and software, as well as peripheral devices such as audio or video monitors. It has the purpose of connecting people (subjects) with elements, objects, or other people that are not present in their physical vicinity, but that can thus be perceived as being there, through the use of special interfaces. As already mentioned, this type of environment is generally known by the terms *virtual reality* or *virtual environment*.

With a certain connotation, the idea of virtual reality has always existed. As a combination between a story or amount of information presented to us and the work of our imagination, it was always present through books and paintings, or

since the end of the 19th century through movies, photographs, or recorded music for instance.

1.3 A Historical Overview on Sound Reproduction

The aural aspect is a major part of a virtual environment and of recreating the presence and immersion sensations the subject would normally experience. Sound reproduction has come a long way since its beginnings, and nowadays, especially due to digital technology in terms of computing power and storage capabilities, one can develop extensively complex surround sound reproduction systems. To understand how Ambisonics appeared as one solution to surround sound reproduction, we present here a brief overview of the history of sound reproduction and its impact on listeners' expectations over time.

In 1895 in Paris, *The Arrival of a Train at the Station* was the first movie ever to be shown to an audience. The performance frightened the audience through the realism induced by its images and sound effects [3]. Nowadays throughout the world, a large amount of research is involved in developing new visual and aural rendering systems that are able to provide a sense of realism to a reproduced/virtual environment, so that the audience can fully experience the illusion of being present in the space where the *action* is taking place. What happened in the last over 120 years, that in spite of the improved technology the sense of realism and presence that is sought for in virtual environments seems to be less credible? As Jerome Daniel says in [4], the human perception has the capacity of not letting itself fooled for a too long time:

By multiple experiences, by comparison and once overcoming the surprise effect (like the audience of the first movie watching a train entering a train station), the person learns how to distinguish these illusions of real or simulated events, and to identify their procedures.

This could mean that no matter how much we improve rendering systems (of whatever nature: visual, aural etc.), humans, due to their learning capabilities, will always be able to differentiate between reality and simulation. This could be part

of our self defense system and we probably cannot and should not fight against it. However, even if aware of not being part of the immediate reality, reproducing or simulating a space as accurately as possible is a goal for both communication and entertainment purposes.

The story of recorded sound started in 1877 when Edison invented the first phonograph (i.e. the tinfoil cylinder phonograph) [5]. It was followed by the gramophone and telegraphone (i.e. magnetic recorder using steel wire). All these sound systems reproduced the sound through a horn type loudspeaker; only in 1911 the moving-coil transducer concept was introduced through the production of a moving coil loudspeaker called the “Magnavox”. Up to this point all sound recording was done on one single track and reproduced in mono. Once the magnetophone (i.e. magnetic tape recorder) was invented in 1928, the next big step consisted in its capability of recording and reproducing synchronized discrete sound tracks. Due to this new technology, in 1949 Magnecord added a second head to its PT-6 tape recorder, creating one of the first open reel stereo tape recorders.

In parallel with the sound reproducing devices described above, new recording techniques were developed. In the early 1930s British scientist Alan Blumlein was experimenting with stereo spaced microphone techniques, developing the coincident microphone recording. Named Blumlein after its inventor, the technique was based on a crossed bidirectional microphone configuration that was able to recreate a “stable and articulate stereo image” [6]. The technique is based on differences of intensity between the two microphones. At the same time, Harvey Fletcher at Bell Laboratories in the US designed a different stereo recording system, which was using spaced microphones. The stereo miking techniques along with the possibility of recording the signals on independent tape channels opened a new way of sound reproduction. This started to be commercially used only in the early 1950s. The sound, recorded stereophonically, was encoded and transmitted to a multi-loudspeaker system, where it recreated a new aspect of sound: its spatiality.

The overall goal in constantly changing and improving audio reproduction systems is surrounding the spectator with sound in such a way as to improve the sentiment of realism in the virtual environment he/she is immersing into, either in

the form of a television show, movie, records listening, teleconferencing or other types of computer interactive scenarios. One of the most important factors the human hearing mechanism is responsible for, is the aural perception leading to the spatial localization of objects or events that are either in contact or in close vicinity, “surrounding” a person. When trying to accurately reproduce a virtual space, one must find a way of recreating its sound field by encoding directional information of the sound sources, along with the spectral, temporal and amplitude information. The goal is to capture their directivity and virtual location in relation to the listening position. As sound reproduction via a mono system did not render any spatial information, the need for multichannel reproduction systems naturally earned its place in the entertainment and music industry market. And even if television continued, probably for economical reasons, to transmit in mono until the early 1980s, music records and cinema embraced the idea of multichannel sound rendering as soon as it was developed.

When dealing specifically with sound reproduction over a multichannel system, besides localization but related to it (as spatiality), many other issues can be raised. Michael Gerzon addressed some of them in [7]:

Is the localization sharp or diffuse? Is the image single or double? Is it in the head or elevated? Is the bass quality clean or lumpy? Is the treble quality clear or harsh? When two sounds in different directions occur together, are they both well located in their respective directions? Is there any sensation of “pumping”? Is the ambience uniform around the listener, or is there a “tunnel” in one particular direction? Is there any front/ back ambiguity? When listening to music, does listener fatigue set in, or does the sound have an unobtrusive quality that makes one forget the technical means of reproduction?

The ideal surround sound reproduction system should be able to successfully answer all these questions, and only afterwards one can start thinking what could confer a sentiment of presence in that environment, and from a semantic, philosophical or physiological point of view, what can be done to delude the participant into thinking he or she is in the house of “imagination”, in Alice’s Wonderland.

1.4 The Two-Channel Stereophonic System: First Spatial Sound Reproduction Technique

In audio-only applications (by this we refer mainly to record production and radio), the main form of sound encoding and decoding is the two-channel stereo system. The term *stereo* - an abbreviation for *stereophonic* - has its origins from Greek, and it can be translated as “solid sound”. Hence, a stereophonic system should theoretically be a system that makes use of a multiple number of loudspeakers in order to reproduce a solid, natural sound image. In spite of this, there are only a few types of what we call stereophonic formats using more than two loudspeakers (such as Dolby Stereo for instance); when one generally refers to the term “stereo”, unless otherwise specified, one understands a two-loudspeaker sound reproduction system.

Jens Blauert reviews in [1] theories and research results on sound localization developed up to this time. Here it is showed that if one and only one sound event is produced by two sources, the resulting auditory event appears at a position that depends on the positions of the two sound sources and the signals radiated by them. This situation occurs when the level and times of arrival of two identical signals produced by two distinct sound sources (loudspeakers) differ by very little, but enough to differentiate position. Quoting from [1]:

In establishing the position of the auditory event, the auditory system interprets the resulting two ear input signals approximately as if they arose at a single “phantom” sound source.

If a delay or weakening of one of the two signals takes place, the auditory event shifts towards the source radiating the earlier or stronger signal. This phenomenon is also called *summing localization*. If the signals arrive with a difference of more than one millisecond at the listener’s ears, the signal arriving first is taken into consideration, while the other signal is completely ignored in the interpretation process, and the auditory event appears as coming from the loudspeaker generating the respective first signal. This is called the *law of the first wavefront*. The *summing localization* and the *law of the first wavefront* form the phenomenon known as the *precedence*

effect, which is at the basis of the stereo rendering system.

Let us consider a traditional stereo set up: one has two loudspeakers placed in front of a listener in such a way that the three (two loudspeakers plus listener) form an equilateral or isosceles triangle. The two loudspeakers are fed the same input signal. However, if the signal directed to one of the loudspeakers is either delayed or weakened in amplitude, the auditory event will shift towards the other loudspeaker. By choosing the amount of delay and/or amplitude difference between the signals, one can determine the azimuth at which the listener perceives the auditory event (between the two loudspeakers). A zero azimuth translates into an auditory event positioned in the median plane, and is achieved when there is no delay or amplitude difference between the two signals.

There are two main drawbacks of the stereo technique: first, the phantom image (simulated location of the auditory event) is perceived as intended only if the listener is seated in a particular spot in relation to the loudspeakers, as described above. If listening from a different position, the phantom image shifts and the perceived direction of the auditory event is distorted. Because of this, stereo rendering is not a viable solution for cinema productions, where the audience is spread over a very large area, and where it is important to match the visual position of the sources with the sound without any directional shifting. With this in mind, Dolby laboratories developed a multitrack sound system specifically designed for movie theaters: the system encodes spatial (surround) information into the left and right channels through a difference signal, while encoding at the same time a center channel across the two [8]. The left and right speakers are used for music and special effects, while the center speaker proves to be essential for the dialog reproduction. In this way, no matter where the viewer is located in the theater (either of larger or smaller dimensions), the dialog comes always from the center (the center source being actually located behind the screen), eliminating thus any possibility of confusion between the visual and aural locations (which would have happened if the reproduction was a stereo phantom image). On the other hand, the sound image is exclusively frontal, not being able to render directional information of the rear events, and to immerse the listener into a 360° virtual space. For this reason, interest increased for sound

reproduction techniques that would involve more than two channels, providing thus more enveloping surround sound. The following section will very briefly overview the transition to multichannel reproduction systems.

1.5 The Beginnings of Surround Sound

The interest in surround sound was steered as early as the beginning of the 1950s. At the time, new reproduction systems such as *mono plus room* (M+R) or the *mid/side* method (M/S) were experimented with, being considered more aesthetically pleasing solutions than the regular two-channel stereophony [9]. The M+R system essentially replaced the left and right channels with a mono channel for the direct sound and a separate channel for “room reverberation”. On the other hand, the M/S is a method of recording of a sound field: while the mid channel consists of the signal recorded by a cardioid microphone capturing the direct sound, the side signal is recorded by a bi-directional microphone placed closely to the cardioid microphone, but at a 90° angle, so that its null is pointing towards the direct sound. By using a sum and difference matrix, these two signals are mixed and a variable amount of room signal (from the side channel) can be added into the mix.

Regardless, the big step forward is taken only when sound reproduction systems consisting of more than two channels start being considered. Among the first attempts of such systems we can mention the *stereo-ambiophony* developed in the early 1960s and based on a 4-2-4 matrix and consisting of two channels of direct sound and two channels of reverberation, followed in the 1970s by the quadraphonic and holophonic systems [9]. These early techniques are followed by new ones: the *binaural* and *transaural talk cancellation* systems, which in spite of using only two channels, simulate a 3-D spatial sound environment due to the head related transfer functions filtering; *Ambisonics*, which as we will immediately see, is inspired by the original quadraphonic system; *wave field synthesis*, developed by Berkhout in the late 1980s; *5.1, 7.1 to n-channel surround sound systems*, which are currently commercially utilized in the music as well as the film industry.

We particularly want to mention here the quadraphonic system because from

it, the ambisonics matrix was later developed. This technique, introduced in the early 1970s, consisted of placing four loudspeakers in the four corners of the listening space and feeding each of them its own discrete channel. The system proved to be unsuccessful because of its lack of compatibility with mono or stereo systems, strict set-up requirements and poor longevity of the recording media. Later on, the quadraphonic technique approached different types of encoding in matrix forms instead of discrete signals, allowing it to be compatible with other rendering systems. The Ambisonics system developed by initially by Michael Gerzon was one of these modified quadraphonic techniques, based on matrix encoding. This will be discussed in further detail in Chapter 3.

CHAPTER 2

SOUND LOCALIZATION: CONCEPTUAL OVERVIEW AND PRIOR WORK

2.1 Localization and Localization Blur

In acoustics language, the term *localization* can be defined as the process that seeks to identify an auditory event experienced by a subject with the location of the sound event that is causing it. The amount of spatial error between the location of the sound source and the location where the correspondent auditory event is actually perceived is defined as the *localization blur*. The localization blur depends on the human hearing mechanism, the architectural conditions of the space where the sound event is produced, and the type of sound stimulus. Added to all the elements described above, in the case of virtual sound sources, the localization blur is also a function of the audio reproduction system's characteristics.

2.2 A History of Sound Localization Research

Starting with the end of the 19th century, sound localization started to be a topic of interest in acoustics research. John Strutt (Lord Rayleigh) was the first scientist to study in detail the way human beings are able to localize auditory events around them and how the hearing mechanism influences localization. After doing localization experiments with humans, Lord Rayleigh concluded that sound localization is a function of the combined use of both ears [10]. Even if he admitted that the calculations and observations of his tests were incomplete, his goal was to “clear the ground” and to induce the interest of researchers to pursue the subject of sound localization. His first experiments on localization topics tried to find out “at what degree of accuracy the direction of a sound could be determined”. He realized that in order to determine this, no other material for the subject judgment “should be contemplated”.

In their most simple, incipient format, localization experiments were organized

as follows: the test took place in the middle of a lawn (free field) during quiet evenings; five or six people (live sources) placed around the subject and continuously shifting their positions were uttering words or sentences while the subject, keeping his/ her eyes closed, had the task to point with the hand towards the direction he/ she thought the sound was coming from. The results determined that the human voice (sentence, word or just a vowel) could be followed with precision and localized within a few degrees. With other sounds, the result raised the issue of localization confusions between the front and the back. They found out that the less complex the sound, the bigger the front to back confusion in localizing it [10]. Even if very empirically constructed, at the end of this particular research, the experimenters came up with a very preliminary idea of interaural level difference, and concluded that this is a function of the dimensions of the wavelength compared to the circumference of the head. The longer the wavelength compared to the head's dimensions the smaller the level difference of the sound arriving at the two ears was. Thus, it was expected that acute sounds would be localized with more precision because the difference of intensity at the two ears would help discriminating the location of the sound. Rayleigh determined that the human hearing system appeared to use different mechanisms to localize sound at frequencies below 700 Hz.

A number of other scientists such as Stevens and Newmann, or Roffler and Buttler took Lord Rayleighs work further, continuing his localization experiments. Blauert summed up a large number of research experiments and results on sound localization and spatial hearing in [1].

2.3 The Concepts of Sound Event vs. Auditory Event

According to [1], the concepts of sound event and auditory event have different meanings that cannot be interchanged. In these two terms lie the foundation and the reason behind the Ambisonics concept as well as all the other multichannel sound reproduction configurations. The term *sound event* is used to describe the physical aspect of the acoustic phenomena. By this, one refers to the mechanical vibrations and waves of an elastic medium, particularly in the frequency range of the human hearing (20 Hz – 20 kHz). On the other hand, an *auditory event* consists

of what is aurally perceived by a human being. This is usually determined by sound events, but it can be distorted by hearing affections (e.g. hearing loss), or other internal disease conditions (e.g. tinnitus), where an auditory event does not correspond to any external sound event but it is produced when an acoustic nerve is internally stimulated. Nevertheless, in normal conditions an *auditory event* is always caused by a *sound event*, and it is spatially and temporally correlated to it. Another element that stays between the *sound event* produced in a space and its corresponding *auditory event* (affecting/ influencing the latest one) is the space in which the sound is produced. Between the source and the ears, the sound wave is be changed/filtered by several conditions, like the spatial conditions of the room, the location of the receiver in that room in relation to the sound source, and the physical characteristics of the receiver's body (torso, head, and pinnae). In this context we introduce a particular type of such a filter: head-related transfer functions.

Also abbreviated as HRTFs, the head-related transfer functions are important in the process of localization of both real and virtual sources. They constitute the filtering process of incoming waves scattered by the head, torso and external ears (pinnae). The external ears along with the head and torso superimpose linear distortions on the incoming signals, distortions that depend on the direction of incidence of the sound wave and the source distance [1]. The term will be often encountered later in the text due to its role in the process of localization of both real and virtual sources.

2.4 Localization in the Horizontal Plane

Minimum localization blur occurs in the forward direction [1]. The more the sound source shifts towards the sides of the listener, the more the localization blur increases, reaching a peak at 90° from the frontal direction. If the sound source continues moving towards the back of the listener, the localization blur decreases again. However, at 180° the blur reaches values that double the amount at 0° , but is smaller than the blur specific to the 90° direction.

More detailed information on localization in the horizontal plane as a function of the human auditory system is described in Section 2.6, as part of the baseline

Type of Signal	Localization Blur
Continuous Speech by an Unfamiliar Person	17°
Continuous Speech by a Familiar Person	9°
White Noise	4°

Table 2.1: Localization in the median plane as a function of signal content [1]

listening tests for directional hearing in the horizontal plane (also see [11]). The perceptual localization experiments of an Ambisonics rendering system presented in Chapter 4 are in part developed taking into consideration these results.

2.5 Localization in the Median Plane

In the median plane, all auditory events consist of equal interaural time and level differences, making impossible to distinguish between different sound locations using only these cues. Therefore, localization in the median plane depends on different properties of the stimulus, such as frequency content, or familiarity of the subject with the sound source. Blauert, summarizing in [1] several auditory localization tests by Damaske & Wagener, and Roser, categorizes the localization blur in the median plane depending on the type of stimulus, as seen in Table 2.1.

Other elements that affect localization in the median plane are the length of the signal and very narrow frequency band content. Thus for very brief signals with impulsive content, the auditory event shifts to the rear of the subject, while for sounds having a frequency content narrower than two thirds of an octave band, localization cannot be determined by the subject.

2.6 Baseline localization test: Investigation of the Auditory System Localization Blur in the Horizontal Plane

2.6.1 Experimental Design

The baseline localization test is designed to determine the auditory system's ability to localize sound events in the horizontal plane for an azimuthal range of approximately 180°. This experiment, reviewed in the following sections was realized

as part of [11].

2.6.1.1 Test environment

The tests were led in a hemi-anechoic environment using discrete sound sources for localization in the horizontal plane. The sources were concealed by black acoustically transparent curtain, in order to avoid any visual triggered bias.

2.6.1.2 Test Sound Sources

Set in a rectangular shaped hemi-anechoic chamber, the discrete sources consisted of 22 individual loudspeakers (see Figure 2.1). In a 360° clockwise direction with 0° located where the subject's head was pointing, the twenty-two loudspeakers were positioned from the left-most position at 272° to the right-most position of 92.86°. The loudspeakers were placed along the walls of the chamber at various distances from the subject. Using equalization filters and delays this distance asymmetry was compensated for, and virtually the sources were placed equidistantly from the subject in a circle. The loudspeakers were concealed by acoustically transparent black cloth to avoid visual bias in the localization.

2.6.1.3 Subject Population

The subjects were seated during the test and not allowed to move their heads. In order to keep their head position consistently facing towards an azimuth of 0°, the subjects wore on their foreheads a headband with a laser pointer. Before listening to each new stimulus, they had to make sure the laser was pointing on a blue strip taped to the black cloth at 0° azimuth. The sources were placed at the same height as the subjects' ears.

2.6.1.4 Test Stimuli

The baseline localization test had two main goals:

1. To test for human auditory localization resolution in the horizontal plane;
2. To test if there is a difference in the localization blur as a function of the type of aural information presented to the listener. For this reason, the experimenter

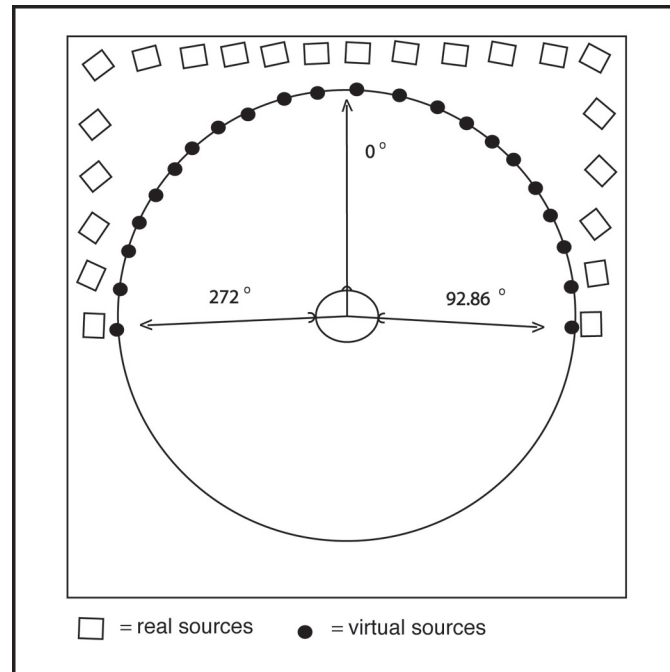


Figure 2.1: Baseline localization test set-up

chose a variety of stimuli types (e.g. task sentences, non-sense sentences, inverted speech, pink noise, diagnostics rhyme test words).

2.6.1.5 Test Procedure and Data Acquisition

The stimuli were played randomly using Matlab code. Each stimulus was played twenty times through each loudspeaker over the whole subject population. Each person heard each loudspeaker eight times, every time for a different stimulus. The playing order was random. The subjects used a laser pen to point on a horizontal scale at the position where they localize the sound source. A measuring tape was placed along the curtain, and the experimenter (in the chamber with the subject) would record in inches the position on the measuring tape where the subject pointed as the location of the sound source. These values were later translated in azimuth degrees.

CHAPTER 3

AMBISONICS: THEORY

3.1 Overview

The goal of a surround sound system is either to recreate the sound field of a real space with all its localization information, or to simulate the sound field of a virtual space. Theory shows that in order to exactly recreate a sound field over a two-meter diameter listening area for frequencies up to 20 kHz, around 400,000 speakers are needed [12]. As this is not achievable in practice, one has to find a way of encoding the sound field that one wants to reproduce with all the distance, position, and direction components in a limited number of available channels.

Ambisonics is a sound encoding system that consists of recording or simulating a *sound event* produced in a physical or modeled space in such a way that when reproduced, the decoded information transmitted through a specially configured array of loudspeakers can recreate at the listener's position the *auditory event* that would have been experienced in the original space. Let us consider the ordinary sound signal path as consisting of a sound source, a physical space in which the sound is radiated, the receiver's filtering characteristics (which can be represented by HRTFs) and mental processing of the event – through the auditory nerve and the brain. An Ambisonic system aims to encode the sound event at the receiver area to later reproduce it as if the resulting auditory event was experienced in its original context. In order to achieve this, the system includes directivity and other spatial information, which are derived from psychoacoustic theories on hearing and localization mechanisms. Ambisonics essentially attempts to recreate the first two steps of the signal path described above, with an emphasis on the psycho-acoustical elements that are known to be used by the human hearing mechanism. The HRTF filtering is inherently processed by the listener at the time of the decoding and reproduction of the ambisonically recorded event.

Besides looking to achieve correct localization of the sound sources, Ambisonics also aims to provide a feeling of enveloping, surrounding information of the room

(if one considers it as being a larger instrument for the sound), which with its own characteristics can create a particular sonority, a specific timbre and coloration.

3.2 The Psychoacoustics Behind the Ambisonics Concept

As described in detail in [13], the Ambisonics technique originates from theories stating that the hearing mechanism has three different methods to localize sound information as function of the frequency range: below 700 Hz, between 700 Hz and 5 kHz, and above 5 kHz.

Due to longer wavelengths, localization at low frequencies (below 700 Hz) is based primarily on phase differences, as the head is too small¹ to present an obstacle, and thus the amplitude reaching the two ears is essentially the same. Considering two theories on how localization is perceived at low frequencies, Gerzon introduced a new theory that would embed both [13]: One school of thought, represented by Clark, Dutton & Vanderlyn, and Bauer, based its theory on the idea that the listener would not move his/her head in order to perceive the phase differences at the two ears, which helped in localizing the sound. The second theory sustained by Makita, Leaky and Tager claims that to localize low frequencies “the brain uses additional information from variations at the two ears caused by rotations of the head within the sound field” (also referred to by Gerzon as the “Makita theory”). Gerzon combined the two, proposing a new solution in [13]: by recording the sound field with an omni-directional microphone at the position where the listener is supposed to be located, he could encode the sum of the waveforms reaching the two ears if we ignored the presence of the head (at very low frequencies, under 700 Hz). The remaining directional information at low frequencies (based on phase differences) is actually the velocity of the sound field along the ear-axis, which can be picked up by a sideways-pointing figure of eight microphone. This set-up would correspond to the first theory where the information is obtained without any head movements from the listener. However, as the head in practice will rotate, the information obtained this way from the figure of eight microphone would be used to sustain the “moving

¹The size of the head between the two ears equals on average about half a wavelength at 700 Hz.

head” theory. In both situations it is assumed that the velocity microphone information that is 90° out of phase with the omni-directional microphone information is used in deducing the direction of sound.

Therefore, the direction of a sound whose frequency is lower than 700 Hz can be determined by a vector r_v equal to the ratio between the overall acoustical vector velocity gain and the acoustical pressure gain of a reproduced sound at the listener’s position [14]. Let us consider a circular sound reproduction array of n number of loudspeakers. Assuming the center of the circle as the listening position, each loudspeaker l will have an azimuth θ_l , and a gain g_l . The total acoustical pressure gain reaching the listener will be:

$$P = \sum_{l=1}^n g_l \quad (3.1)$$

The velocity gain, according to [14], is the vector sum of the n vectors with their respective lengths g_l and pointing towards azimuth θ_l , with x- and y- components:

$$v_x = \sum_{l=1}^n g_l \cos \theta_l \quad (3.2)$$

$$v_y = \sum_{l=1}^n g_l \sin \theta_l \quad (3.3)$$

Thus, the velocity localization vector r_v pointing in the direction azimuth θ_v will be:

$$r_v \cos \theta_v = v_x / P \quad (3.4)$$

$$r_v \sin \theta_v = v_y / P \quad (3.5)$$

r_v is the velocity localization vector magnitude and ideally should equal unity (only for singular sound sources), while θ_v is the velocity vector localization azimuth or “the apparent direction of sound at low frequencies if one turns ones head to face the apparent direction” [14].

For mid-frequencies, the wavelengths are shorter and thus the phase differences

are no longer of singular importance. Instead, it is “the directional behavior of the energy field around the listener” that mainly contributes to the localization [13]. In this sense, localization will be determined by an energy localization vector, which is obtained similarly to the velocity localization vector, with the difference that g_l is replaced by g_l^2 . The vector indicating the location of sound above 700 Hz and up to around 5 kHz equals the ratio between the vector sound-intensity gain and the acoustical energy gain of the reproduced signal at the listener’s position [14].

In this case Gerzon determines the total energy gain at the listener’s position (in the center of the system) to be:

$$E = \sum_{l=1}^n g_l^2 \quad (3.6)$$

The sound-intensity gain, according to [14], is the vector sum of the n vectors with their respective lengths g_l^2 and pointing towards azimuth θ_l , with x- and y-components:

$$E_x = \sum_{l=1}^n g_l^2 \cos \theta_l \quad (3.7)$$

$$E_y = \sum_{l=1}^n g_l^2 \sin \theta_l \quad (3.8)$$

Thus, the energy localization vector r_E pointing in the direction azimuth θ_E will be given by the equations:

$$r_E \cos \theta_E = E_x/E \quad (3.9)$$

$$r_E \sin \theta_E = E_y/E \quad (3.10)$$

r_E will rarely equal unity², but one should aim to maximize its value as close to unity as possible. It is important to mention though that there is a transition range between 250 Hz and 1.5 kHz where both methods apply. The decoders have to be designed such that in this range they are able to cover both pressure/velocity

² r_E can equal unity only if the sound comes from only one loudspeaker

information as well as energy vector information. This is realized by designing shelf filters, which give optimum velocity magnitude at low frequencies while giving optimum energy vector magnitude at high frequencies, such that the Makita and energy vectors are identical at all frequencies [15].

Above 5 kHz, multiple experiments starting with those undertaken by Lord Rayleigh show that one's ability to recognize sound direction at high frequencies is due to our pinnae and the way the short wavelength of high frequencies scatter them (HRTFs). Thus, the high-frequency information will be deduced by the listener at a decoding level using his/her own HRTFs.

3.3 The Encoding Process

Taking into account the psychoacoustics methods and localization theories described above, the encoding of a sound field is done by recording simultaneously the pressure of the sound at a particular (virtually the listener's) location using an omni-directional microphone (the signal noted as W), while recording the velocity components (or gradient pressure) with bidirectional microphones placed in exactly the same location as the omni-directional one. Depending on the number of velocity components measured/recorded, there are several Ambisonic encoding methods.

Ambisonics encoding can be regarded as the decomposition of the sound field into spherical harmonics centered at the listener's position [6]. The encoding can be designed to use first- or higher-order harmonics. The amount of directivity information encoded is directly proportional to the degree of the harmonics order.

The original method, based only on first-order harmonics, encodes the directivity of the sound field through two or three velocity components along with the omni-directional one. The signals encoded with first-order Ambisonics are also known as B-format: W - pressure, and X , Y , Z - pressure gradient. The Z component is necessary only if encoding height information; for an exclusively horizontal representation of the sound field, Z is not necessary. As a rule in Ambisonics the X coordinate is pointing forward, while the Y coordinate is pointing to the left, as seen in Figure 3.1.

Later at the reproduction status, through an optimized decoding technique

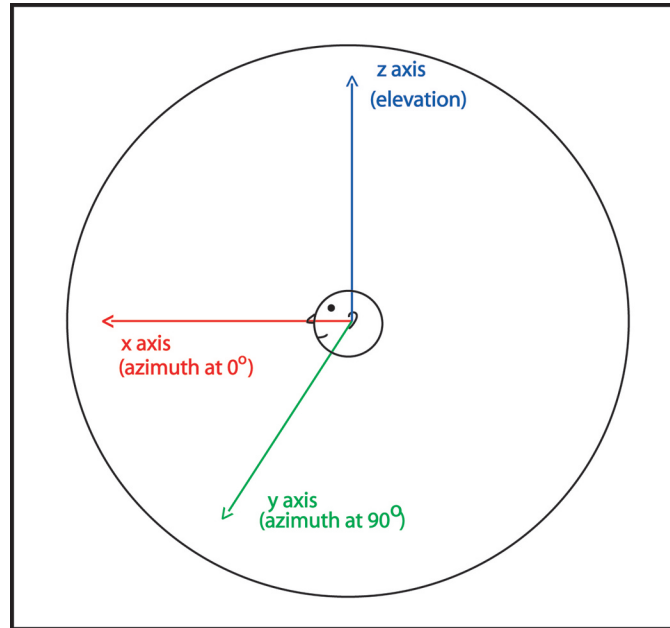


Figure 3.1: Ambisonics Coordinate System

one can obtain a coherent and homogenous image of the sonic space, either as a two-dimensional or three-dimensional representation. The encoding technique is independent of the number of loudspeakers and their set-up, being configurable in variable layouts. Regardless, for good results the number of loudspeakers used should be larger than the number of encoded signals (W, X, Y etc.) If the encoding covers only the 360° of the horizontal field, the system is called *panthophonic*; if it covers a whole sphere, the encoding will be three-dimensional and the system is called *periphonic*.

Bamford and Vanderkooy derived in [8] the first-order pantophonic encoding equations. This is the simplest way to exemplify an Ambisonics encoding method. Let us consider a plane wave :

$$S_\psi = P_\psi e^{ikr \cos(\theta - \psi)} \quad (3.11)$$

where ψ is the incidence angle of the plane wave with respect to the x-axis, P_ψ is the peak amplitude of the wave, k is the wave number ($2\pi f/c$), and r is the radial distance from the listener position to the source at an angle θ . When rewriting the

plane wave equation in terms of spherical harmonics we get:

$$S_\psi = P_\psi J_0(kr) + P_\psi \left(\sum_{m=1}^{\infty} 2i^m J_m(kr) [\cos(m\psi) \cos(m\theta) + \sin(m\psi) \sin(m\theta)] \right) \quad (3.12)$$

If we assume that the signal coming from each speaker in the Ambisonics decoding configuration is a plane wave arriving at the listener, the resulting signal at this particular position is the sum of each of the signals coming from each speaker n positioned at an angle θ , and where P_n is the amplitude coming from each speaker, located at an angle ϕ_n . Thus, the signal coming from each individual loudspeaker is:

$$S_n = P_n J_0(kr) + P_n \left(\sum_{m=1}^{\infty} 2i^m J_m(kr) [\cos(m\phi_n) \cos(m\theta) + \sin(m\phi_n) \sin(m\theta)] \right) \quad (3.13)$$

The sum of all loudspeakers is:

$$S_{total} = \sum_{n=1}^N P_n J_0(kr) + \sum_{m=1}^{\infty} 2i^m J_m(kr) \left(\sum_{n=1}^N P_n \cos(m\phi_n) \cos(m\theta) + \sum_{n=1}^N P_n \sin(m\phi_n) \sin(m\theta) \right) \quad (3.14)$$

The coefficient m represents the order of the Ambisonic reproduction. For a first-order Ambisonics system m equals 1. For second-order Ambisonics m equals 1 and 2, and so on. Hence, the Ambisonics first-order pantophonic components in the horizontal plane are:

$$W = P_\psi = \sum_{n=1}^N P_n \quad (3.15)$$

$$X = \sqrt{2} P_\psi \cos \psi = \sqrt{2} \sum_{n=1}^N P_n \cos(1\phi_n) \quad (3.16)$$

$$Y = \sqrt{2} P_\psi \sin \psi = \sqrt{2} \sum_{n=1}^N P_n \sin(1\phi_n) \quad (3.17)$$

To extend the encoding to the second-order, two new components are generated:

$$U = \sqrt{2} P_\psi \cos(m\psi) = \sqrt{2} \sum_{n=1}^N P_n \cos(2\phi_n) \quad (3.18)$$

$$V = \sqrt{2}P_\psi \sin(m\psi) = \sqrt{2} \sum_{n=1}^N P_n \sin(2\phi_n) \quad (3.19)$$

A normalization factor $\sqrt{2}$ is introduced to make sure the channels (W, X, Y) have equivalent mean power.

Additionally, for first-order Ambisonics, the Z-component provides the information for a periphonic reproduction (the verticality of the sound field), but also ensures a homogenous representation of the sound events, without favoring any direction over another. For the second order with periphonic information, three more signals are added: R, S and T.

As we saw above, higher-order Ambisonics are built on the same principles as first-order. But if first-order Ambisonics have a very small effective listening area where the reproduced auditory event can be accurately perceived from a directional point of view, higher-order Ambisonics address this problem as well. As stated in [4], higher-order harmonics bring contributions in the expansion of the area of sound field reconstruction (the effective listening area), resulting in a larger “sweet-spot” (very important for contexts where multiple listeners participate to the event), or allowable head movement. The research presented in the following chapters investigates this issue through localization tests using both first- and second-order Ambisonics.

CHAPTER 4

PERCEPTUAL INVESTIGATION OF THE AMBISONICS RENDERING SYSTEM

This chapter describes a set of tests designed to investigate the rendering characteristics of a periphonic Ambisonics system. The tests involve the participation of human subjects. Below there are described the goals, methods, and results developed in this research.

4.1 Goals for the Perceptual Experiments

The experiment developed for investigating the perceptual attributes of an Ambisonics rendering system consists of localization tests distributed to a group of fourteen subjects. The main goal in designing this experiment is to determine the localization accuracy specific to an Ambisonics system. More in detail, we want to compare and determine the localization accuracy in particular circumstances such as:

1. First- vs. Second-Order Ambisonics;
2. Anechoic vs. Reverberant environment;
3. Localization in the horizontal plane and localization in the vertical plane.

4.2 Experimental Design

This section describes the procedures leading to the design and distribution of the experiment.

4.2.1 Test Environment

The experiment is led in a hemi-anechoic environment, using an Ambisonics audio rendering system and a large screen projection. The Ambisonics rendering system is configured as a periphonic set-up, and consists of twelve loudspeakers

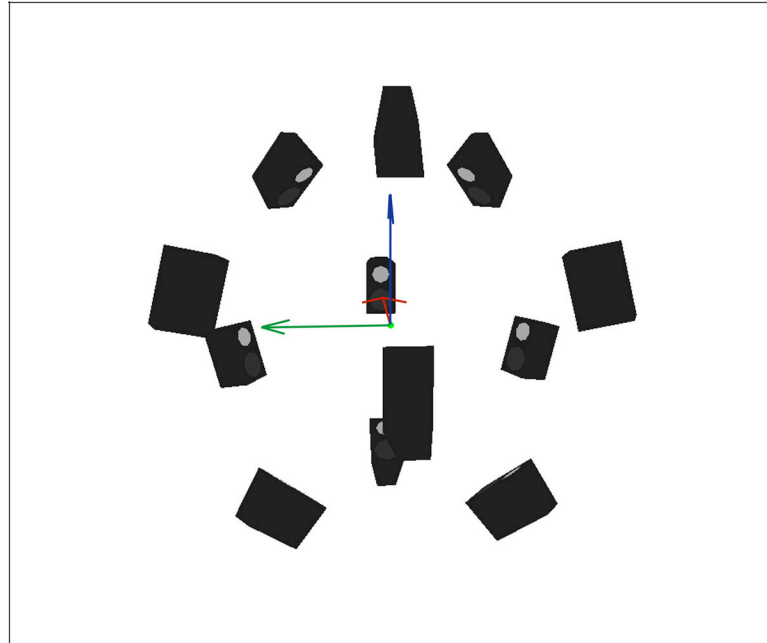


Figure 4.1: Ideal dodecahedron configuration

spherically distributed (see Figure 4.1). The listening position is at the origin of the Ambisonics coordinate system, which is virtually the center of the sphere as well. This is, however, not achievable in practice in our rectangular listening space, and the loudspeakers could not be located equidistantly from the subject/origin. Figure 4.2 shows the warped version, used in our experiment. To compensate for the distance differences, equalization filters and delays are implemented in the decoding path to each loudspeaker (see Section 4.2.7).

The loudspeakers are only partially seen by the subjects: The ones hanging from the ceiling and the ones on the floor around the listener are visible, while the loudspeakers around the walls are masked by black, acoustically transparent curtains.

4.2.2 Test Virtual Sound Sources

The choice of the virtual sources location was determined by the following:

1. the loudspeaker configuration used in the decoding of the signals;
2. considerations on the characteristics of the human auditory system.

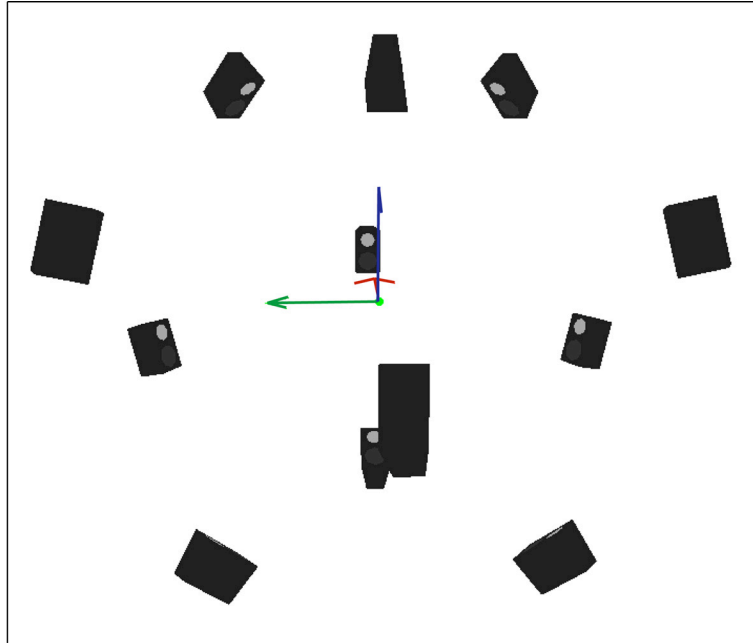


Figure 4.2: Warped dodecahedron configuration

From the first point of view, in a symmetrically distributed audio rendering system composed of an even number of loudspeakers, the Ambisonics decoding consists of assigning the same signal to two diametrically opposite loudspeakers, but in anti-phase. For this reason we choose to locate our virtual sources in only one hemisphere, while assuming that the localization blur specific to the system will be the similar in the second hemisphere as well.

On the other hand, regarding the characteristics of the human hearing system we see both in [1] and [11] that due to their bodies' symmetry, humans have a similar localization response towards their left and right sides, in the horizontal plane. For this reason, we decided to place our sources only towards the right of the listener, assuming that the responses at their left would have been similar. We also see in [11] (also see Section 2.6) that the most consistent responses and the smallest localization blur is encountered within an azimuth of $\pm 35^\circ$ in front of the listener. The virtual sources in our experiment are placed up to an azimuth of 40° to the right of the listener. Regarding the localization blur in the vertical plane, we positioned our virtual sources in the median plane within an elevation range of

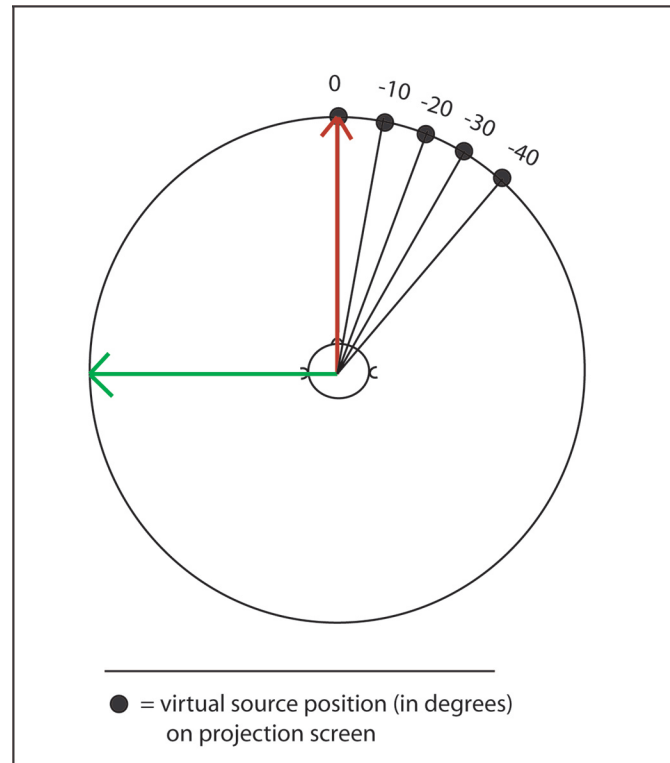


Figure 4.3: Virtual Sources Location in the Horizontal Plane

0° to 30° . The distance of all virtual sources to the listener is kept constant. The exact angular location of the virtual sources with respect to the listener can be seen in Figures 4.3 and 4.4.

The distribution of the sources for localization in the horizontal plane (elevation = 0°) in the Ambisonics coordinate system was:

1. azimuth = 0° ;
2. azimuth = -10° ;
3. azimuth = -20° ;
4. azimuth = -30° ;
5. azimuth = -40° .

For localization in the median plane (azimuth = 0°), the sources were distributed as follows:

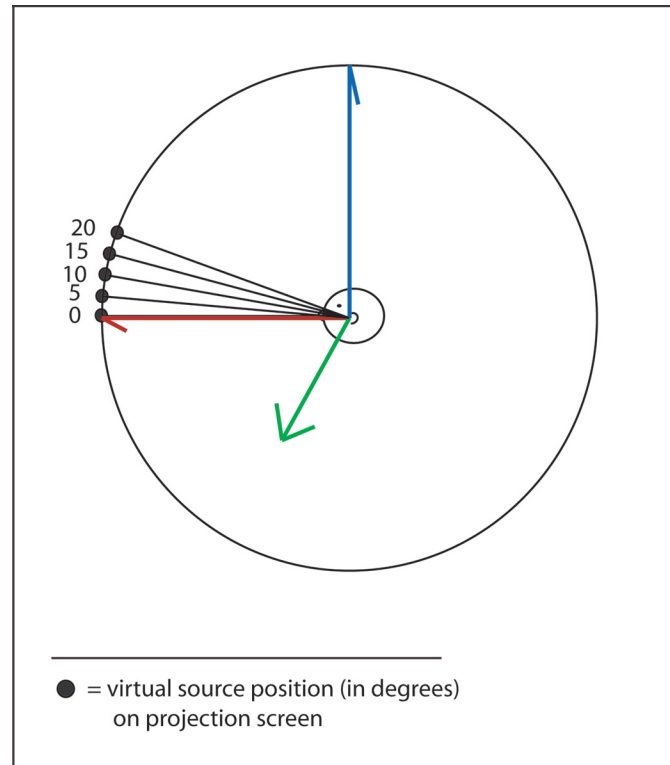


Figure 4.4: Virtual Sources Location in the Vertical Plane

1. elevation = 0° ;
2. elevation = 5° ;
3. elevation = 10° ;
4. elevation = 15° ;
5. elevation = 20° .

4.2.3 Test Stimuli

The source signal consists of two pink noise generated sound bursts (each of them of two seconds long) with two seconds of silence in-between. Each sequence can be of two types: anechoic or with reverberation. One of the goals of this research is to find out if there is a difference in the localization of the sources when the sound contains reverberation as opposed to anechoic sound.

The reverberant field simulations, we used was a generic reverberation tail of 1.85 seconds; the initial time delay gap between the direct sound and the reverberation was 1 ms long, while the direct-to-reverberant sound level ratio was 0dB.

The sound level for all stimuli conditions was adjusted so that its A-weighted value in the hemi-anechoic field at the listener's location, measured 60 dBA for all cases presented during the test.

4.2.4 Subject Population

The subject population was drawn from both experienced listeners consisting of the graduate students and faculty in the Acoustics program at RPI, as well as unexperienced listeners, all within an age range of 21 to 33 years old.

The listeners were positioned at the origin of the Ambisonics coordinate system. They were given written instructions explaining the motivation of the experiment as well as its procedures. They were informed that their task was to localize sound sources either in the horizontal or median plane. A computer graphic interface was projected on the wall in front of the subjects, as seen in Figure 4.5. In order to record their answers, they had to watch the visual projection and click with the mouse on the screen at the perceived location. For each stimulus, either a horizontal or vertical line would appear on the projection screen (see Figures 4.6 and 4.7). Depending on which of the two types of lines would appear, the subjects were instructed to identify if they were supposed to localize the sound in the horizontal or vertical plane, to point the mouse arrow along the shown line, and click on the perceived location. Subjects were also allowed to move their head as necessary to localize the sound; this test is not designed as a baseline localization test, but aims to determine the localization accuracy during normal listening conditions.

Before starting the test, all listeners had to take a training session in order to familiarize themselves with the experiment. The training session consisted of ten stimuli, randomly chosen from the stimuli used in the test, and played in the same conditions as during the actual experiment. Once the training session was completed, the subjects could start the actual session that was logged for later statistical evaluation. This session consisted of eighty samples and took between

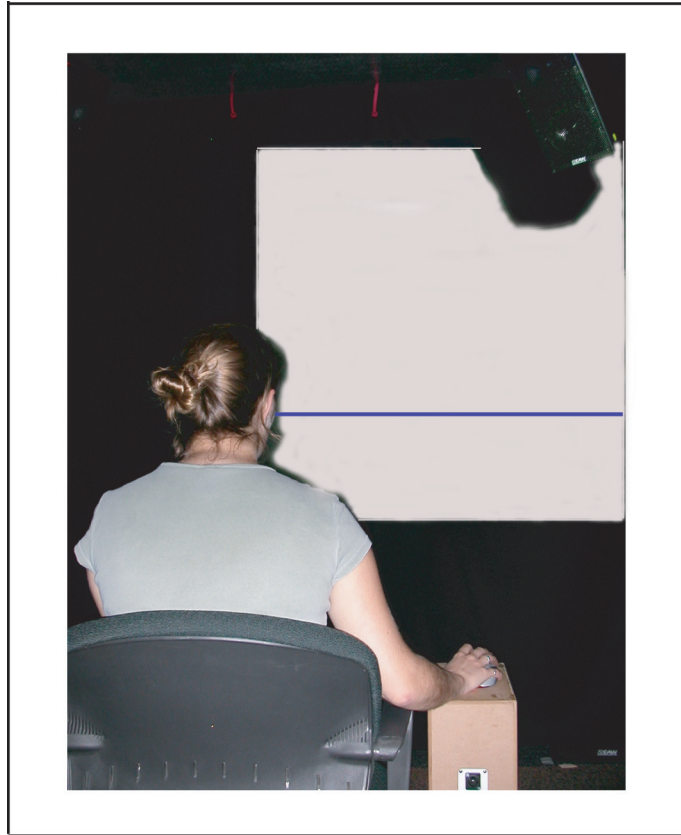


Figure 4.5: Photograph of the testing environment

fifteen and twenty minutes to be completed by each subject.

4.2.5 Test Cases and Task Distribution

The experiment was composed of 40 different cases, each repeated twice for a total of 80 stimuli played to each subject. With a total of fourteen subjects we have 1120 cases distributed over the whole subject population. A summary of the cases is showed in Table 4.1.

4.2.6 The User Interface

The experiment was run using Matlab code. The user interface consisted of a mouse and a projection screen displayed on the wall in front of the listener as seen in Figure 4.5. To start the test, the subject had to click the mouse anywhere on the screen. Next, a horizontal or vertical line appears on the screen (see Figures

Ambisonics Order	Simulation	Azimuth Resolution	Elevation Resolution
First Order	Anechoic	5 cases, one every 10 degrees	
			5 cases, one every 5 degrees
	With Reverberation	5 cases, one every 10 degrees	
			5 cases, one every 5 degrees
Second Order	Anechoic	5 cases, one every 10 degrees	
			5 cases, one every 5 degrees
	With Reverberation	5 cases, one every 10 degrees	
			5 cases, one every 5 degrees

Table 4.1: Localization Test Cases

4.6 and 4.7) and the first stimulus is played. According to the written instructions, the subject has to click along the line on the screen wherever he thinks the sound is coming from. Once he clicks the mouse at the chosen location, the answer is logged in a text file. Subsequently, a new stimulus is played and the process repeated.

4.2.7 Test's Design Procedure and Methodologies

This section overviews the methodology used to create the sound stimuli for the localization tests. The sound simulations and encoding were developed using Matlab code. For decoding and play-back the Lexicon's LARES Signal Processor was used.

All the sources are located on a virtual sphere having the same origin as the Ambisonics coordinate system, and whose radius equals the distance from the origin to the loudspeakers' virtual positions. Once we have obtained the Cartesian coordinates for each virtual source, we can determine the Ambisonics channels' encoding coefficients. The encoding coefficients for a sound source located at coordinates (x, y, z) are reviewed in Table 4.2 (also see [16]).

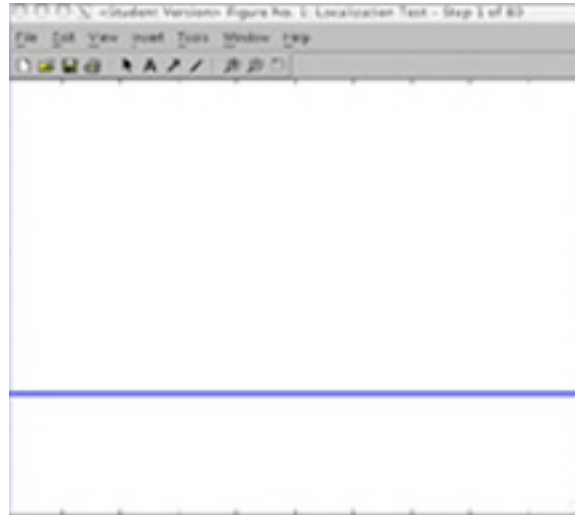


Figure 4.6: Projection Screen, for localization in the Horizontal Plane

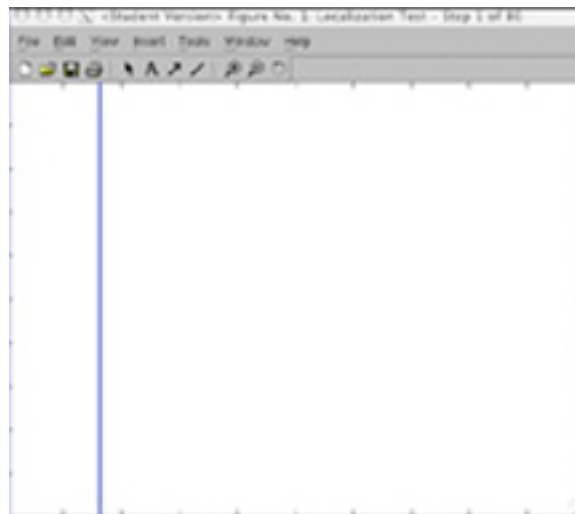


Figure 4.7: Projection Screen, for localization in the Median Plane

After the Matlab protocol is started and the subject clicks the mouse to start the test, a test case is picked at random by the program. According to the chosen case characteristics, the code will choose between an anechoic pink noise sound file or one with reverberation. Subsequently, depending on the chosen source location and Ambisonics order, the encoding is performed in real time and the resulting nine-channel matrix is sent at a sampling rate of 48 kHz and 16 bit quantization, via the MOTU Audio 24 I/O interface, to two LARES Signal Processor units. The

Encoding Channel	Cartesian Representation
W	0.707107
X	x
Y	y
Z	z
R	1.55zz-0.5
S	2zx
T	2yz
U	xx-yy
V	2xy

Table 4.2: Ambisonics Encoding Coefficients, up to the Second Order

Loudspeaker	x-coordinate	y-coordinate	z-coordinate
#1	1	0	0
#2	-1	0	0
#3	0.4472	0	-0.8944
#4	-0.4472	0	0.8944
#5	0.4472	0.8507	-0.2764
#6	-0.4472	-0.8507	0.2764
#7	0.4472	-0.8507	-0.2764
#8	-0.4472	0.8507	0.2764
#9	0.4472	0.5257	0.7236
#10	-0.4472	-0.5257	-0.7236
#11	0.4472	-0.5257	0.7236
#12	-0.4472	0.5257	-0.7236

Table 4.3: Cartesian Coordinates of the Loudspeakers in the Ambisonics Rendering System

rendering system chosen for this experiment consists of a tilted dodecahedron (twelve loudspeakers), distributed in its ideal configuration as shown in Table 4.3 and Figure 4.1.

Once inside the LARES Signal Processor, each of the four or nine encoded channels (depending on the Ambisonics order) is routed to an FIR filter. The use of the shelf filters is due to the fact that there is a transition range between 250 Hz and 1.5 kHz where both pressure/velocity information as well as energy vector information are needed for localization (see Section 3.2). By designing shelf filters,

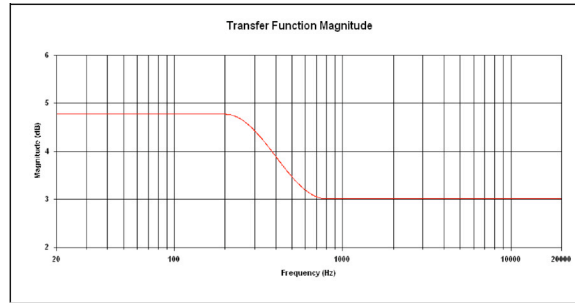


Figure 4.8: Shelf Filter applied to the velocity components

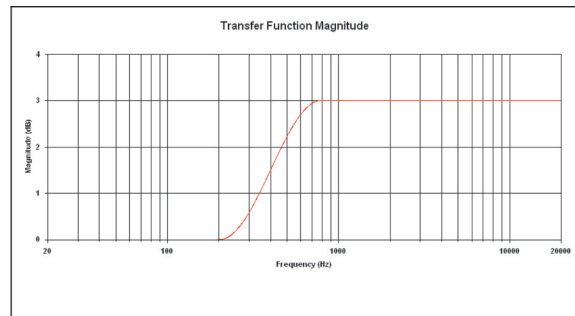


Figure 4.9: Shelf Filter applied to the energy components

we provide optimum velocity magnitude at low frequencies while giving optimum energy vector magnitude at high frequencies. The filters characteristics can be seen in Figures 4.8 and 4.9. Next, each filtered signal is sent to twelve individual virtual mixers, and after being decoded in real time, they will eventually constitute the source signal for each of the twelve loudspeakers. For decoding, each of the nine encoded channels is applied a certain gain and subsequently sum up to create each channel of a loudspeaker feed. Table 4.4 shows the gain coefficients applied to each signal in the path towards each of the twelve loudspeakers, as part of the Ambisonics decoding process.

However, as our hemi-anechoic chamber dimensions and shape did not permit the assemblage of the loudspeakers set-up in its ideal form (as it was shown in Table 4.3), we had to modify the location of the loudspeakers according to Table 4.5. In order to compensate for the resulting distance inequality between each loudspeaker and the origin, and to recreate thus the ideal spherical/dodecahedron configuration,

Lsp	W	X	Y	Z
#1	0.1179	0.2500	0.0000	0.0000
#2	0.1179	-0.2500	0.0000	0.0000
#3	0.1179	0.1118	0.0000	-0.2236
#4	0.1179	-0.1118	0.0000	-0.2236
#5	0.1179	0.1118	0.2127	-0.0691
#6	0.1179	-0.1118	-0.2127	0.0691
#7	0.1179	0.1118	-0.2127	-0.0691
#8	0.1179	-0.1118	0.2127	0.0691
#9	0.1179	0.1118	0.1314	0.1809
#10	0.1179	-0.1118	-0.1314	-0.1809
#11	0.1179	0.1118	-0.1314	0.1809
#12	0.1179	-0.1118	0.1314	-0.1809

Lsp	R	S	T	U	V
#1	-0.2083	0.0000	0.0000	0.3135	0.0000
#2	-0.2083	0.0000	0.0000	0.3135	0.0000
#3	0.2917	-0.2500	0.0000	0.0625	0.0000
#4	0.2917	-0.2500	0.0000	0.0625	0.0000
#5	-0.1606	-0.0773	-0.1469	-0.1636	0.2378
#6	-0.1606	-0.0773	-0.1469	-0.1636	0.2378
#7	-0.1606	-0.0773	0.1469	-0.1636	-0.2378
#8	-0.1606	-0.0773	0.1469	-0.1636	-0.2378
#9	0.1189	0.2023	0.2378	-0.0239	0.1469
#10	0.1189	0.2023	0.2378	-0.0239	0.1469
#11	0.1189	0.2023	-0.2378	-0.0239	-0.1469
#12	0.1189	0.2023	-0.2378	-0.0239	-0.1469

Table 4.4: Ambisonics Decoding Coefficients, up to the Second Order, for a Tilted Dodecahedron Loudspeaker Configuration

additional filtering and equalization have to be applied in the signal path leading to each individual loudspeaker. In this sense, the monophonic output signal from each mixer is subsequently sent to its own delay processor, for a total of twelve virtual delay units. Each one of these units will assign a particular amount of delay to each of the twelve decoded signals, so when assigned to the loudspeakers the dodecahedron is correctly recreated, by simulating an equal distance between each loudspeaker and the origin. At the end of each delay processor, there is a gain component for each of the twelve signals where, additionally to the delay, the signals will be equalized in gain so they recreate the original spherical configuration.

Loudspeaker	x-coordinate	y-coordinate	z-coordinate
#1	1.88	0	0
#2	-1.88	0	0
#3	0.6149	0	-1.2298
#4	-0.6149	0	1.2298
#5	0.7884136	1.4997841	-0.4872932
#6	-0.7884136	-1.4997841	0.4872932
#7	0.7884136	-1.4997841	-0.4872932
#8	-0.7884136	1.4997841	0.4872932
#9	0.76015056	0.89358486	1.22997528
#10	-0.76015056	-0.89358486	-1.22997528
#11	0.76015056	-0.89358486	1.22997528
#12	-0.76015056	0.89358486	-1.22997528

Table 4.5: Cartesian Coordinates of the Loudspeakers in the Modified Decoding Configuration

Test Conditions and Variables:	
Stimulus type	Free field simulated pink noise Reverberant field simulated pink noise
Number of source positions	9
Rendering technique	First Order Ambisonics Second Order Ambisonics
Number of times each case is played	2

Table 4.6: Varying Components of the Experiment

Each gain unit has a mono output which is directly assigned to its corresponding loudspeaker.

Thus, as described in Section 4.2.2, there are nine source locations for which we want to test the localization blur. To review, the localization test variables are shown in Table 4.6. The eighty possible stimuli of the experiment are played in random order. The stimuli are not previously encoded as Ambisonics files. The encoding as well as decoding process are performed in real time.

4.3 Results of the Perceptual Localization Experiment

The experiment consisted of four main categories as seen in Table 4.1. For each category of the experiment, there were a total of 280 stimuli distributed over the subject population. Out of the 280 stimuli, half consisted of free field environment simulations, while the other half consisted of reverberant environment simulated stimuli. For each source position in each category we had 56 answers over the subject population; 28 being stimuli simulated in the free field, and 28 in a reverberant field. However, this did not apply to the categories testing for localization blur in the horizontal plane. In these categories, we eventually eliminated the data associated with the stimuli located at a -40° azimuth (on the y axis of the Ambisonics coordinate system). The reason is that many of the subjects commented that they mostly had heard the sound coming from the outside of our visual projection area, and thus the answers could not be properly recorded. In these conditions we are left with a total of 224 stimuli for both first- and second-order Ambisonics in the horizontal plane categories.

If we analyzed the data for most subjects, we would notice that there is no consistency in the response trends of each subject for all positions which could allow us to determine outliers or to make a difference between the experienced vs. unexperienced listeners. Out of a total of 14 subjects, an exception was observed for only one listener, whose results showed consistency in the responses at all positions and also for stimuli repetitions, while localizing very closely to the stimuli locations; especially in the horizontal plane. This particular subject was also the most experienced listener, being an active player in orchestras and chamber groups as well as an experienced acoustician, more versed into the taking of listening tests than all the other subjects. We can reason that there could be a learning curve that would eventually influence this perceptual experiment, and the possibility exists that if all subjects went through a longer learning process, the experiment results could produce smaller similar blurs. These results will be analyzed in parallel and compared with the mean responses over the whole subject population.

In order to present the results in this chapter, we will refer to the stimulus azimuth angle as θ_s , and to the response/perceived azimuth angle as θ_r . For the

localization in the median plane categories, we will refer to the stimulus elevation angle as ϕ_s , while the perceived angle will be referred to as ϕ_r .

4.4 Perceptual Investigation Results with Mean Signed Error Values

One approach in analyzing the data resulted from this experiment is considering the mean perceived localization response for each condition tested. The mean perceived position is calculated according to Equations 4.1 and 4.2.

$$\bar{\theta}_{abserror} = 1/n \sum_{i=1}^n \theta_{ri} \quad (4.1)$$

$$\bar{\phi}_{abserror} = 1/n \sum_{i=1}^n \phi_{ri} \quad (4.2)$$

Along with the mean error values we will present the standard deviation values per each conditions, as well.

4.4.1 Mean Signed Localization Error Results in the Horizontal Plane, for a First-Order Ambisonics Sound Rendering System

This section reviews the responses in the category investigating the localization accuracy in the horizontal plane with first-order Ambisonics. For this section of the experiment, the subjects listened to stimuli located at a θ_s between -30° and 0° on the y-axis of the ambisonics coordinate system (see Figure 4.3). Subsequently, they had to determine their perceived azimuth angle, θ_r . The resulted data is shown in Figure 4.10.

Looking at Table 4.7, one can see that the θ_r mean values for the anechoic stimuli are close to their respective θ_s directions. However, the standard deviation by position values are large, varying from 7.4° for $\theta_s = 0^\circ$, up to 13.9° for $\theta_s = -20^\circ$.

Analyzing the data for the reverberant field simulated stimuli localization, one observes that there is not a considerable difference in the mean positions nor in standard deviation, compared to the anechoic cases. There is a difference of less than 1° in the standard deviation at all θ_s positions, except for $\theta_s = -30^\circ$, where a

Azimuth Stimulus Location	-30°	-20°	-10°	0°
<i>Free Field Environment</i>				
Mean of Perceived Position	-30.0°	-24.7°	-15.8°	-0.7°
Standard Deviation of Perceived Position	11.9°	13.9°	13.7°	7.4°
(Mean θ_r) $- \theta_s$	0.0°	-4.7°	-5.8°	-0.7°
<i>Reverberant Environment</i>				
Mean of Perceived Position	-28.6°	-20.5°	-12.0°	-1.1°
Standard Deviation of Perceived Position	14.2°	14.4°	12.4°	7.5°
(Mean θ_r) $- \theta_s$	1.4°	-0.5°	-2.0°	-1.1°

Table 4.7: Mean and Standard Deviation Values of the Azimuth Localization Results, with First Order Ambisonics

Azimuth Stimulus Location	-30°	-20°	-10°	0°
<i>Free Field Environment</i>				
Mean of Perceived Position	-26.2°	-21.2°	-7.7°	-0.4°
Standard Deviation of Perceived Position	2.6°	2.2°	1.0°	6.2°
(Mean θ_r) $- \theta_s$	3.8°	-4.7°	-2.3°	-0.4°
<i>Reverberant Environment</i>				
Mean of Perceived Position	-28.0°	-17.4°	-11.3°	0.7°
Standard Deviation of Perceived Position	3.8°	17.6°	16.6°	1.0°
(Mean θ_r) $- \theta_s$	2.0°	2.6°	-1.3°	0.7°

Table 4.8: Best-case Mean and Standard Deviation Values of the Azimuth Localization Results, with First Order Ambisonics

difference between the anechoic and reverberant stimuli standard deviations of 2.3° occurred.

Table 4.8 and Figure 4.11 show the results for the subject with best localization performance. We observe very high localization accuracy in a free field environment, with small mean error and standard deviations values. Larger localization errors are noticed for the reverberant field simulations, especially in terms of standard deviations.

4.4.2 Mean Signed Localization Error Results in the Horizontal Plane, for a Second-Order Ambisonics Sound Rendering System

This section reviews the responses in the category investigating the horizontal plane localization with second-order Ambisonics. Similarly to the first-order Ambisonics azimuth localization cases, the subjects were played stimuli from virtual sources located at a θ_s between -30° and 0° on the y-axis of the ambisonics coordinate system (see Figure 4.3). The data resulted from their answers is shown in Figure 4.12.

Table 4.9 shows that while the θ_r mean values for both $\theta_s = 0^\circ$ and $\theta_s = -30^\circ$ are within a localization error of less than 3° from the original source position, the θ_r means corresponding to $\theta_s = -10^\circ$ and $\theta_s = -20^\circ$ are within mean localization errors of approximately 9° . Moreover, from the θ_r mean values we can conclude that the perceived positions are mostly shifting to the right of their corresponding θ_s , and away from the center. The standard deviation is reaching a minimum of 4.7° at $\theta_s = 0^\circ$ and a maximum of 13.7° at $\theta_s = -30^\circ$.

Regarding the localization results in a simulated reverberant environment, the most interesting aspect is a loss in the accuracy of the θ_r corresponding to $\theta_s = 0^\circ$: the mean perceived position is at -6° to the right of the actual stimuli source location, while the standard deviation increases by more than 5° compared to the results for the simulation in the free field. The difference between the localization errors in the free and reverberant environments decreases the farther away θ_s is from the center position.

Table 4.10 and Figure 4.13 show the results for the subject with best localization performance. Once again, very high localization accuracy characterizes these results, especially in a free field environment, with small mean error and standard deviations values. Larger localization errors are noticed again for the reverberant field simulations.

Azimuth Stimulus Location	-30°	-20°	-10°	0°
<i>Free Field Environment</i>				
Mean of Perceived Position	-32.2°	-29.3°	-19.9°	-1.7°
Standard Deviation of Perceived Position	13.7°	11.7°	13.1°	4.7°
(Mean θ_r) - θ_s	-2.2°	-9.3°	-9.9°	-1.7°
<i>Reverberant Environment</i>				
Mean of Perceived Position	-33.8°	-27.4°	-14.9°	-6.0°
Standard Deviation of Perceived Position	11.5°	12.5°	11.0°	10.4°
(Mean θ_r) - θ_s	-3.8°	-7.4°	-4.9°	-6.0°

Table 4.9: Mean and Standard Deviation Values of the Azimuth Localization Results, with Second Order Ambisonics

Azimuth Stimulus Location	-30°	-20°	-10°	0°
<i>Free Field Environment</i>				
Mean of Perceived Position	-27.1°	-21.6°	-7.9°	0.8°
Standard Deviation of Perceived Position	8.4°	5.2°	0.7°	0°
(Mean θ_r) - θ_s	2.9°	-1.6°	2.1°	0.8°
<i>Reverberant Environment</i>				
Mean of Perceived Position	-28.5°	-27.3°	-10.0°	-10.5°
Standard Deviation of Perceived Position	5.7°	1.9°	6.4°	5.4°
(Mean θ_r) - θ_s	1.5°	-7.3°	0.0°	-10.5°

Table 4.10: Best-case Mean and Standard Deviation Values of the Azimuth Localization Results, with Second-Order Ambisonics

4.4.3 Comparison of the localization accuracy in the horizontal plane between first and second order Ambisonics rendering

After reviewing the localization error results in the horizontal plane for both first- and second-order Ambisonics renderings (see Sections 4.4.1 and 4.4.2) we cannot determine a significant difference in the perceptual localization accuracy between these two categories. One can compare the results between the two Ambisonics orders against each other in Table 4.11 for the entire population, and in Table 4.12 for the subject with best-performance.

Regarding the results over the entire subject population, for the free field environment simulations the mean localization error values in the second-order Ambisonics category are larger than the values for the first order Ambisonics cases,

Azimuth Stimulus Location (deg)	-30°	-20°	-10°	0°
<i>Free Field Environment</i>				
FIRST ORDER AMBISONICS				
Mean of Perceived Position	-30.0°	-24.7°	-15.8°	-0.7°
Standard Deviation of Perceived Position	11.9°	13.9°	13.7°	7.4°
(Mean θ_r) $- \theta_s$	0.0°	-4.7°	-5.8°	-0.7°
SECOND ORDER AMBISONICS				
Mean of Perceived Position	-32.2°	-29.3°	-19.9°	-1.7°
Standard Deviation of Perceived Position	13.7°	11.7°	13.1°	4.7°
(Mean θ_r) $- \theta_s$	-2.2°	-9.3°	-9.9°	-1.7°
<i>Reverberant Environment</i>				
FIRST ORDER AMBISONICS				
Mean of Perceived Position	-28.6°	-20.5°	-12.0°	-1.1°
Standard Deviation of Perceived Position	14.2°	14.4°	12.4°	7.5°
Localization Blur by Position	1.4°	-0.5°	-2.0°	-1.1°
SECOND ORDER AMBISONICS				
Mean of Perceived Position	-33.8°	-27.4°	-14.9°	-6.0°
Standard Deviation of Perceived Position	11.5°	12.5°	11.0°	10.4°
Localization Blur by Position	-3.8°	-7.4°	-4.9°	-6.0°

Table 4.11: Azimuth Localization Comparison for First and Second Order Ambisonics Renderings, over the entire subject population

while the stimulus image is consistently shifted to the right of θ_s and away from the center. However, in terms of the standard deviation by position for the same cases, the largest standard deviation values are encountered with the first order Ambisonics rendering, except the values corresponding to $\theta_s = -30^\circ$, where the standard deviation for the second order rendering is larger by 1.8° than the first order's one. As for the stimuli simulated in a reverberant environment, the results are once again similar to the results of the free field cases. The localization errors for the first-order Ambisonics category are once again close to their corresponding θ_s , while the standard deviation differences between the first- and second-order are smaller than 3° . Even if overall the mean localization errors are smaller for the first-order Ambisonics cases, their corresponding standard deviation values show that the responses given for the second-order Ambisonics rendered stimuli are generally (but not always) closer to their corresponding θ_s than the responses for the first-order Ambisonics rendered stimuli.

Azimuth Stimulus Location	-30°	-20°	-10°	0°
<i>Free Field Environment</i>				
FIRST ORDER AMBISONICS				
Mean of Perceived Position	-26.2°	-21.2°	-7.7°	-0.4°
Standard Deviation of Perceived Position	2.6°	2.2°	1.0°	6.2°
(Mean θ_r) - θ_s	3.8°	-4.7°	-2.3°	-0.4°
SECOND ORDER AMBISONICS				
Mean of Perceived Position	-27.1°	-21.6°	-7.9°	0.8°
Standard Deviation of Perceived Position	8.4°	5.2°	0.7°	0°
(Mean θ_r) - θ_s	2.9°	-1.6°	2.1°	0.8°
<i>Reverberant Environment</i>				
FIRST ORDER AMBISONICS				
Mean of Perceived Position	-28.0°	-17.4°	-11.3°	0.7°
Standard Deviation of Perceived Position	3.8°	17.6°	16.6°	1.0°
(Mean θ_r) - θ_s	2.0°	2.6°	-1.3°	0.7°
SECOND ORDER AMBISONICS				
Mean of Perceived Position	-28.5°	-27.3°	-10.0°	-10.5°
Standard Deviation of Perceived Position	5.7°	1.9°	6.4°	5.4°
(Mean θ_r) - θ_s	1.5°	-7.3°	0.0°	-10.5°

Table 4.12: Azimuth Localization Comparison for First and Second Order Ambisonics Renderings, for the subject with best-performance

Regarding the best-performance subject results, one can notice almost no difference in the free field simulated results between the two Ambisonics orders. As for the reverberant field simulations, once again the results are very similar between first- and second-order Ambisonics for the -30° and -10° stimuli locations. However, for the stimuli located at -20° and 0° , errors of around 10° occurred between the two conditions.

4.4.4 Mean Signed Localization Error Results in the Median Plane, for a First-Order Ambisonics Sound Rendering System

This section reviews the responses investigating localization in the median plane with first-order Ambisonics. Under this condition, the subjects were played stimuli generated by a virtual source located at a ϕ_s between 0° and 20° degrees on the z-axis of the ambisonics coordinate system (see Figure 4.4), while keeping

Elevation Stimulus Location	0°	5°	10°	15°	20°
<i>Free Field Environment</i>					
Mean of Perceived Position	9.1°	8.7°	6.9°	7.4°	9.9°
Standard Deviation of Perceived Position	12.3°	11.9°	11.4°	8.4°	11.1°
(Mean θ_r) - θ_s	9.1°	3.7°	-3.1°	-7.6°	-10.1°
<i>Reverberant Environment</i>					
Mean of Perceived Position	13.5°	12.1°	16.7°	11.8°	11.8°
Standard Deviation of Perceived Position	10.5°	9.6°	10.0°	12.0°	9.9°
(Mean θ_r) - θ_s	13.5°	7.1°	6.7°	-3.2°	-8.2°

Table 4.13: Mean and Standard Deviation Values of the Elevation Localization Results, with First-Order Ambisonics

Elevation Stimulus Location	0°	5°	10°	15°	20°
<i>Free Field Environment</i>					
Mean of Perceived Position	10.3°	4.9°	2.0°	16.5°	2.5°
Standard Deviation of Perceived Position	20.8°	13.4°	8.4°	17.0°	11.0°
(Mean θ_r) - θ_s	10.3°	-0.1°	-8.0°	1.5°	-17.5°
<i>Reverberant Environment</i>					
Mean of Perceived Position	20.0°	12.2°	27.5°	14.8°	6.7°
Standard Deviation of Perceived Position	9.7°	18.1°	1.9°	14.4°	4.8°
(Mean θ_r) - θ_s	20.0°	7.2°	17.5°	-0.2°	-13.3°

Table 4.14: Best-case Mean and Standard Deviation Values of the Elevation Localization Results, with First-Order Ambisonics

a constant azimuth of 0°. Subsequently, the listeners had to click and record their perceived elevation angle ϕ_r . The resulted data is shown in Figure 4.14.

Table 4.13 shows the ϕ_r mean and standard deviation by position and type of environment values, over the entire population. We notice that independently of ϕ_s , the perceived ϕ_r mean values are consistently grouped between approximately 8° and 10°, with a standard deviation of 11°-12°. For the stimuli simulated in a reverberant environment, the mean of ϕ_r is above 11.8°, independently of ϕ_s , with a standard deviation around 10° to 12°.

Table 4.14 and Figure 4.15, show that in this case even the best subject's performance is comparable to the localization performance over the whole population, being equally poor.

Elevation Stimulus Location	0°	5°	10°	15°	20°
<i>Free Field Environment</i>					
Mean of Perceived Position	12.0°	11.4°	10.6°	11.3°	12.0°
Standard Deviation of Perceived Position	10.1°	13.2°	13.0°	11.9°	12.9°
(Mean θ_r) - θ_s	12.0°	6.4°	0.6°	-3.7°	-8.0°
<i>Reverberant Environment</i>					
Mean of Perceived Position	12.9°	11.7°	15.9°	15.8°	16.7°
Standard Deviation of Perceived Position	10.4°	13.2°	11.5°	12.2°	11.7°
(Mean θ_r) - θ_s	12.9°	6.7°	5.9°	0.8°	-3.3°

Table 4.15: Mean and Standard Deviation Values of the Elevation Localization Results, with Second Order Ambisonics

4.4.5 Mean Signed Localization Error Results in the Median Plane, for a Second-Order Ambisonics Sound Rendering System

This section reviews the responses investigating the localization accuracy in the median plane using second-order Ambisonics. Similarly to the first-order Ambisonics median plane localization cases, the subjects were played stimuli from virtual sources located at a ϕ_s between 0° and 20° on the z-axis of the Ambisonics coordinate system, while keeping a constant azimuth of 0°. The data resulted from their answers is shown in Figure 4.16.

Table 4.15 shows that independently of ϕ_s , the means of the perceived locations of the anechoic stimuli are all gathered around a ϕ_r of approximately 11°, with a standard deviation between 10.1° and 13.2°. The responses corresponding to the reverberant environment simulations have higher values than the anechoic cases, their ϕ_r values ranging between 11.7° and 16.7°.

Having the same trend, Table 4.16 and Figure 4.17, show that for this condition the localization errors for best-performance subject were also very large, as they were for the entire subject population, both for free and reverberant field simulated stimuli.

Elevation Stimulus Location	0°	5°	10°	15°	20°
<i>Free Field Environment</i>					
Mean of Perceived Position	14.2°	23.5°	-5.1°	2.6°	7.2°
Standard Deviation of Perceived Position	18.0°	3.9°	2.5°	7.3°	10.3°
(Mean θ_r) - θ_s	14.2°	18.5°	-15.1°	-12.4°	-12.8°
<i>Reverberant Environment</i>					
Mean of Perceived Position	13.2°	7.9°	14.1°	17.1°	4.9°
Standard Deviation of Perceived Position	15.5°	11.2°	10.9°	16.2°	7.3°
(Mean θ_r) - θ_s	13.2°	2.9°	4.1°	2.0°	-15.1°

Table 4.16: Best-case Mean and Standard Deviation Values of the Elevation Localization Results, with Second-Order Ambisonics

4.4.6 Comparison of the localization accuracy in the median plane between 1st and 2nd order Ambisonics rendering

After reviewing the results for both first- and second-order Ambisonics localization blurs in the median plane (see Sections 4.4.4 and 4.4.5) we cannot determine a significant difference in the rendering accuracy between these two categories.

Table 4.17 compares the results between the two Ambisonics orders against each other, over the entire population. Looking at the mean positions in the free field, one can notice that the means for the first-order Ambisonics category are grouping approximately between 7° and 10° no matter of the actual stimulus location, while in the second order Ambisonics category the mean positions are shifting up reaching values of 10.6° to 12°. Regarding the reverberant environment category, we notice that all ϕ_r mean values are grouped between approximately 12° and 16°, independently of ϕ_s or Ambisonics order.

Similar results are observed with the best-performance subject, where we cannot establish a relationship between stimuli localtions and perceived locations, for any of the testing conditions. Table 4.18, shows these results once again, for comparison purposes.

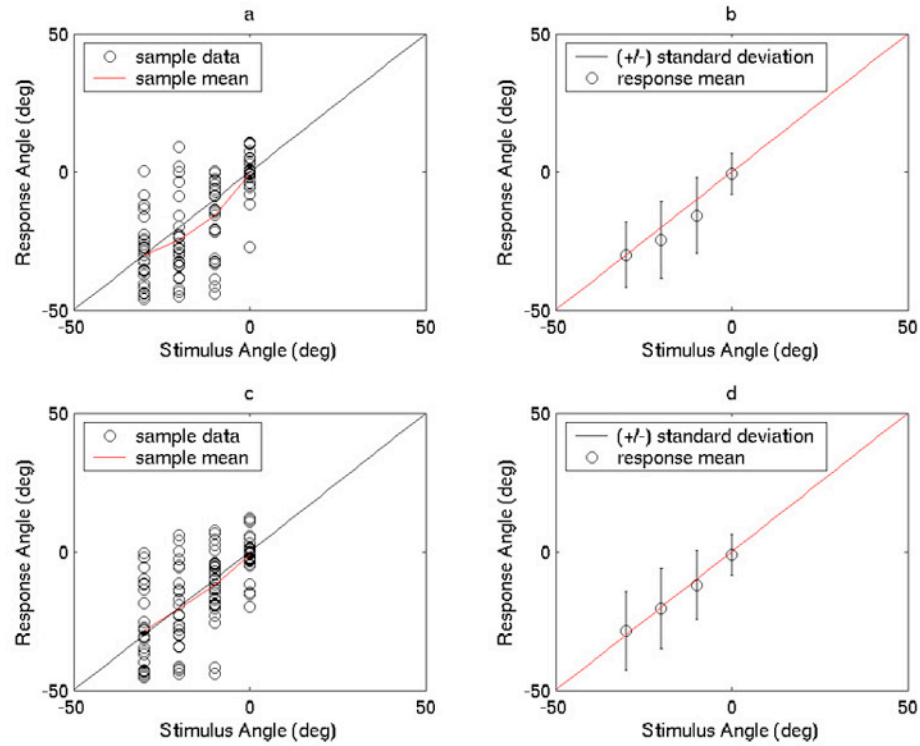


Figure 4.10: Localization Results in the Horizontal Plane for First Order Ambisonics: a & b: free field simulations; c & d: reverberant field simulations

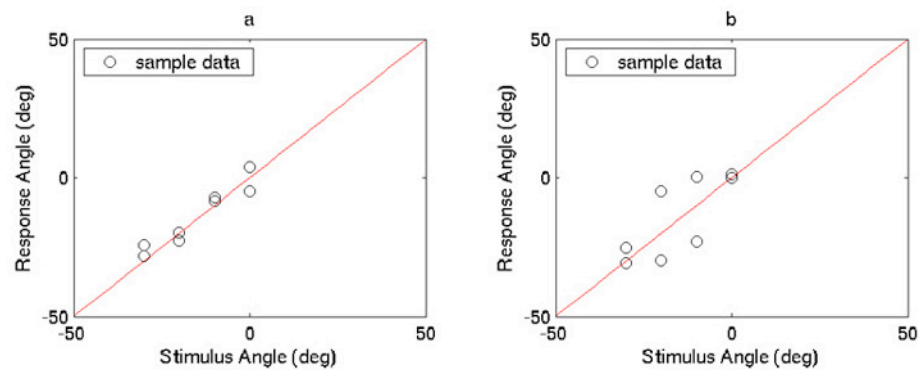


Figure 4.11: Best-case localization results in the Horizontal Plane for first-order Ambisonics: a - free field simulations; b - reverberant field simulations

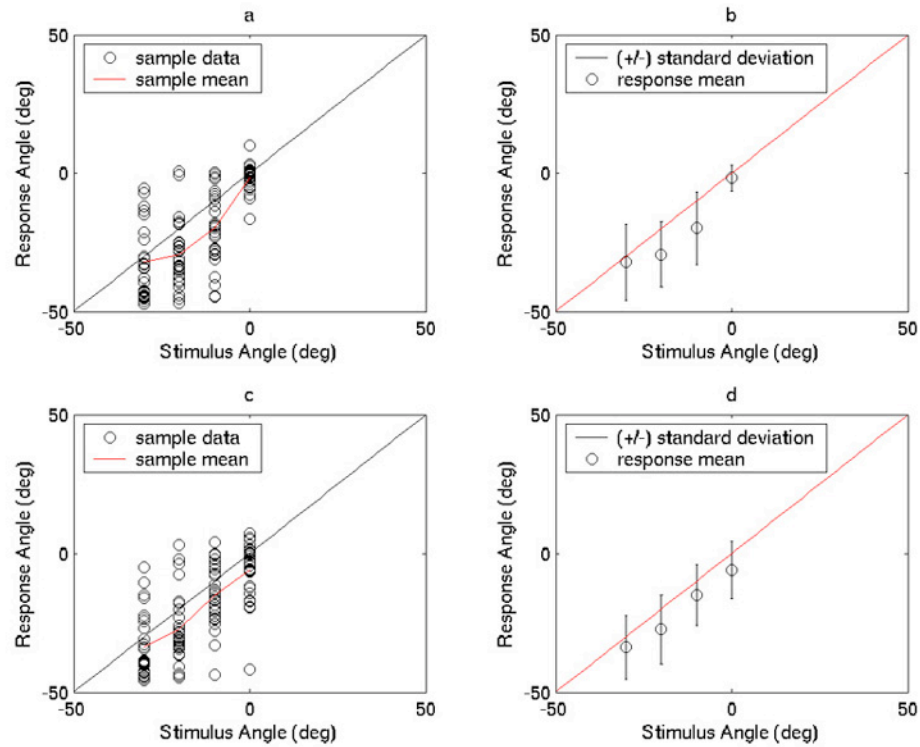


Figure 4.12: Localization Results in the Horizontal Plane for Second Order Ambisonics: a & b: free field simulations; c & d: reverberant field simulations

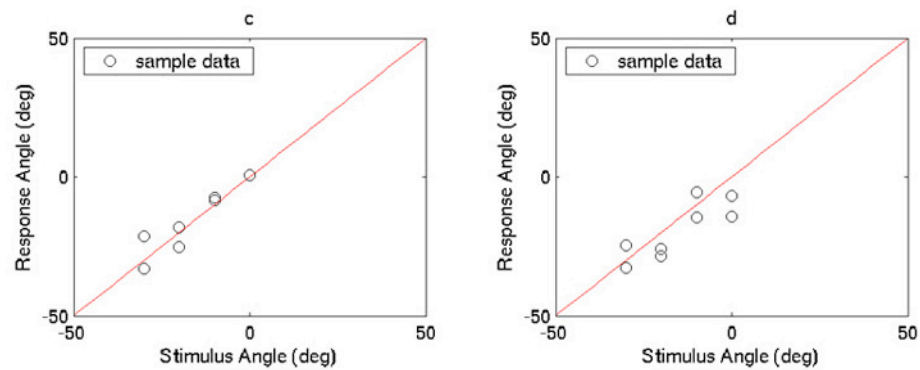


Figure 4.13: Best-case localization results in the Horizontal Plane for second-order Ambisonics: a - free field simulations; b - reverberant field simulations

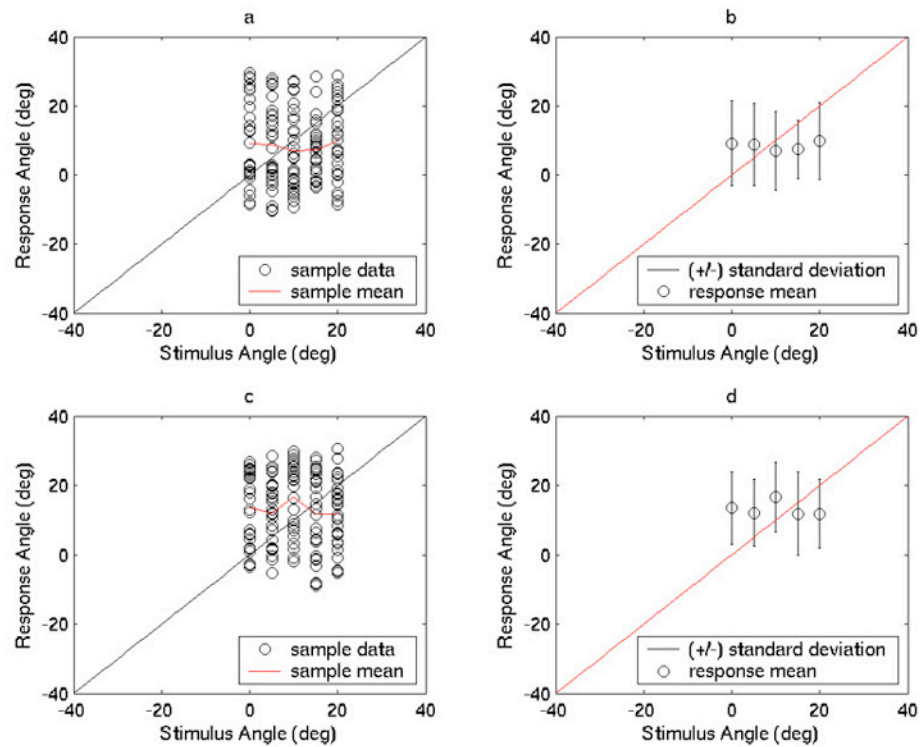


Figure 4.14: Localization Results in the Median Plane for First Order Ambisonics: a & b: free field simulations; c & d: reverberant field simulations

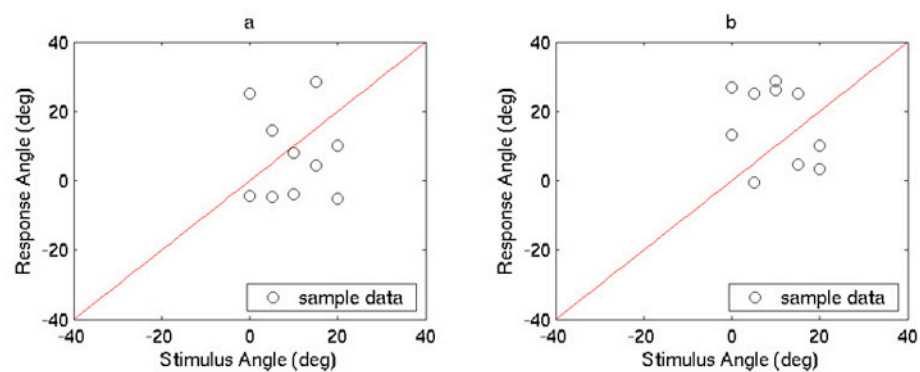


Figure 4.15: Best-case localization results in the Median Plane for first-order Ambisonics: a - free field simulations; b - reverberant field simulations

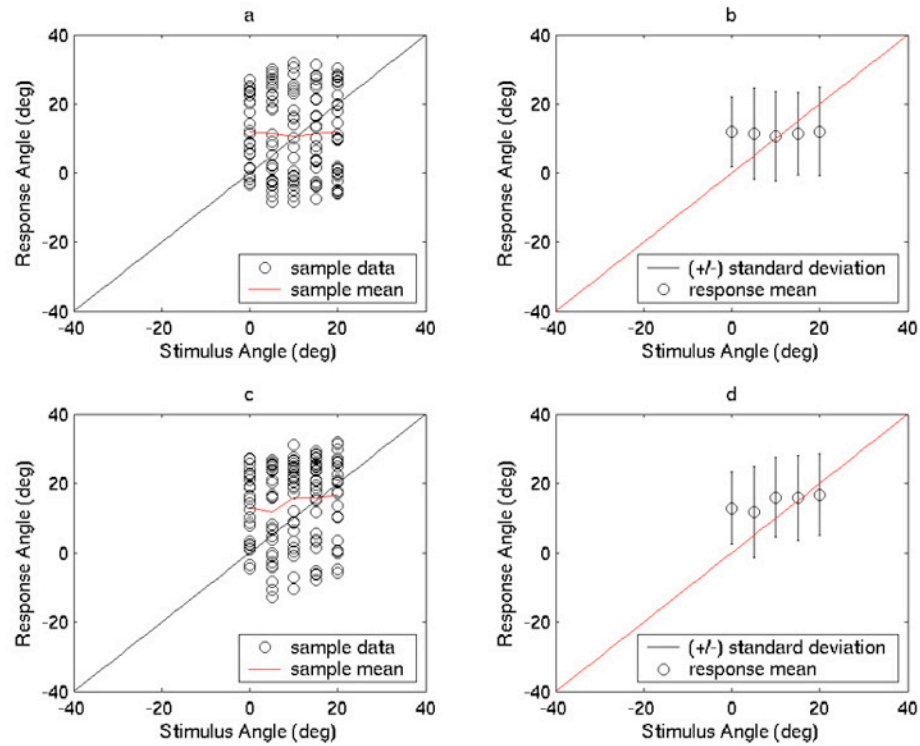


Figure 4.16: Localization Results in the Median Plane for Second Order Ambisonics: a & b: free field simulations; c & d: reverberant field simulations

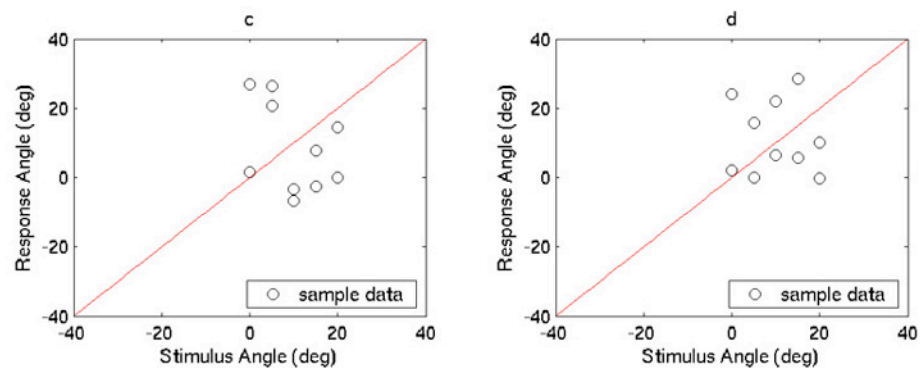


Figure 4.17: Best-case localization results in the Median Plane for second-order Ambisonics: a - free field simulations; b - reverberant field simulations

Elevation Stimulus Location (deg)	0°	5°	10°	15°	20°
<i>Free Field Environment</i>					
FIRST ORDER AMBISONICS					
Mean of Perceived Position	9.1°	8.7°	6.9°	7.4°	9.9°
Standard Deviation of Perceived Position	12.3°	11.9°	11.4°	8.4°	11.1°
(Mean θ_r) - θ_s	9.1°	3.7°	-3.1°	-7.6°	-10.1°
SECOND ORDER AMBISONICS					
Mean of Perceived Position	12.0°	11.4°	10.6°	11.3°	12.0°
Standard Deviation of Perceived Position	10.1°	13.2°	13.0°	11.9°	12.9°
(Mean θ_r) - θ_s	12.0°	6.4°	0.6°	-3.7°	-8.0°
<i>Reverberant Environment</i>					
FIRST ORDER AMBISONICS					
Mean of Perceived Position	13.5°	12.1°	16.7°	11.8°	11.8°
Standard Deviation of Perceived Position	10.5°	9.6°	10.0°	12.0°	9.9°
(Mean θ_r) - θ_s	13.5°	7.1°	6.7°	-3.2°	-8.2°
SECOND ORDER AMBISONICS					
Mean of Perceived Position	12.9°	11.7°	15.9°	15.8°	16.7°
Standard Deviation of Perceived Position	10.4°	13.2°	11.5°	12.2°	11.7°
(Mean θ_r) - θ_s	12.9°	6.7°	5.9°	0.8°	-3.3°

Table 4.17: Median Plane Localization Comparison for First and Second Order Ambisonics Renderings

4.5 Perceptual Investigation Results with Absolute Error Values

Another approach in analyzing the data resulted from the perceptual investigation is considering the average absolute localization error per condition. This analysis choice can be explained as being able to characterize the localization blur specific to the system, while ignoring the auditory system generated localization blur. In contrast to the signed error, which reflects more the auditory performance of the human participants, the absolute error indicates more the system performance, i.e., how much blur is introduced by the simulation technique/rendering system, as compared to reality. The average absolute localization error for each condition is calculated according to Equations 4.3 and 4.4.

$$\bar{\theta}_{abserror} = 1/n \sum_{i=1}^n |\theta_{ri} - \theta_{si}| \quad (4.3)$$

Elevation Stimulus Location	0°	5°	10°	15°	20°
<i>Free Field Environment</i>					
FIRST ORDER AMBISONICS					
Mean of Perceived Position	10.3°	4.9°	2.0°	16.5°	2.5°
Standard Deviation of Perceived Position	20.8°	13.4°	8.4°	17.0°	11.0°
(Mean θ_r) - θ_s	10.3°	-0.1°	-8.0°	1.5°	-17.5°
SECOND ORDER AMBISONICS					
Mean of Perceived Position	14.2°	23.5°	-5.1°	2.6°	7.2°
Standard Deviation of Perceived Position	18.0°	3.9°	2.5°	7.3°	10.3°
(Mean θ_r) - θ_s	14.2°	18.5°	-15.1°	-12.4°	-12.8°
<i>Reverberant Environment</i>					
FIRST ORDER AMBISONICS					
Mean of Perceived Position	20.0°	12.2°	27.5°	14.8°	6.7°
Standard Deviation of Perceived Position	9.7°	18.1°	1.9°	14.4°	4.8°
(Mean θ_r) - θ_s	20.0°	7.2°	17.5°	-0.2°	-13.3°
SECOND ORDER AMBISONICS					
Mean of Perceived Position	13.2°	7.9°	14.1°	17.1°	4.9°
Standard Deviation of Perceived Position	15.5°	11.2°	10.9°	16.2°	7.3°
(Mean θ_r) - θ_s	13.2°	2.9°	4.1°	2.0°	-15.1°

Table 4.18: Median Plane Localization Comparison for First and Second Order Ambisonics Renderings, for the subject with best-performance

$$\bar{\phi}_{abserror} = 1/n \sum_{i=1}^n |\phi_{ri} - \phi_{si}| \quad (4.4)$$

Table 4.19 shows the average absolute error results for all for conditions covered in the previous sections.

One can see that under all conditions there is a localization average absolute error larger than 9°, up to a maximum of 12°.

Table 4.20 shows the average absolute localization error per condition for the best-performance subject. These results are characterized by very good performance in the horizontal plane with free field simulations for both first- and second-order Ambisonics (around 3°). A very good localization is noticed for the reverberant field simulations in the horizontal plane as well, but of double error if compared to the free field conditions. However, the responses are again consistent between the first-

Testing Condition	Type of environment	$\theta_{abserror}$	$\phi_{abserror}$
Azimuth Localization with First Order Ambisonics	Free Field	9.4°	
Azimuth Localization with First Order Ambisonics	Reverberant Field	9.4°	
Azimuth Localization with Second Order Ambisonics	Free Field	10.3°	
Azimuth Localization with Second Order Ambisonics	Reverberant Field	10.2°	
Elevation Localization with First Order Ambisonics	Free Field		10.7°
Elevation Localization with First Order Ambisonics	Reverberant Field		11.0°
Elevation Localization with Second Order Ambisonics	Free Field		12.0°
Elevation Localization with Second Order Ambisonics	Reverberant Field		11.6°

Table 4.19: Absolute Error Analysis of Localization Results

and second-order Ambisonics. In the median plane, localization errors similar to the errors resulted over the entire subject population are noticed. From these results we can conclude the following:

- In the horizontal plane better performance results may be achieved with training or as the subjects improve with repetition, if we compare with the data shown in Table 4.19;
- In the median plane the results are similar to the average response over the whole population (by comparison with Table 4.19), and this proves that the system performance along with the environmental conditions (i.e. floor reflections) and the auditory system characteristics are probably not going to allow for a better localization accuracy, even if more training is applied to the listeners.

Testing Condition	Type of environment	$\theta_{abserror}$	$\phi_{abserror}$
Azimuth Localization with First Order Ambisonics	Free Field	3.0°	
Azimuth Localization with First Order Ambisonics	Reverberant Field	6.9°	
Azimuth Localization with Second Order Ambisonics	Free Field	3.1°	
Azimuth Localization with Second Order Ambisonics	Reverberant Field	6.6°	
Elevation Localization with First Order Ambisonics	Free Field		12.4°
Elevation Localization with First Order Ambisonics	Reverberant Field		14.7°
Elevation Localization with Second Order Ambisonics	Free Field		14.6°
Elevation Localization with Second Order Ambisonics	Reverberant Field		11.1°

Table 4.20: Absolute Error Analysis of Localization Results, for subject with best performance

4.6 Correlation between stimuli location and perceived location

Correlation is the interdependence between two or more variables [17]. If two random variables are such that when one changes the other does so in a related manner they are said to be correlated. In this section we are trying to determine if there was any relationship/correlation between the stimuli original locations and the perceived responses in our experiment. To achieve this goal we calculate the coefficient of determination which equals the square of the correlation coefficient (i.e. a measure of strength of association between two variables X and Y). According to [17], “the coefficient of determination can be interpreted as the proportion of variability in Y that can be accounted for knowing X, or the proportion of variability in X that can be accounted for knowing Y”. The coefficient of determination is calculated according to Equation 4.5.

$$r_{xy}^2 = (Cov_{xy})^2 / s_x^2 s_y^2 \quad (4.5)$$

The coefficient of determination can reach values between 0 and 1. When it equals zero, the two variables are independent from each-other, while the closer it gets to 1, the greater the percentage of variance in X is accounted for by knowing Y and vice-versa.

The data in Table 4.21 shows to what extent the stimulus and response angles over the whole subject population are correlated for each condition covered in our experiment. In this way we are able to determine the validity of the analysis previously done in this chapter. Thus, the coefficients of determination for all conditions testing localization in the horizontal plane are fairly similar, varying between 0.41 to 0.48. The values show a fairly good correlation between the original stimuli locations and responses while, because of the similar values, we can trust that the relationship resulted between conditions is proportionally valid. On the other hand, looking at the correlation resulted in the vertical plane localization conditions, the coefficients show no relationship between stimuli and responses. The coefficients of determination for these cases explain the results described in previous sections, confirming that the response positions (each time grouped around one response angle) were probably a matter of guessing.

Table 4.22 show the coefficients of determination for the subject with best-performance results. These results are characterized by very good correlation in the horizontal plane, while in the median plane no correlation is observed between the two variables. Again these values confirm our results' discussions from previous sections.

Testing Condition	Type of environment	Coefficient of Determination
Azimuth Localization with First Order Ambisonics	Free Field	0.45
Azimuth Localization with First Order Ambisonics	Reverberant Field	0.41
Azimuth Localization with Second Order Ambisonics	Free Field	0.48
Azimuth Localization with Second Order Ambisonics	Reverberant Field	0.48
Elevation Localization with First Order Ambisonics	Free Field	0.00
Elevation Localization with First Order Ambisonics	Reverberant Field	0.00
Elevation Localization with Second Order Ambisonics	Free Field	0.00
Elevation Localization with Second Order Ambisonics	Reverberant Field	0.02

Table 4.21: Correlation Coefficients of Localization Results

Testing Condition	Type of environment	Coefficient of Determination
Azimuth Localization with First Order Ambisonics	Free Field	0.92
Azimuth Localization with First Order Ambisonics	Reverberant Field	0.58
Azimuth Localization with Second Order Ambisonics	Free Field	0.89
Azimuth Localization with Second Order Ambisonics	Reverberant Field	0.70
Elevation Localization with First Order Ambisonics	Free Field	0.00
Elevation Localization with First Order Ambisonics	Reverberant Field	0.10
Elevation Localization with Second Order Ambisonics	Free Field	0.17
Elevation Localization with Second Order Ambisonics	Reverberant Field	0.01

Table 4.22: Correlation Coefficients of Localization Results

CHAPTER 5

AMBISONICS SYSTEMS PERFORMANCE

INVESTIGATION

This chapter describes a series of sound field simulations meant to evaluate the performance of an Ambisonics sound rendering system. This investigation simulates the Ambisonics configuration from our perceptual experiments. We will study this configuration both in its ideal form, as well as in its warped version used in our experiments to fit our rectangular listening space. It consists of a twelve loudspeakers spherical set-up. Below there are described the goals, methods and results developed in this research.

5.1 Goals for the Performance Simulation Investigation

The main goal of this investigation is to simulate the sound wavefronts as reproduced with an Ambisonics system in order to:

- Confirm the theory on the Ambisonics technique performance, described in Chapter 3;
- Support the results of the perceptual investigation experiments, described in Chapter 4;
- Visually determine the size of the effective listening area, when using such a sound reproduction system.

5.2 Simulation Procedure and Design

This section describes the procedures leading to the simulations' development. The simulations were performed using the *Spatial Acoustic Suite* software [18]. The program was custom developed as part of [19], in order to support similar research investigations; [19] also confirms the validity of the *Spatial Acoustic Suite* simulations against real-world wavefield measurements.

By using the *Spatial Acoustics Suite* software, we were able to accomplish the following steps for a successful simulation:

- Construct a physical model of the loudspeaker system. Due to the software’s applications, we were able to implement the physical model of the particular type of loudspeaker used in our perceptual investigations: the EAW JF60 two-way loudspeaker system;
- Introduce variable DSP elements (gain and delay parameters for each individual source/loudspeaker) in order to reproduce the LARES Signal Processor behavior used for the recreation of the dodecahedron configuration in our perceptual experiments;
- Calculate the wave field at specific frequencies.

As described in [19], to create the simulation of the wave field, the *Spatial Acoustic Suite* models a tessellated listening plane using a square grid, and computes the complex pressure generated by each source at each grid position. The pressure from all sources is then summed at each position producing the model of the complex wave field at that point. In this chapter we will view the wave field’s coincident response in the form of $Re[\bar{H}(jw)]$. The coincident response calculation corresponds to computing a spatial wavefront produced by exciting the system with a cosine or sine wave input, and thus showing the shape and curvature of the wavefront at one particular frequency, as produced by the sound rendering system.

Using Matlab code, we compute the impulse responses corresponding to each of the twelve loudspeakers, which eventually recreate the sound field generated at the following virtual sound source positions:

- *Source 1*: on the sphere arc at an azimuth $= -10^\circ$ (to the right of the listener) and elevation $= 0^\circ$; this location coincides with one of the source positions used in the localization tests for determining localization blur in the horizontal plane (see Figure 5.1).

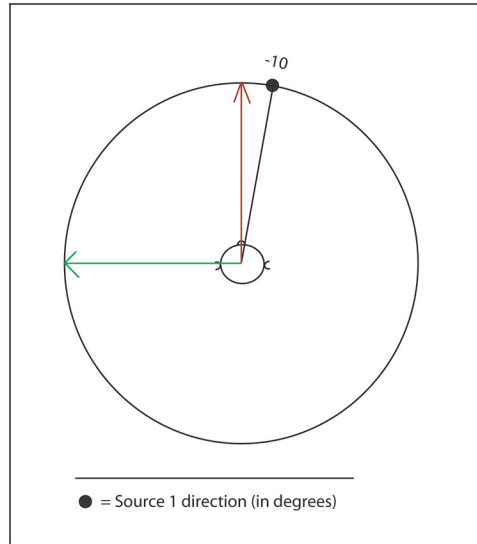


Figure 5.1: Source 1 angular direction

- *Source 2*: on an arc at an elevation = 15° and an azimuth = 0° ; this location is also one of the source locations in the perceptual experiments (see Figure 5.2).
- *Source 3*: at a point on the sphere located in the middle of three loudspeakers, starting from loudspeaker #1 and advancing mid-way between loudspeakers #7 and #11, at an azimuth = -27.93° and elevation = 8.66° (for reference on the loudspeakers location in the coordinate system see Table 4.3). We chose this location as the worst case scenario in terms of reproducing a wavefront coming from that direction as it is not directly supported by any of the loudspeakers (see Figure 5.3).

The distance of all virtual sources to the listener is kept constant. We will compare the rendering of the wavefront generated at these positions with both first- and second-order Ambisonics rendering.

Once we have the impulse responses we want to use, with *Spatial Acoustics Suite* we simulate the wavefield using two loudspeaker configurations:

1. the ideal dodecahedron configuration, where all twelve loudspeakers are equidistant from the origin;

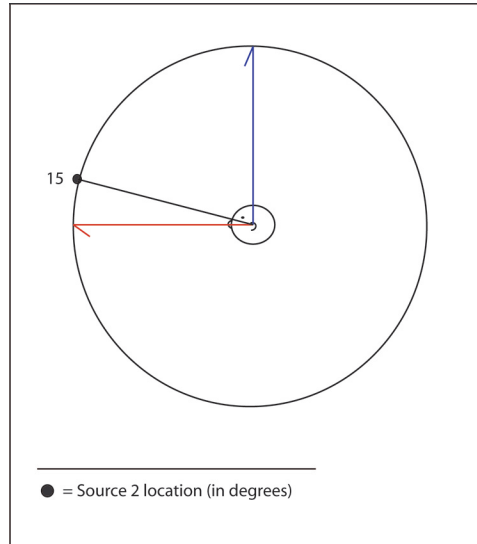


Figure 5.2: Source 2 angular direction

2. the dodecahedron configuration used by us in the perceptual experiments, warped to fit into our rectangular listening space. In this case we individually apply delays and gain changes to each source, within the *Spatial Acoustic Suite* application, to compensate for the distance differences between each loudspeaker and the origin.

To show the performance of the system, we will reproduce a pure tone spherical wave at different frequencies, coming from the source locations described above.

5.3 Simulation Results

5.3.1 Ideal simulation of the wavefront generated by a single sound source

The results of the Ambisonics renderings can be compared to reference cases, consisting of the simulation of similar pure tone wavefronts presented in their ideal form, as if being generated by only one sound source. One can see in Figures 5.4 to 5.6 these single source simulations, generated from the source locations described in the previous section. Such an ideal wavefront is characterized by constant magnitude and phase.

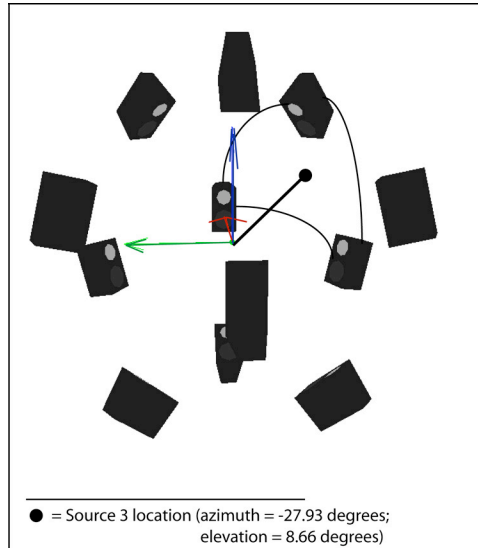


Figure 5.3: Source 3 angular direction

5.3.2 Simulation of the wavefront using the ideal dodecahedron loudspeakers configuration

In this section we study the simulation of a sound field when using the ideal dodecahedron loudspeakers configuration, where all sources are equidistant from the center of the coordinate system.

In comparison with the ideal wavefronts shown in Figures 5.4 to 5.6, one can see the wavefront functions generated using first and second order Ambisonics rendering technique in Figures 5.7, 5.8 and 5.9. The simulations are shown for the frequencies of 250, 500 and 1000 Hz.

First-order Ambisonics is accurately reconstructing the wavefront only for the 250 Hz frequency, for all source locations. Above that, we notice spatial aliasing: The wavefront is not continuously recreated anymore, instead the sound field looks like a multitude of independent wavefronts generated by individual loudspeakers.

Second-order Ambisonics rendering is able to reconstruct the wavefront fairly accurately up to 1000 Hz for both *Source 1* and *Source 2*. For *Source 3* however, which we considered as the most difficult source position to be reproduced by this particular rendering configuration, aliasing is noticed for frequencies above 250 Hz.

In Figures 5.10, 5.11 and 5.12, one views the vertical wave field behavior of

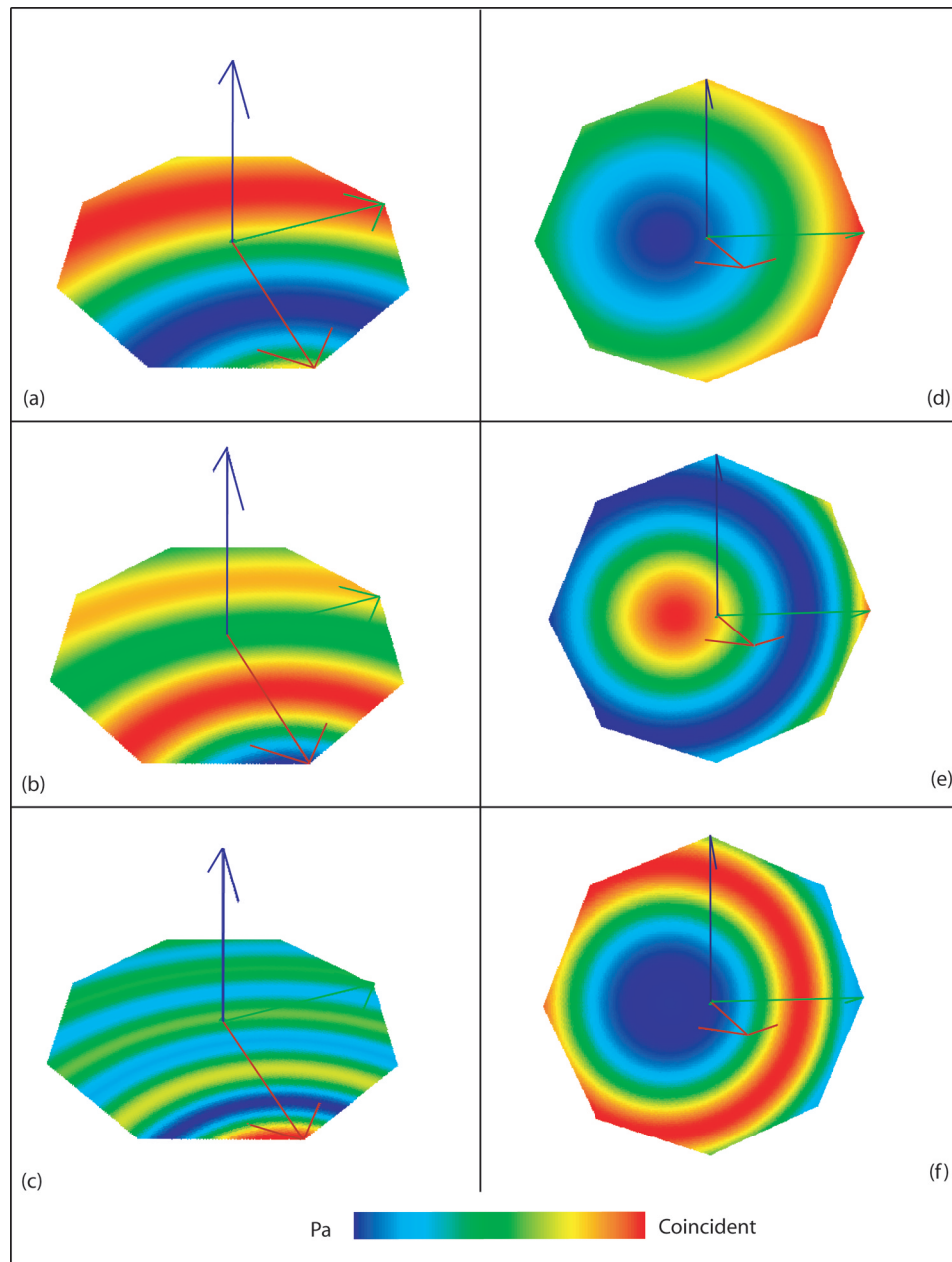


Figure 5.4: Simulation of an ideal wavefront for *Source 1*: (a), (b) and (c) 250Hz, 500 Hz and 1 kHz in the horizontal plane; (d), (e) and (f) 250Hz, 500 Hz and 1 kHz in the vertical plane.

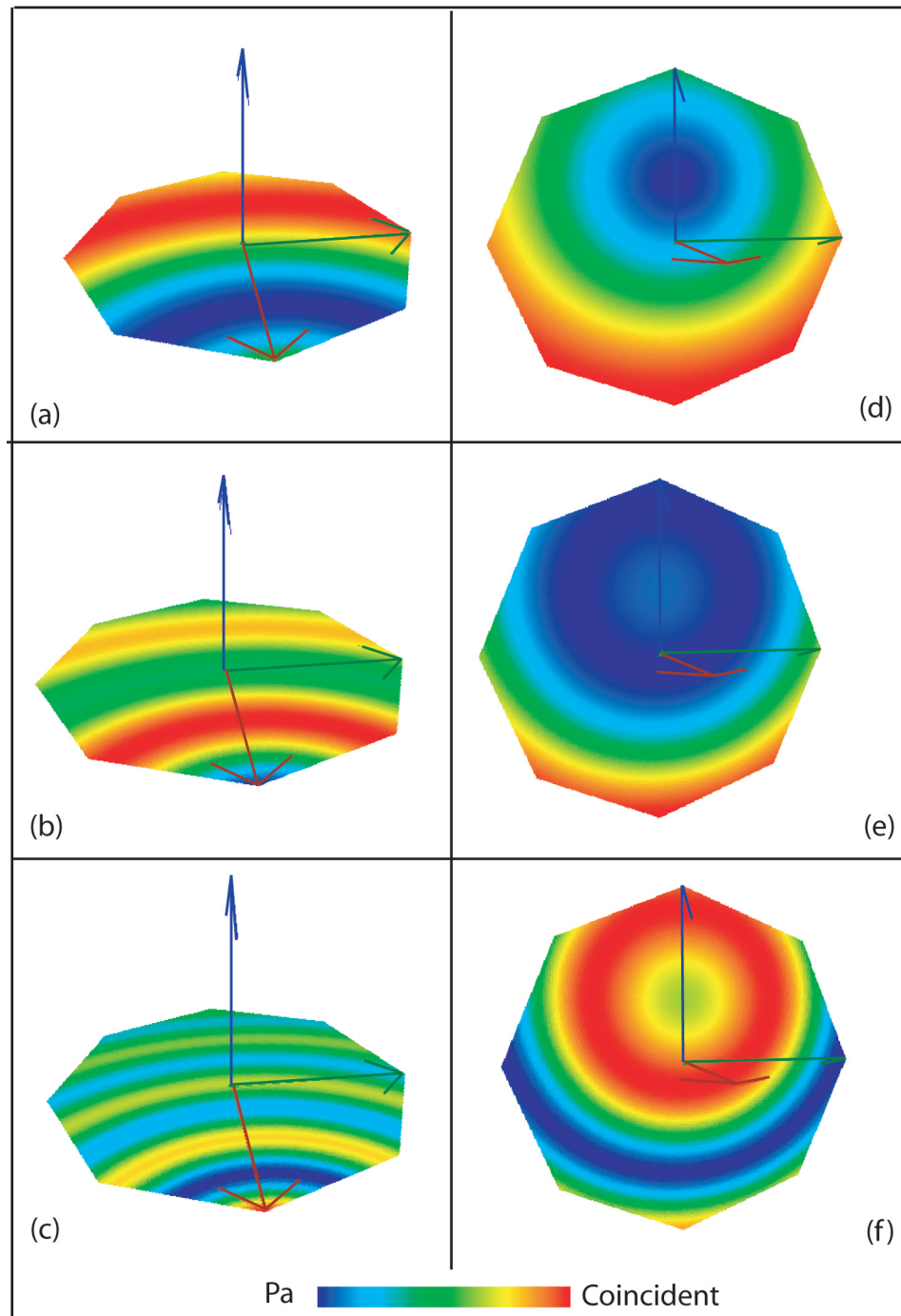


Figure 5.5: Simulation of an ideal wavefront for *Source 2*: (a), (b) and (c) 250Hz, 500 Hz and 1 kHz in the horizontal plane; (d), (e) and (f) 250Hz, 500 Hz and 1 kHz in the vertical plane.

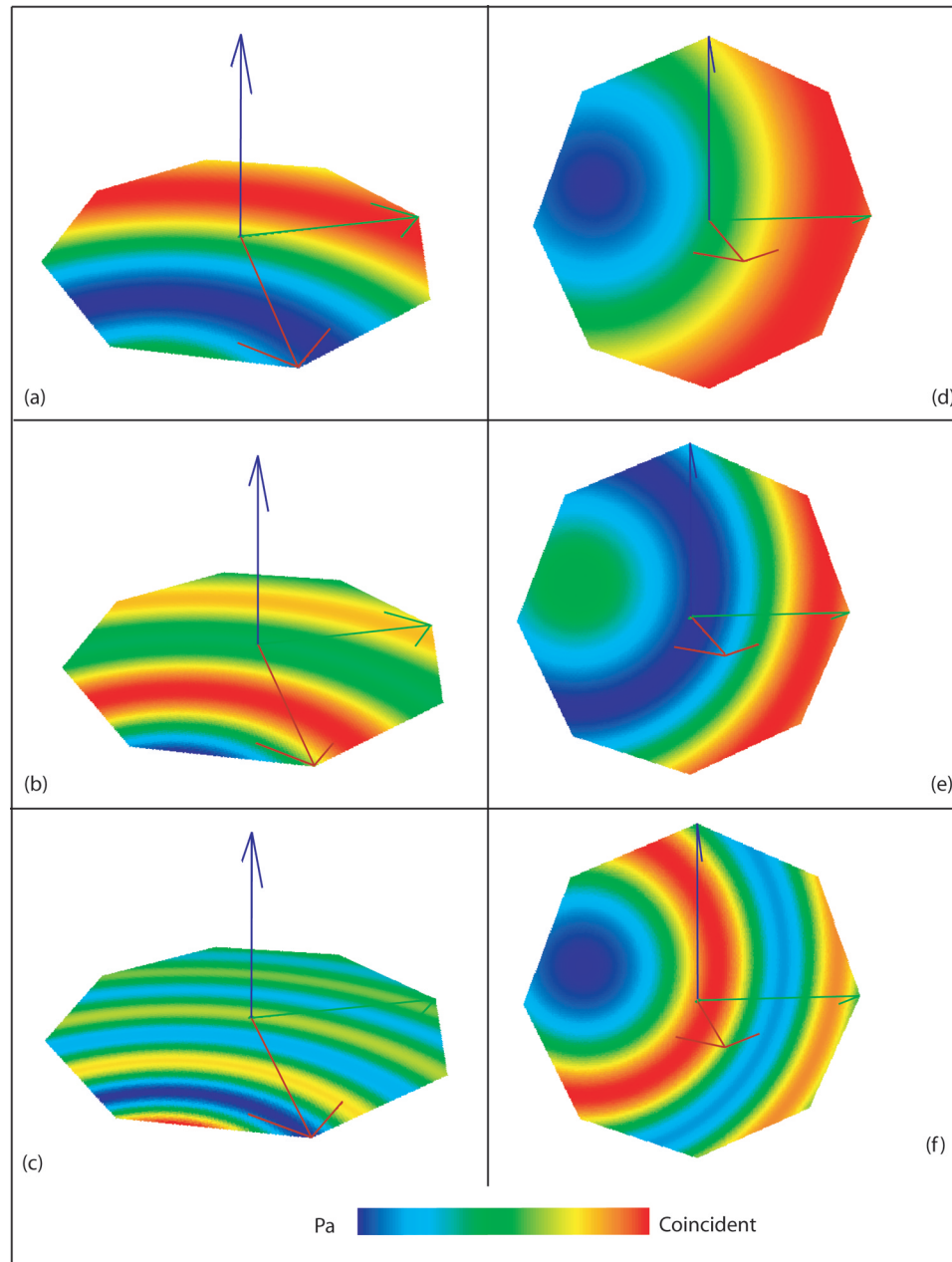


Figure 5.6: Simulation of an ideal wavefront for *Source 3*: (a), (b) and (c) 250Hz, 500 Hz and 1 kHz in the horizontal plane; (d), (e) and (f) 250Hz, 500 Hz and 1 kHz in the vertical plane.

the Ambisonics rendering. The same spatial aliasing effect is noticed with first order Ambisonics as when looking at the horizontal wave field. Aliasing also appears for second-order Ambisonics at 500 Hz and even more predominantly at 1000 Hz. The poor representation of the sound field in the vertical direction can be influenced by non-symmetry of the loudspeaker drivers in the vertical direction.

5.3.3 Simulation of the wavefront using the warped dodecahedron setup

In this section we are looking at the simulations of a sound field using the same decoding configuration as in our perceptual experiments. The decoding matrix is identical to the one used for the ideal dodecahedron configuration, but due to space limitations, we could not place the loudspeakers equidistantly from the origin of the Ambisonics coordinate system. Instead, we compensated for these differences by using appropriate delay amounts and gain levels to each individual source, and virtually recreated the ideal dodecahedron set-up. Some form of warped dodecahedron is probably the most common way one can use this decoding configuration anyway, because surrounding the listener with loudspeakers from all directions (including from below) is practically prohibitive. In order to be able to use the aforementioned configuration, modifications in terms of delay and gain compensation are required, similar to the ones implemented in our research.

As the Figures 5.13 through 5.18 show, due to the gain and delay corrections, some phase differences are encountered in the spatial wavefront. The corrections do compensate for the level and phase differences at the center point, but not necessarily over a large field; thus, the effective listening area is considerably diminished. Figures 5.13, 5.14 and 5.15 show the aliasing resulted from these modifications in the horizontal plane (above 250 Hz), with both Ambisonics rendering orders. Even worse aliasing effects can be noticed in the vertical plane, in Figures 5.16, 5.17 and 5.18.

To confirm that however, even if warped, the system can reconstruct correctly the sound field at least around the center of the coordinate system, a new series of simulations were performed, zooming into an area with a radius of approximately 30 centimeters from the center. These new simulations are showed both in the

horizontal and vertical plane in Figures 5.19 through 5.24. We wanted to confirm this way that at least at the center of the system where the listener's head was located, the wavefront was correctly reproduced. Each figure compares the ideal and warped configuration for each rendering condition. The figures show similar results between the ideal and the warped dodecahedron simulations, proving that indeed, the wavefront could be reproduced correctly, if only over a very small area. Some aliasing starts to be noticed at 1000 Hz, but this holds true for both ideal and warped conditions, excluding the possibility of being an effect of modifying the system, but rather an effect of the Ambisonics order used. From all simulations performed, over a large as well as small area we observe that the practical frequency cut-off seems to be an effect of the Ambisonics order used in the rendering. Another reason can be the loudspeaker characteristics, as described in previous sections.

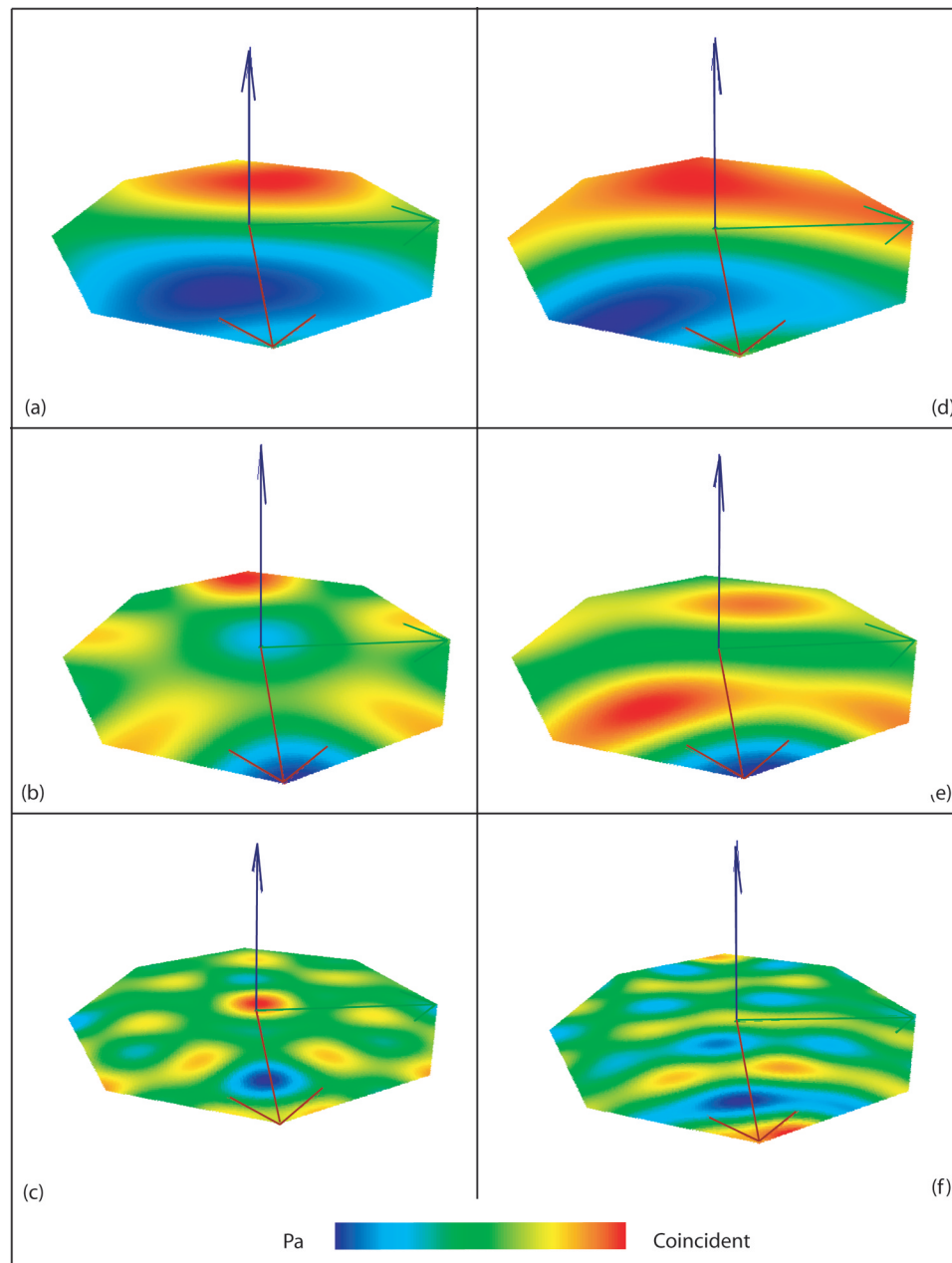


Figure 5.7: Ideal Ambisonics simulation of the wavefront in the horizontal plane for *Source 1*: (a), (b) and (c) 250Hz, 500 Hz and 1 kHz with *1st order Ambisonics*; (d), (e) and (f) 250Hz, 500 Hz and 1 kHz with *2nd order Ambisonics*.

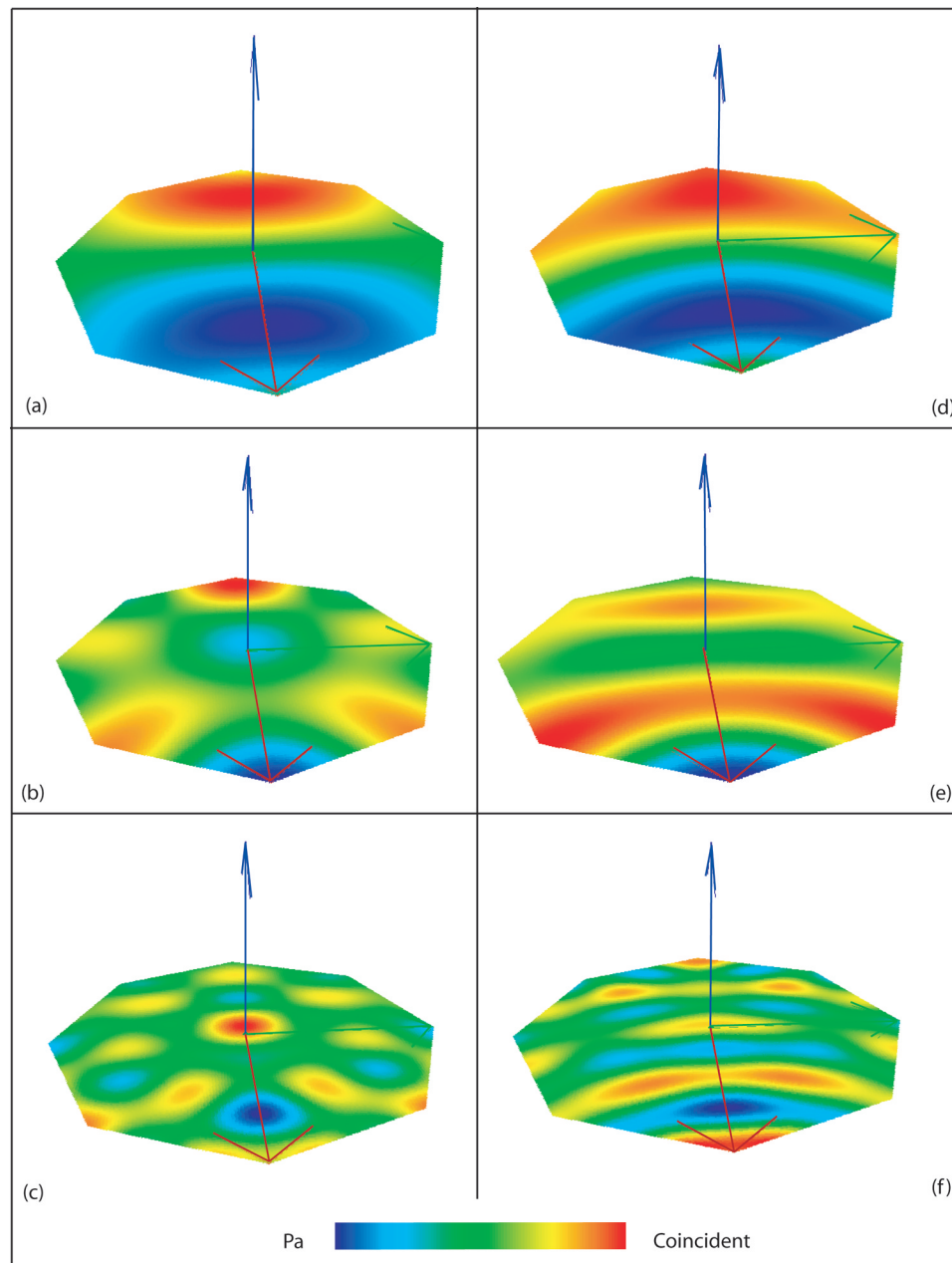


Figure 5.8: Ideal Ambisonics simulation of the wavefront in the horizontal plane for *Source 2*: (a), (b) and (c) 250Hz, 500 Hz and 1 kHz with *1st order Ambisonics*; (d), (e) and (f) 250Hz, 500 Hz and 1 kHz with *2nd order Ambisonics*.

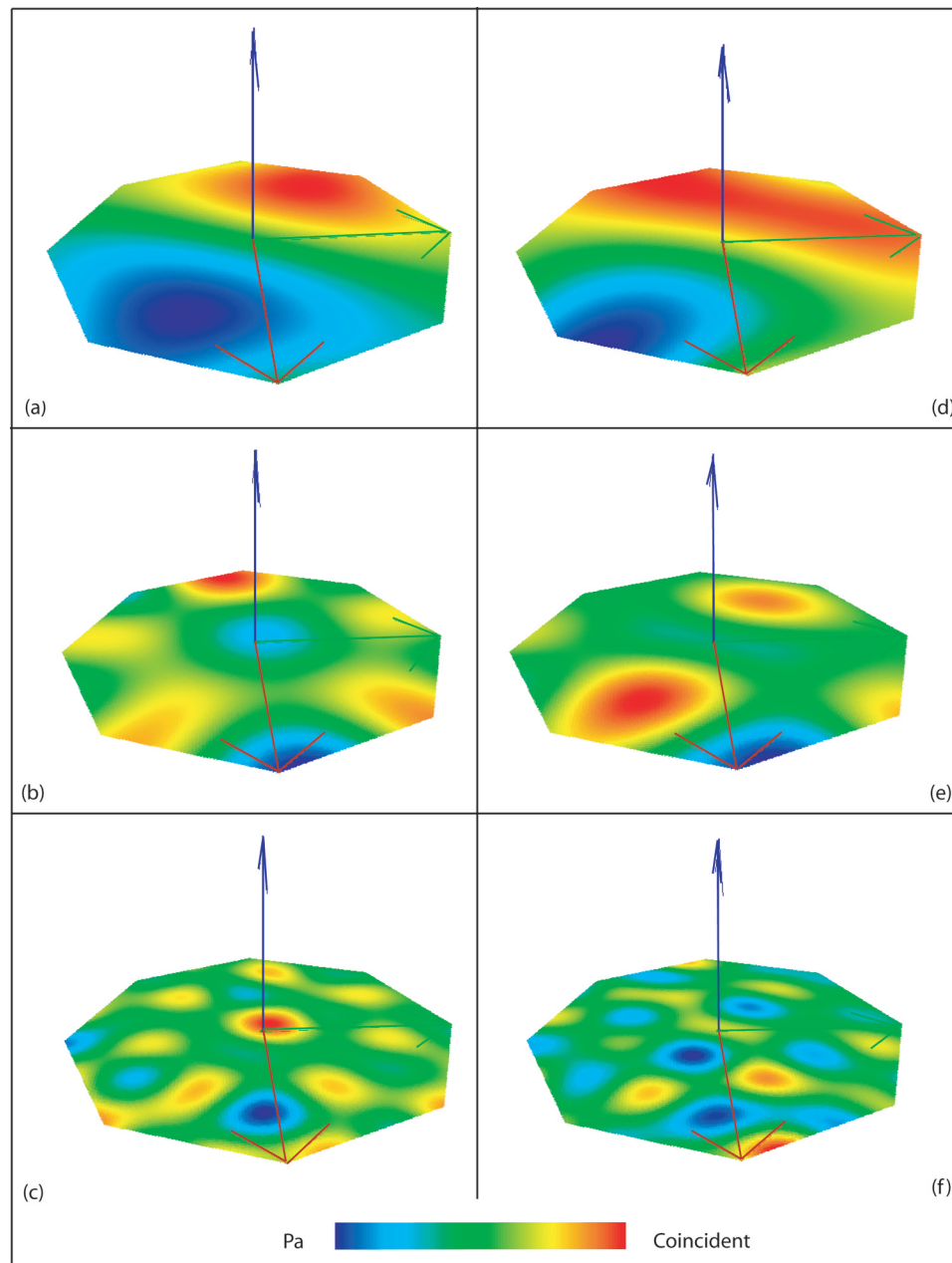


Figure 5.9: Ideal Ambisonics simulation of the wavefront in the horizontal plane for *Source 3*: (a), (b) and (c) 250Hz, 500 Hz and 1 kHz with *1st order Ambisonics*; (d), (e) and (f) 250Hz, 500 Hz and 1 kHz with *2nd order Ambisonics*.

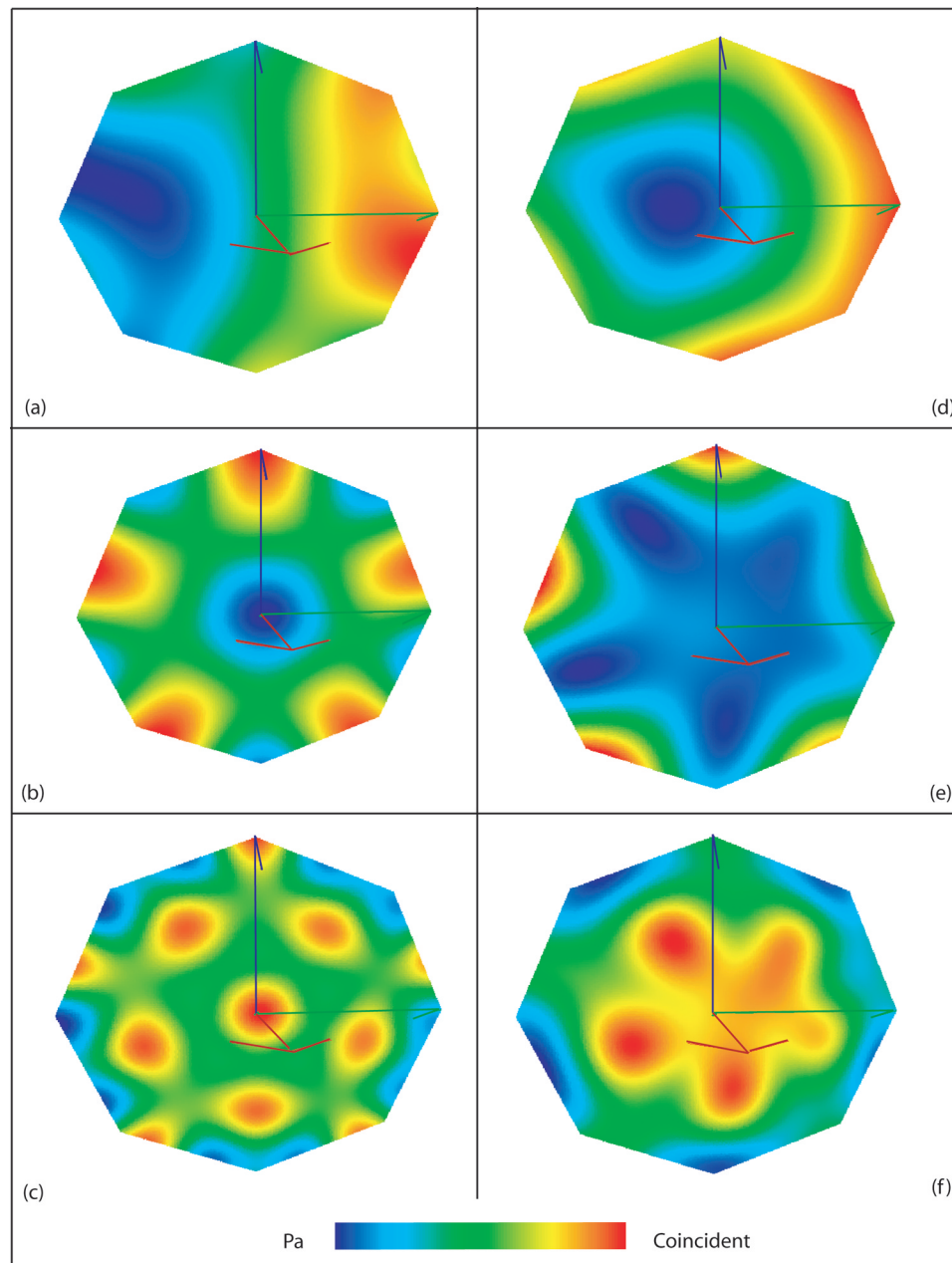


Figure 5.10: Ideal Ambisonics simulation of the wavefront in the vertical plane for *Source 1*: (a), (b) and (c) 250Hz, 500 Hz and 1 kHz with *1st order Ambisonics*; (d), (e) and (f) 250Hz, 500 Hz and 1 kHz with *2nd order Ambisonics*.

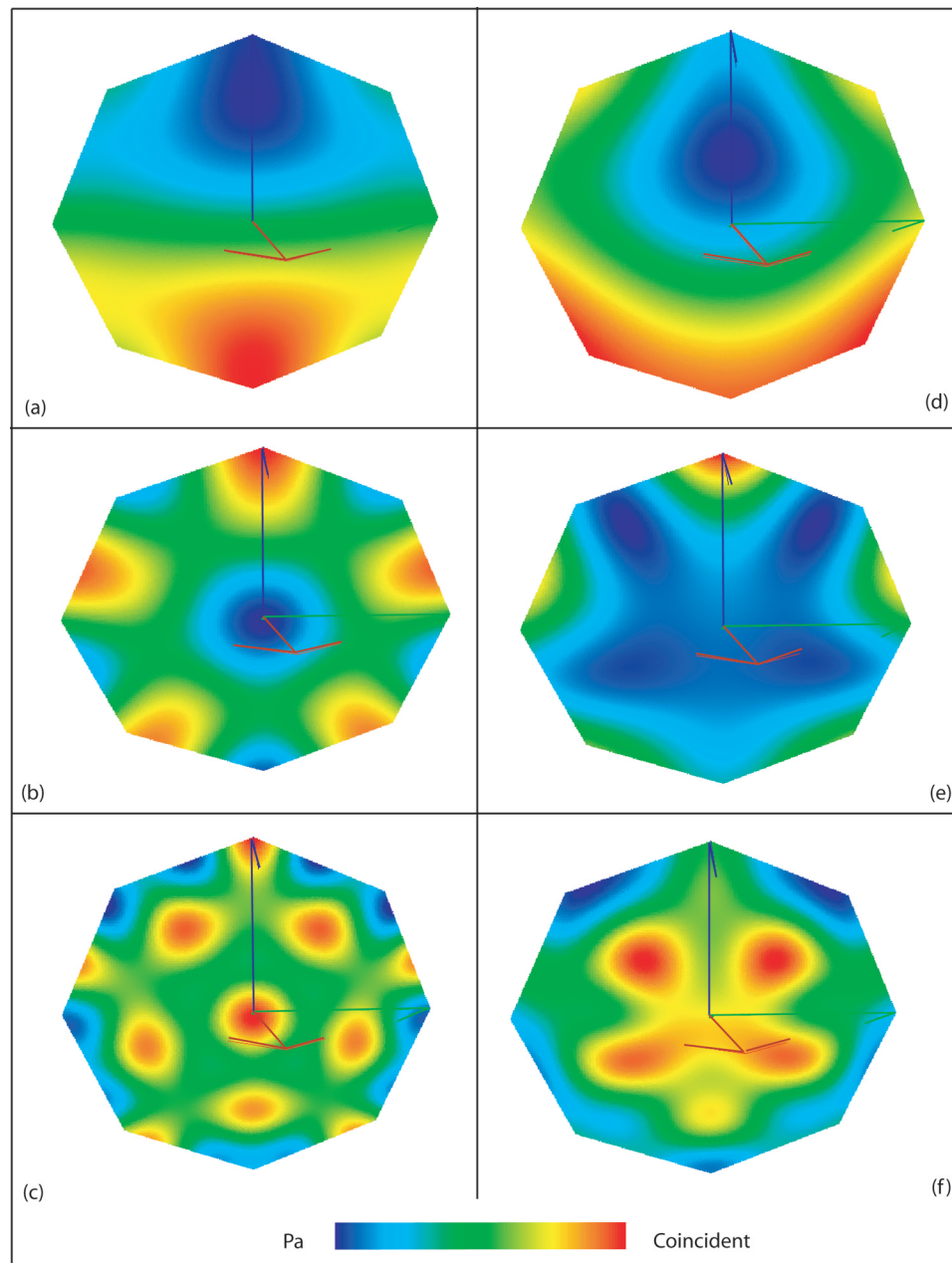


Figure 5.11: Ideal Ambisonics simulation of the wavefront in the vertical plane for *Source 2*: (a), (b) and (c) 250Hz, 500 Hz and 1 kHz with *1st order Ambisonics*; (d), (e) and (f) 250Hz, 500 Hz and 1 kHz with *2nd order Ambisonics*.

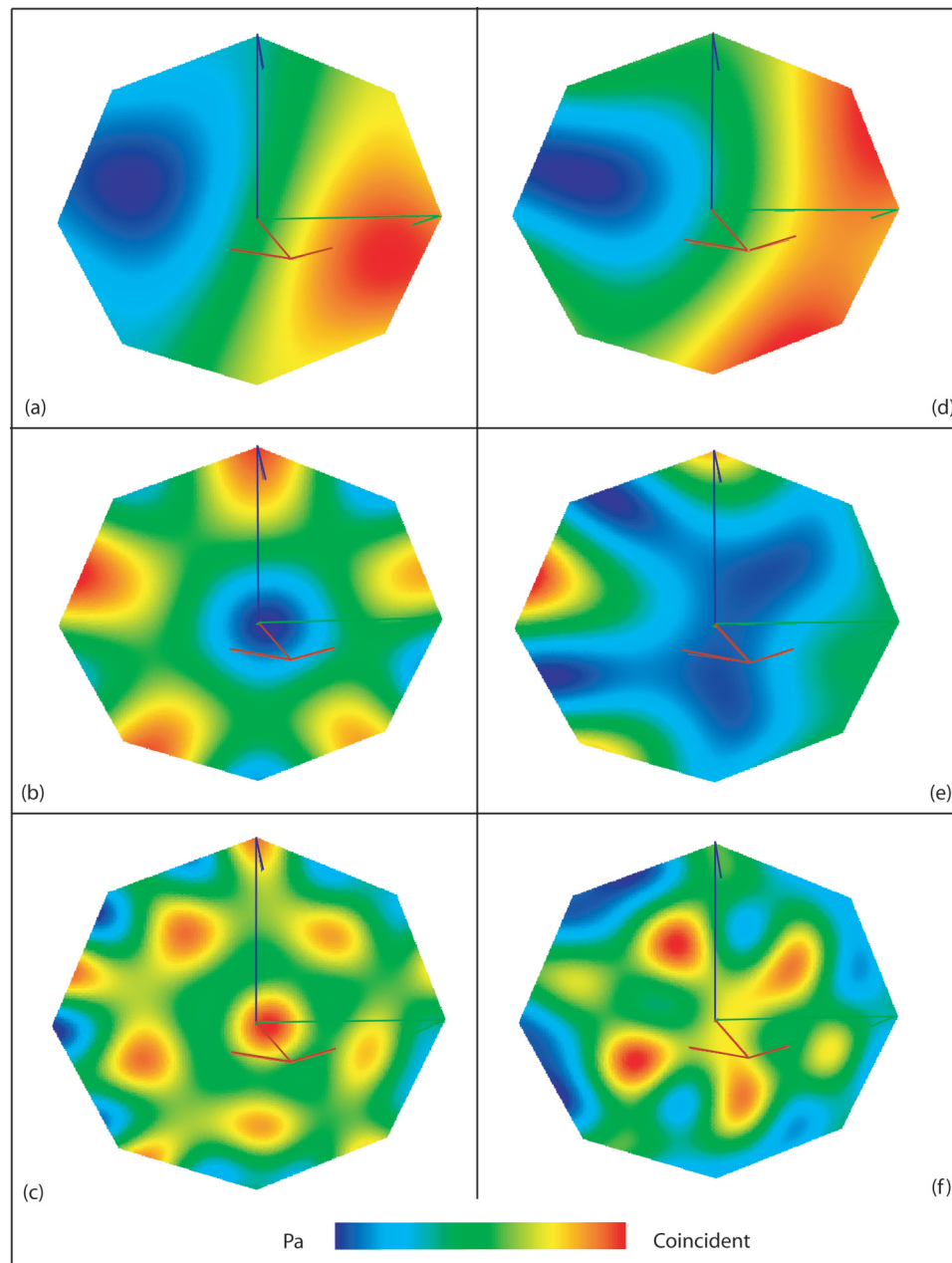


Figure 5.12: Ideal Ambisonics simulation of the wavefront in the vertical plane for *Source 3*: (a), (b) and (c) 250Hz, 500 Hz and 1 kHz with *1st order Ambisonics*; (d), (e) and (f) 250Hz, 500 Hz and 1 kHz with *2nd order Ambisonics*.

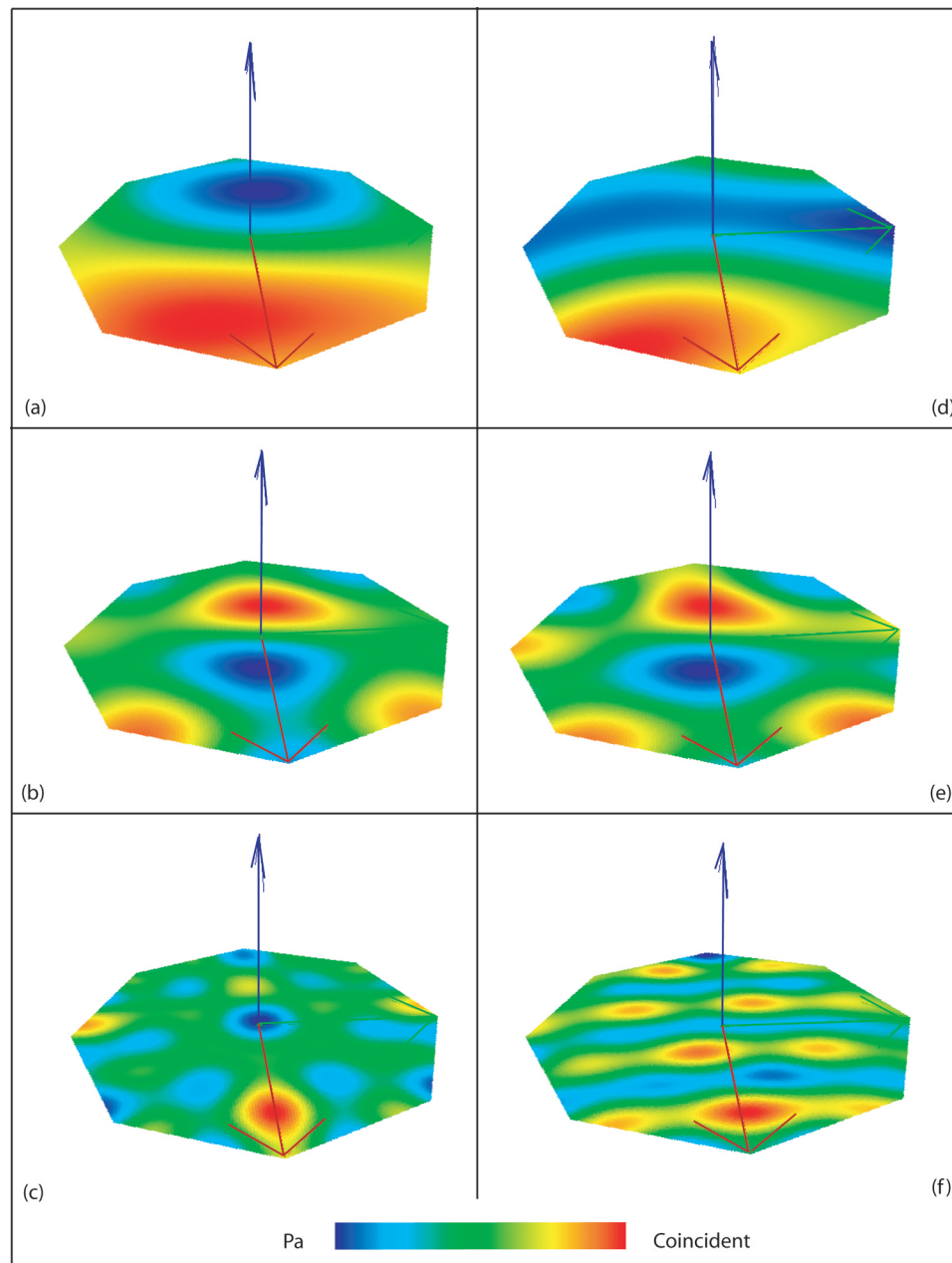


Figure 5.13: Warped Ambisonics simulation of the wavefront in the horizontal plane for *Source 1*: (a), (b) and (c) 250Hz, 500 Hz and 1 kHz with *1st order Ambisonics*; (d), (e) and (f) 250Hz, 500 Hz and 1 kHz with *2nd order Ambisonics*.

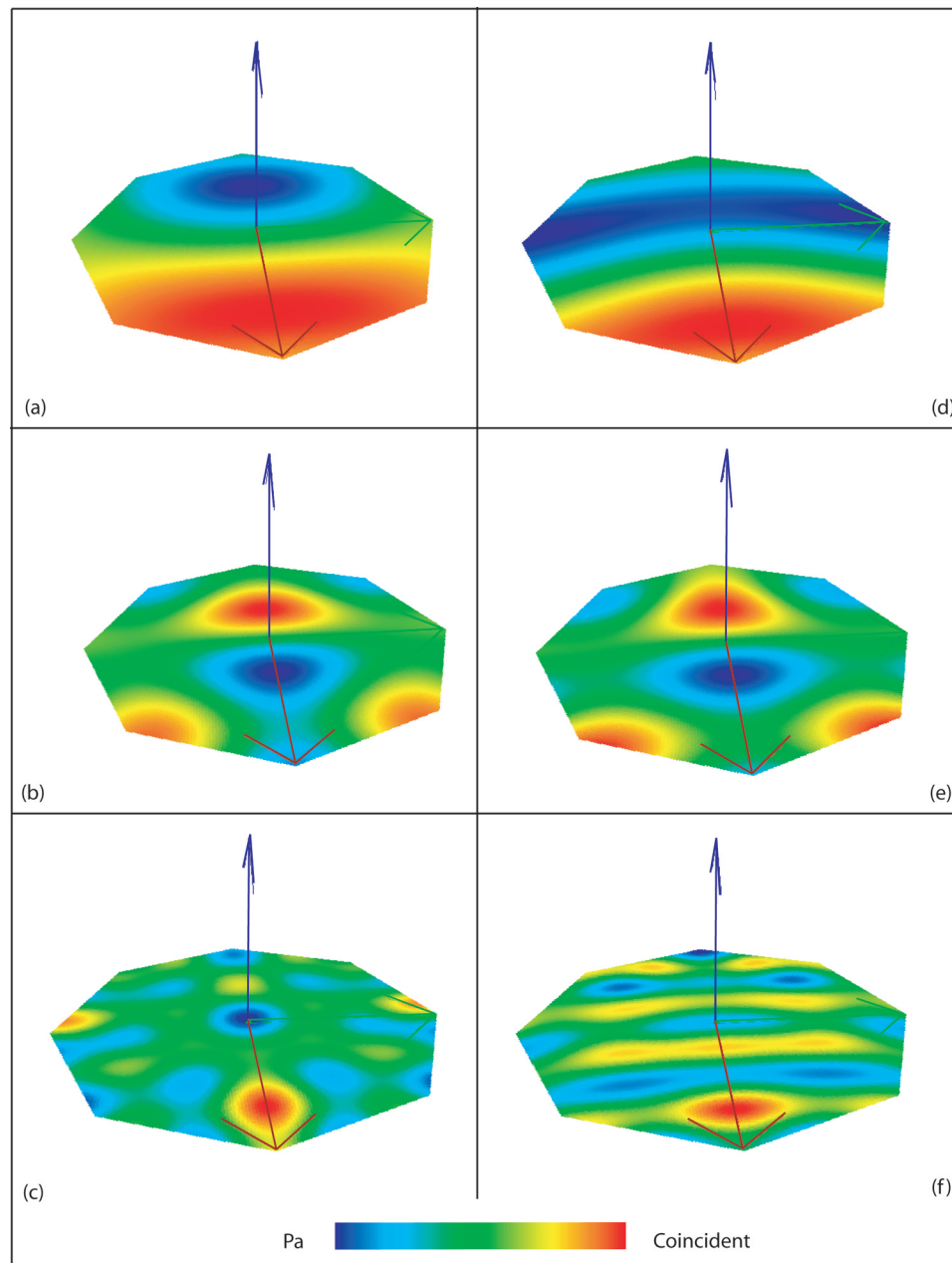


Figure 5.14: Warped Ambisonics simulation of the wavefront in the horizontal plane for *Source 2*: (a), (b) and (c) 250Hz, 500 Hz and 1 kHz with *1st order Ambisonics*; (d), (e) and (f) 250Hz, 500 Hz and 1 kHz with *2nd order Ambisonics*.

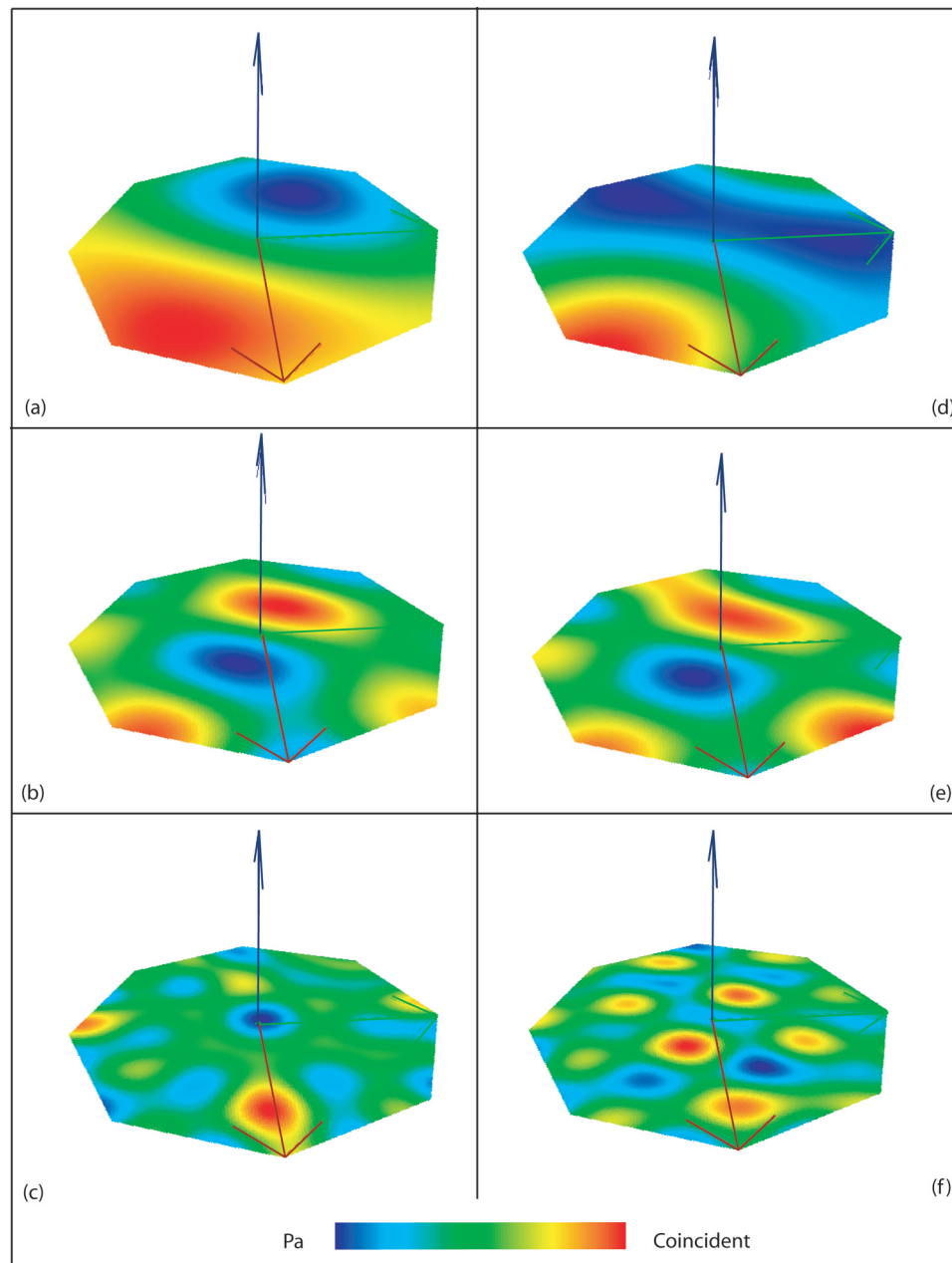


Figure 5.15: Warped Ambisonics simulation of the wavefront in the horizontal plane for *Source 3*: (a), (b) and (c) 250Hz, 500 Hz and 1 kHz with *1st order Ambisonics*; (d), (e) and (f) 250Hz, 500 Hz and 1 kHz with *2nd order Ambisonics*.

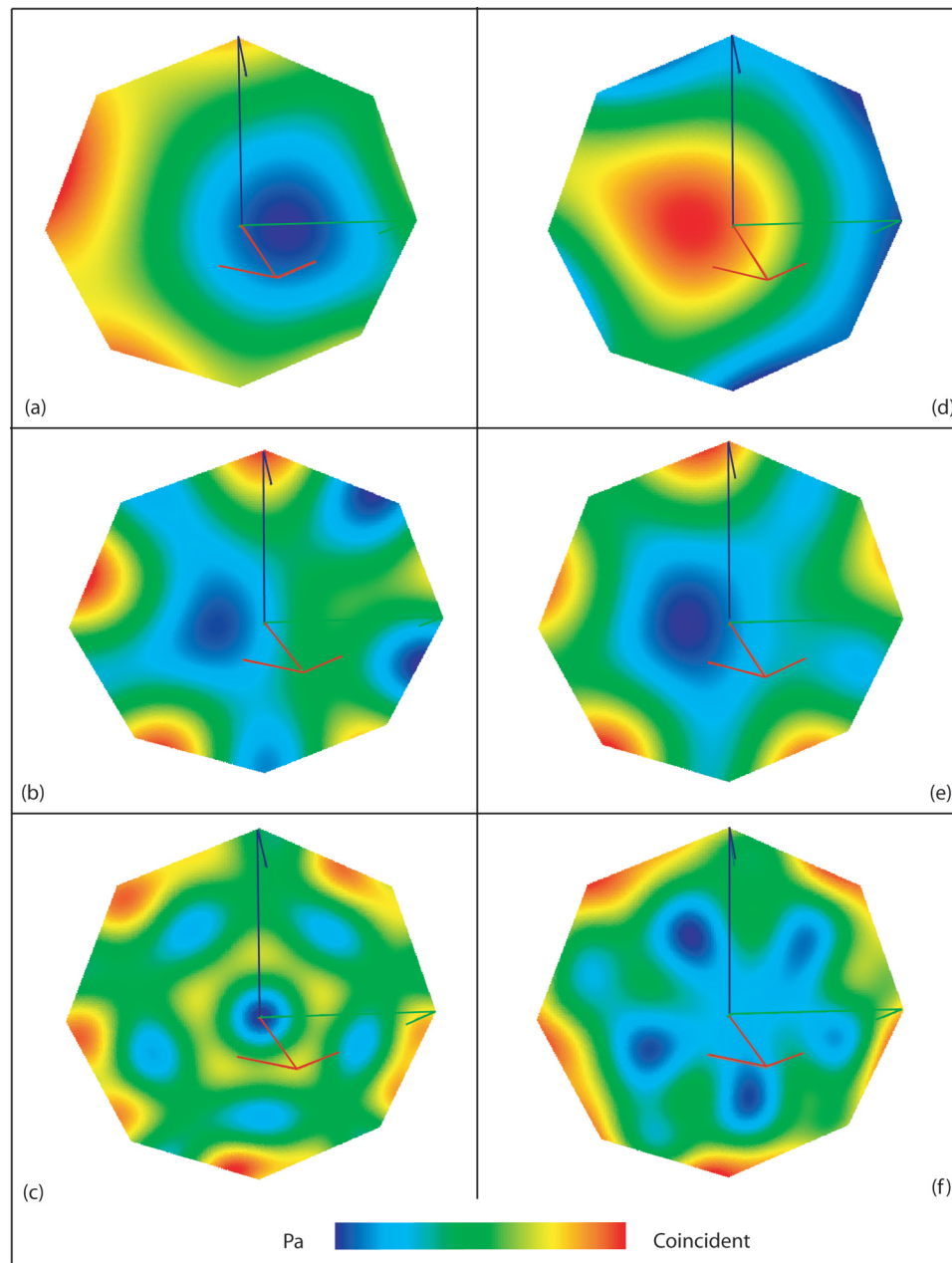


Figure 5.16: Warped Ambisonics simulation of the wavefront in the vertical plane for *Source 1*: (a), (b) and (c) 250Hz, 500 Hz and 1 kHz with *1st order Ambisonics*; (d), (e) and (f) 250Hz, 500 Hz and 1 kHz *2nd order Ambisonics*.

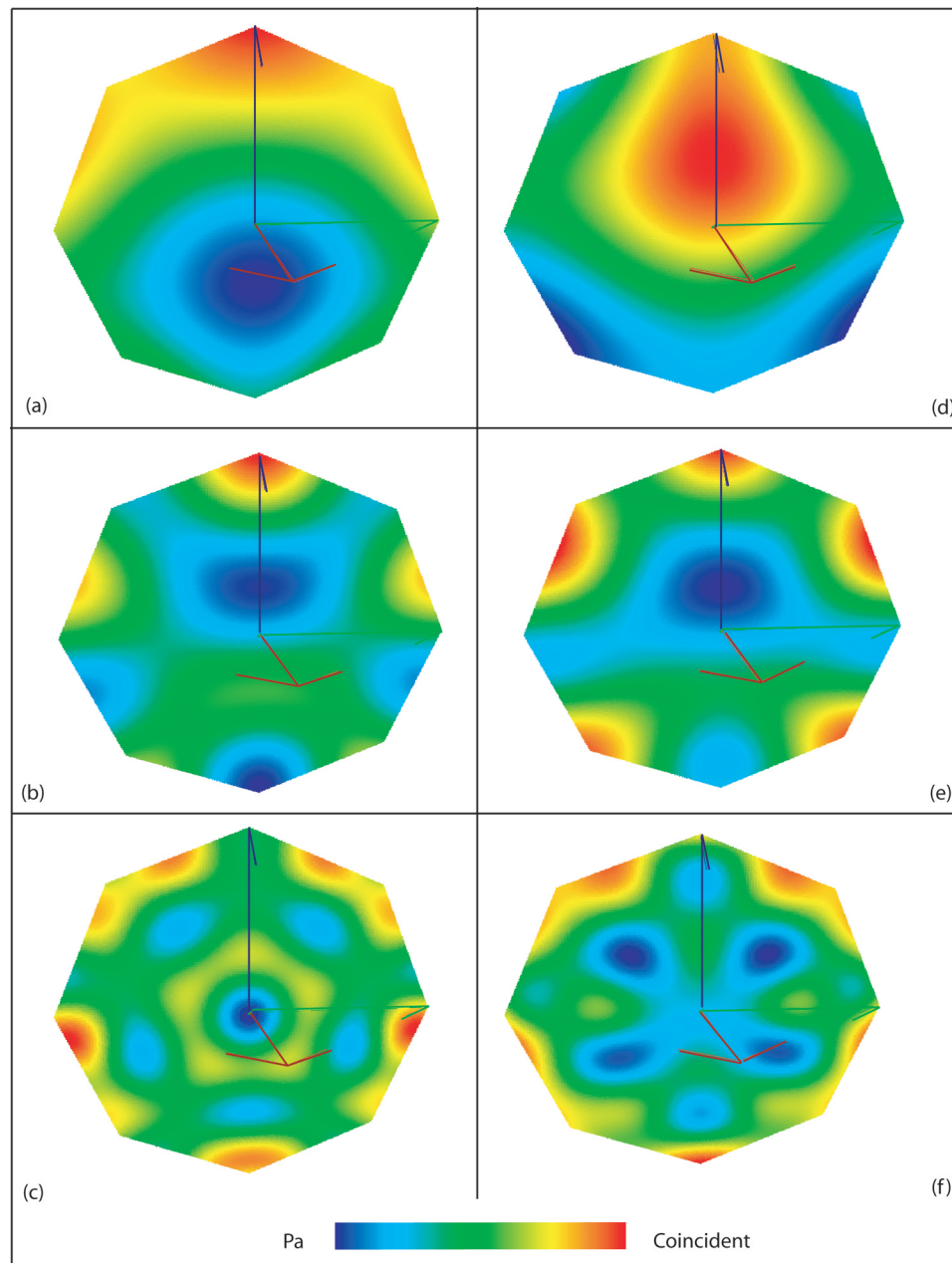


Figure 5.17: Warped Ambisonics simulation of the wavefront in the vertical plane for *Source 2*: (a), (b) and (c) 250Hz, 500 Hz and 1 kHz with *1st order Ambisonics*; (d), (e) and (f) 250Hz, 500 Hz and 1 kHz with *2nd order Ambisonics*.

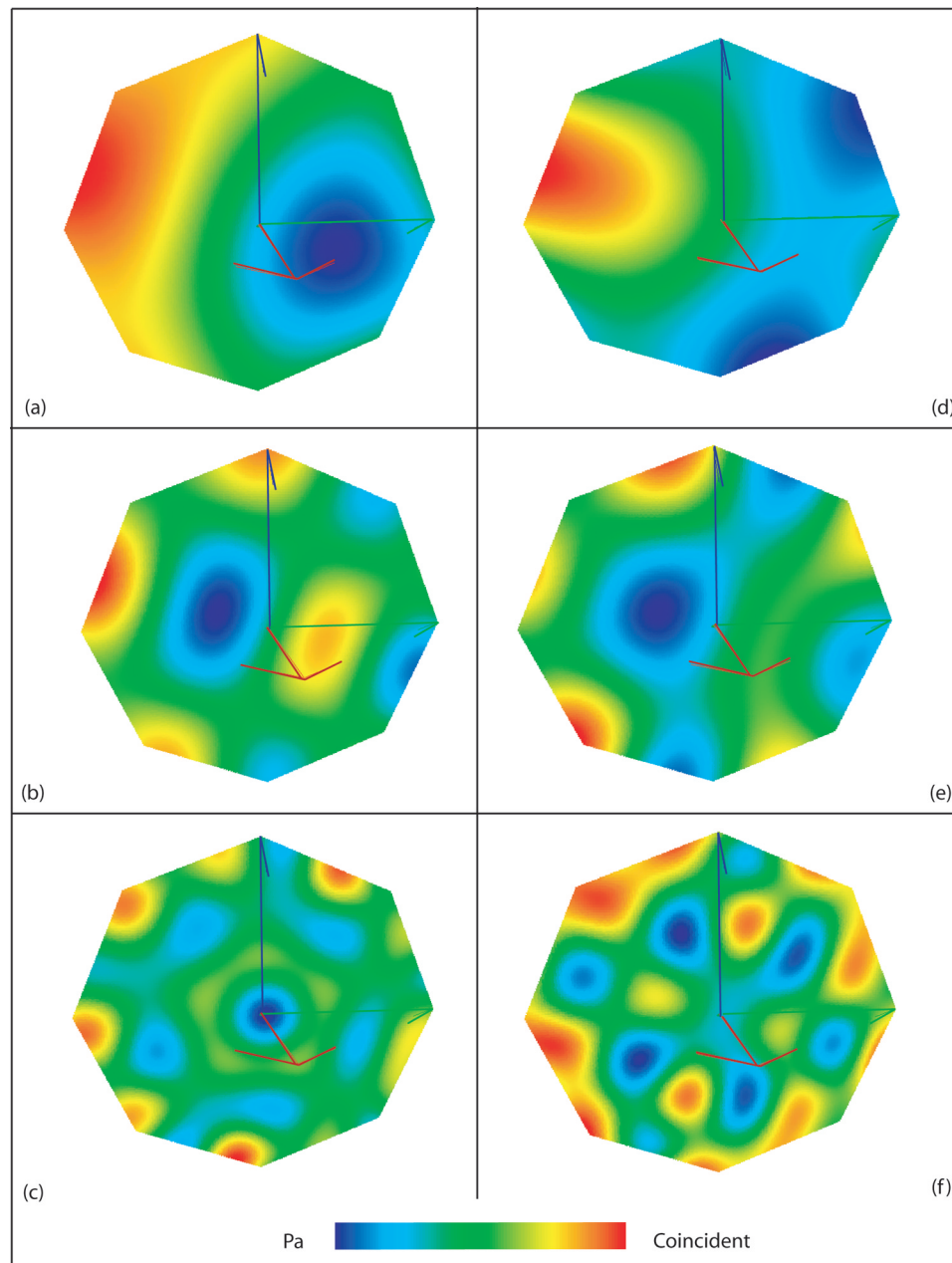


Figure 5.18: Warped Ambisonics simulation of the wavefront in the vertical plane for *Source 3*: (a), (b) and (c) 250Hz, 500 Hz and 1 kHz with *1st order Ambisonics*; (d), (e) and (f) 250Hz, 500 Hz and 1 kHz with *2nd order Ambisonics*.

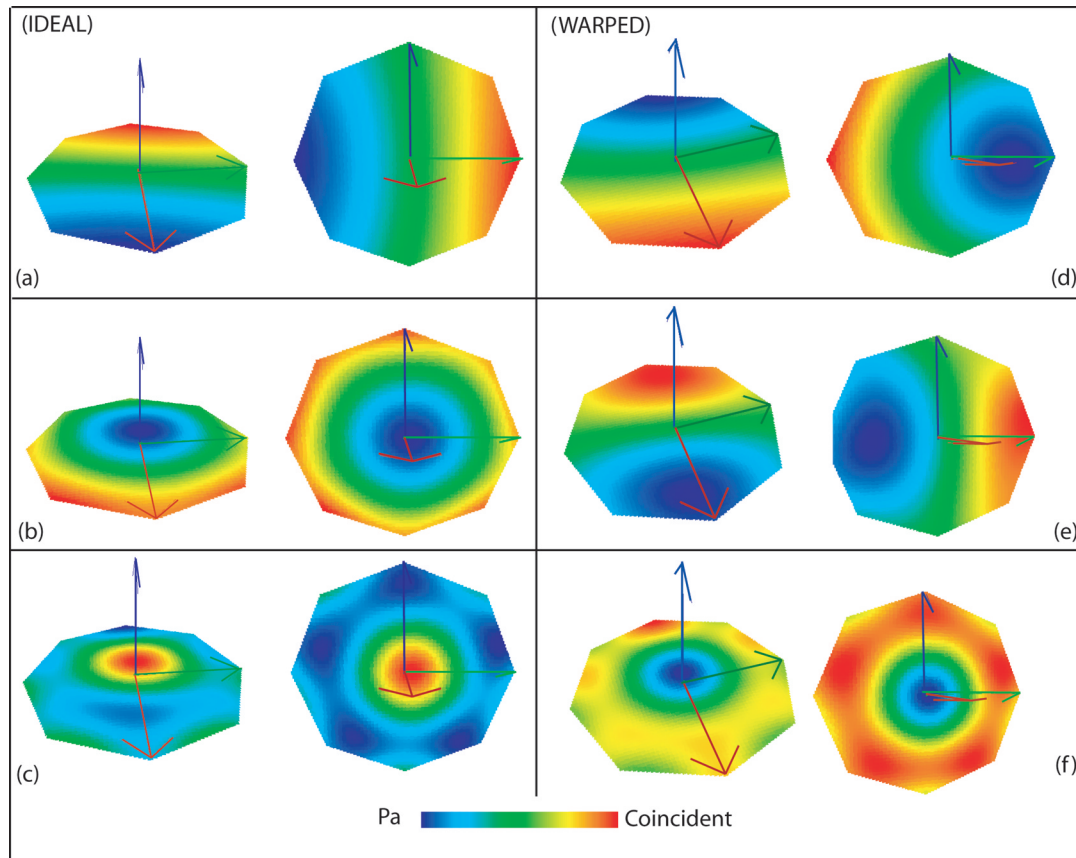


Figure 5.19: First-order Ambisonics, magnified simulation comparison for *Source 1*: (a), (b) and (c) 250Hz, 500 Hz and 1 kHz with an ideal configuration; (d), (e) and (f) 250Hz, 500 Hz and 1 kHz with a warped configuration.

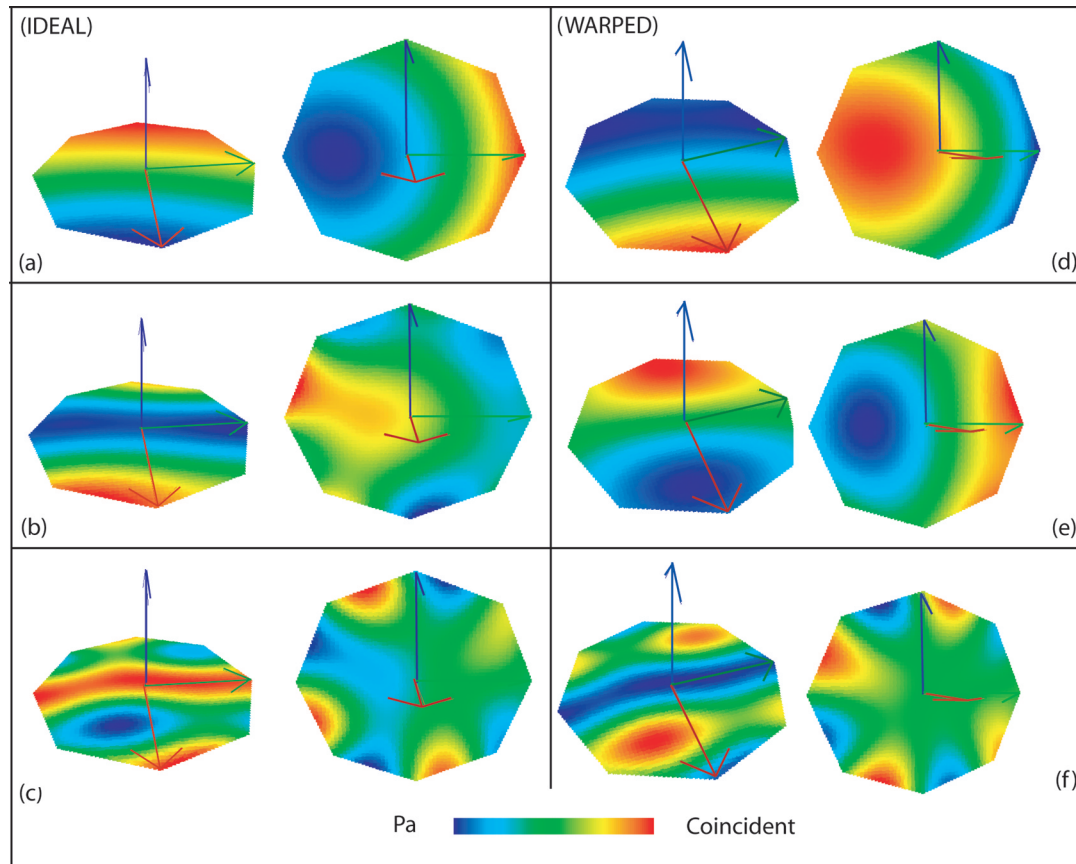


Figure 5.20: Second-order Ambisonics, magnified simulation comparison for *Source 1*: (a), (b) and (c) 250Hz, 500 Hz and 1 kHz with an ideal configuration; (d), (e) and (f) 250Hz, 500 Hz and 1 kHz with a warped configuration.

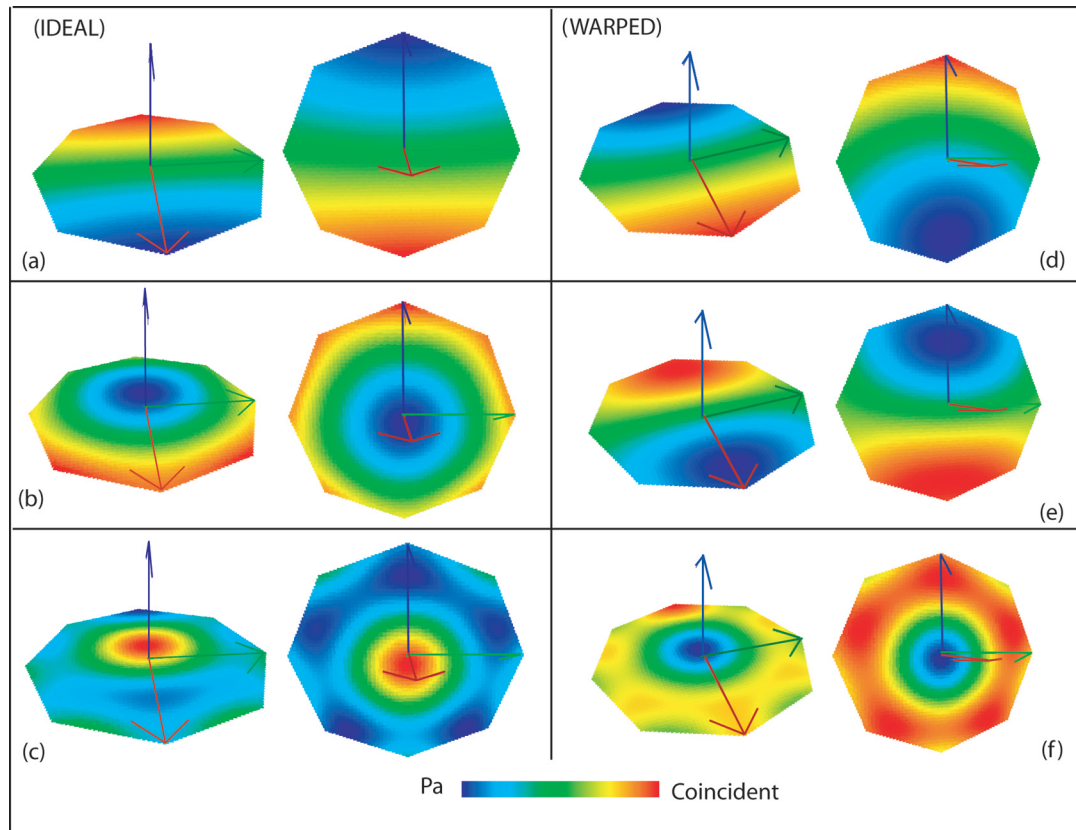


Figure 5.21: First-order Ambisonics, magnified simulation comparison for *Source 2*: (a), (b) and (c) 250Hz, 500 Hz and 1 kHz with an ideal configuration; (d), (e) and (f) 250Hz, 500 Hz and 1 kHz with a warped configuration.

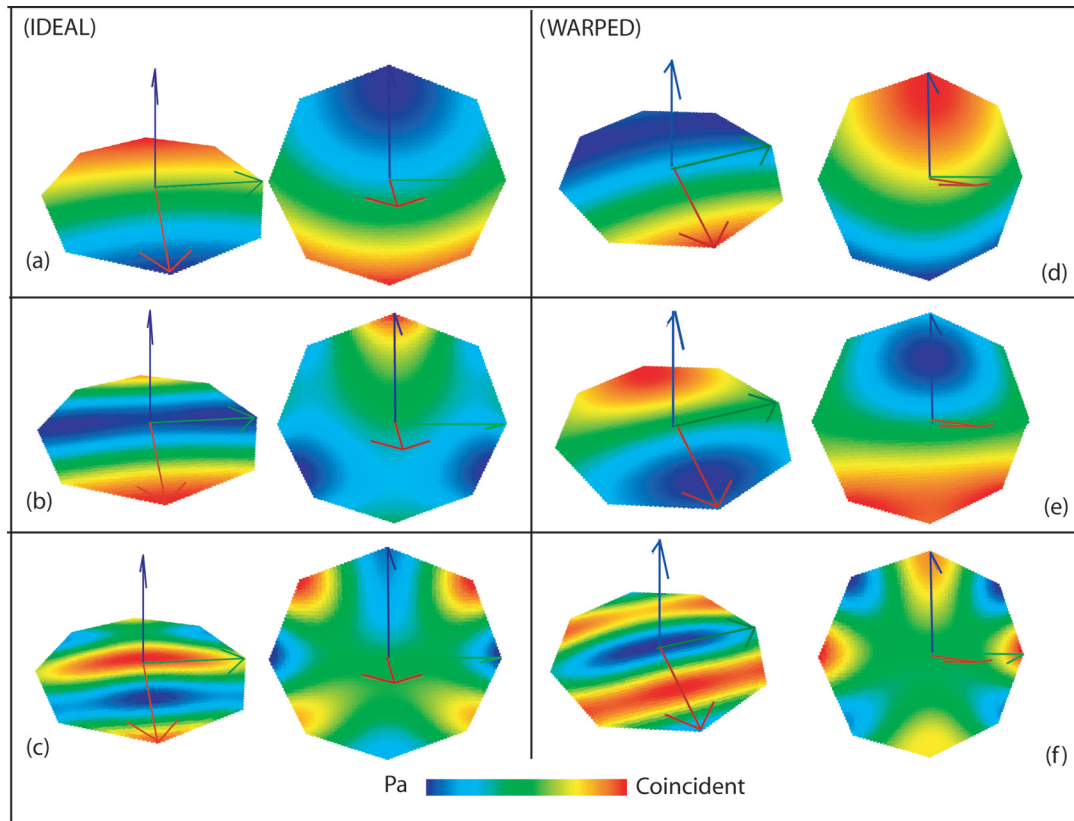


Figure 5.22: Second-order Ambisonics, magnified simulation comparison for *Source 2*: (a), (b) and (c) 250Hz, 500 Hz and 1 kHz with an ideal configuration; (d), (e) and (f) 250Hz, 500 Hz and 1 kHz with a warped configuration.

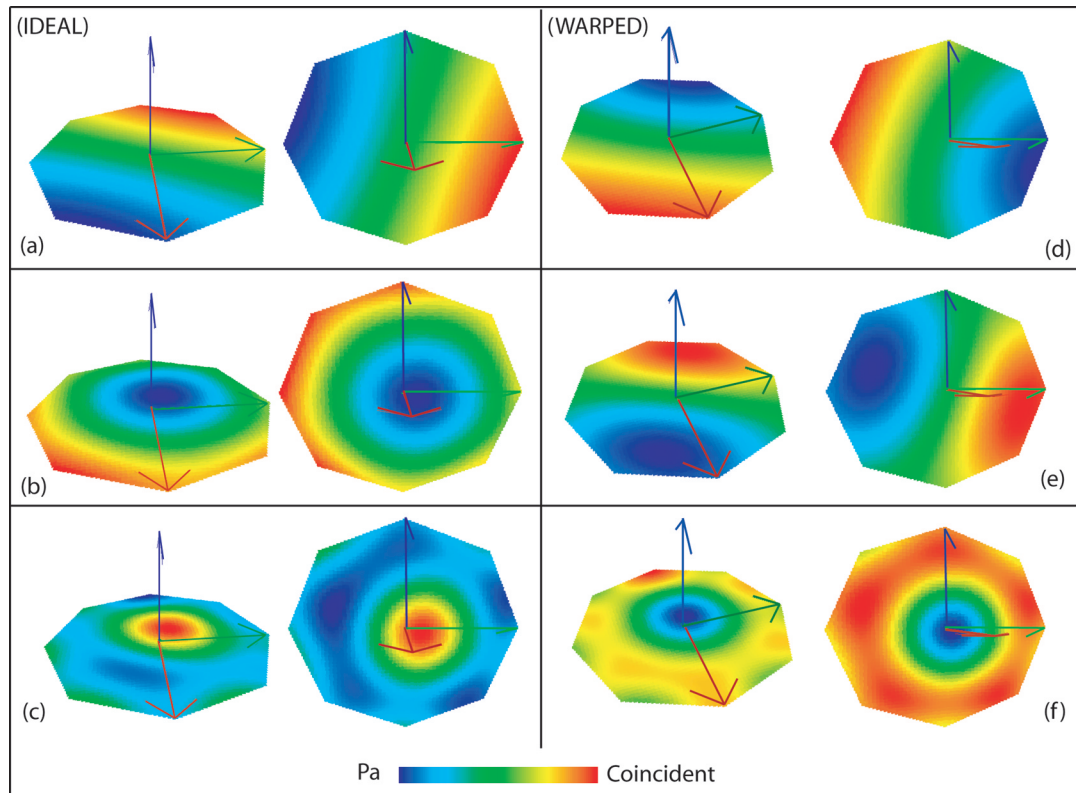


Figure 5.23: First-order Ambisonics, magnified simulation comparison for *Source 3*: (a), (b) and (c) 250Hz, 500 Hz and 1 kHz with an ideal configuration; (d), (e) and (f) 250Hz, 500 Hz and 1 kHz with a warped configuration.

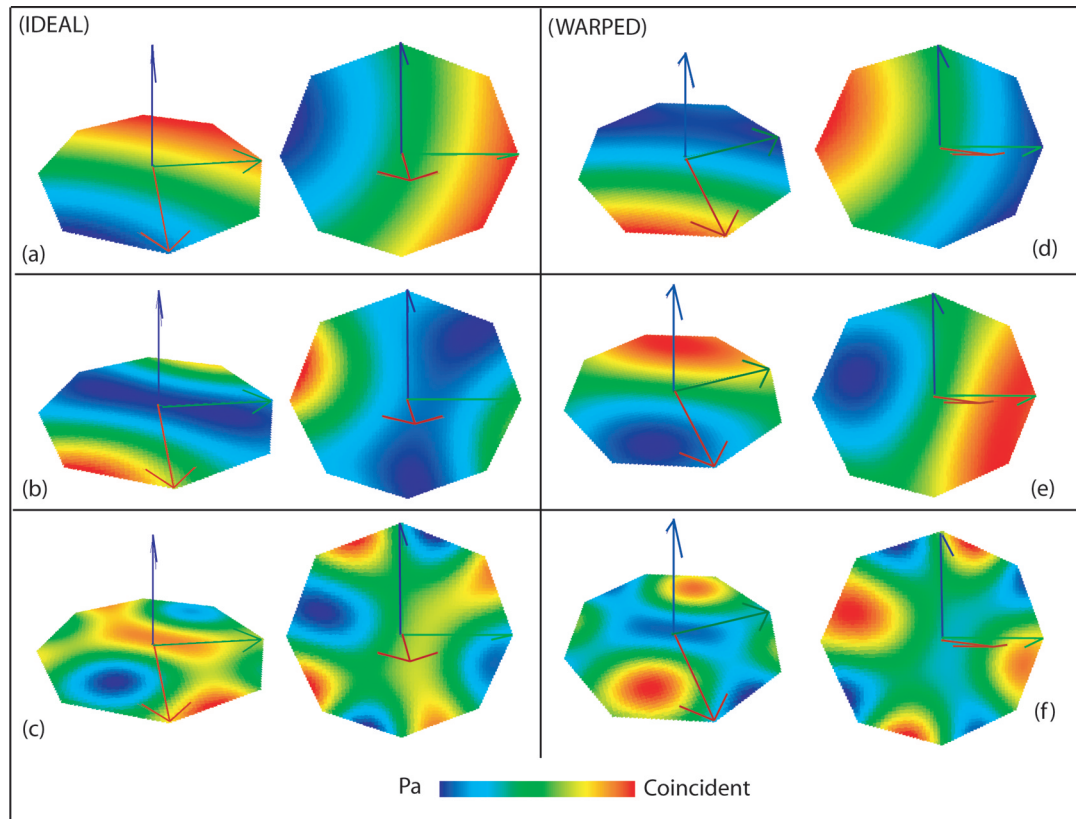


Figure 5.24: Second-order Ambisonics, magnified simulation comparison for *Source 3*: (a), (b) and (c) 250Hz, 500 Hz and 1 kHz with an ideal configuration; (d), (e) and (f) 250Hz, 500 Hz and 1 kHz with a warped configuration.

CHAPTER 6

CONCLUSIONS

This chapter presents the conclusions drawn from the results of the research developed and presented previously. We will summarize both the perceptual investigations and the wavefront simulations findings, and discuss to what extent the Ambisonics sound rendering system used in this research is a viable solution for practical implementation in a virtual environment, as initially discussed.

6.1 Research Results

6.1.1 System Implementation

To review, in this research we constructed and proposed to study a sound rendering system to be implemented in a virtual environment for remote communication. The system chosen was a surround sound reproduction system, based on low-order Ambisonics periphonic encoding; the decoding was configured as a tilted dodecahedron, which was physically modified to meet our testing environment characteristics. Due to our rectangular listening space, the loudspeakers could not be located equidistantly from the system's origin, forcing us to implement instead a modified version of the dodecahedron. A main reason for eventually opting for this "warped" configuration in our research, is that in the real world a sound system that can spherically surround the listeners is often impractical; most rooms are built to be rectangular and the system components that convey the vertical information are to be placed on the floor and ceiling, making it difficult in practice to achieve an equal distance between these sound sources and the listening position.

In order to determine the suitability of the system if implemented in a virtual environment, we carried out perceptual experiments in the form of localization tests, using human subjects, as well as studied the physical characteristics of the system by performing sound field simulations. For the perceptual tests, we designed a real time Ambisonics encoding and decoding protocol, which was complemented in our particular case by a custom developed digital signal processing matrix (in the form

of level gains and delays) implemented to compensate for the physical irregularities of our listening environment. For the sound field simulations, we studied the system both in its ideal decoding configuration as a tilted dodecahedron, as well as in its warped version as it was implemented in our perceptual experiments.

6.1.2 Perceptual Investigation

This research investigated in part the perceptual attributes of low-order Ambisonics rendering systems, consisting of localization tests distributed to human subjects. The experiment was designed to determine the localization accuracy in the horizontal and median planes with both anechoic and reverberant stimuli, when using first- and second-order Ambisonics. All the testing was performed with each subject located at the origin of the Ambisonics coordinate system. Stimuli consisted of pink-noise bursts.

According to the results reviewed below, we conclude that at least from a perceptual point of view, there is not a significant difference in localization between first and second order Ambisonics rendering. However, this statement is valid only if the listener is located at the origin of the system. It would be interesting to investigate in further research the localization accuracy in similar conditions, but with the listener located off-center.

Regarding the free field vs. reverberant field localization, the average absolute error values over the entire subject population show almost no difference in the localization blur between the two categories. However, if we look at the results for the subject with best performance and the only consistent responses, there is a noticeable difference between the two categories in the horizontal plane, showing a doubling in the localization error for the reverberant field, compared to the free-field results.

6.1.2.1 Localization in the Horizontal Plane

This section overviews the results of the perceptual localization accuracy specific to the rendering system for stimuli distributed in the horizontal plane. The subjects were played stimuli located between 0° and -30° on the y-axis of the Ambisonics coordinate system, which translates into an azimuth between 0° and 30° to

the right of the listener.

One can see in Chapter 4 that for the first-order Ambisonics system there is a mean absolute error of 9.4° for both free and reverberant field simulations, over the entire subject population. The mean absolute errors specific to the second-order Ambisonics rendering are 10.3° for the free field simulations and 10.2° for the reverberant field simulations. These numbers appear to show that there is no difference in localization between free field and reverberant field rendering conditions. One also observes a difference of approximately one degree of error between the two Ambisonics orders; however, this may be due to the auditory system's characteristics and not to the rendering system. Similar correlations for all four conditions in the horizontal plane show both consistency in the relationship among the categories, and relatively reliable results as evidenced by the coefficients of determination.

However, if we examine the localization results for the subject with best performance in Chapter 4), mean absolute localization errors of about 3.0° are observed for the free field simulations with first- and second-order Ambisonics, while values of $6^\circ - 7^\circ$ characterize the reverberant field simulations with first- and second-order Ambisonics respectively. Comparing with the baseline localization experiment described in Section 2.6, whose results show a mean absolute localization error of about $2^\circ - 3^\circ$, one can conclude that these results are very similar, and thus consider that in this case both first- and second-order Ambisonics perform reasonably well, especially in the free field rendering.

6.1.2.2 Localization in the Median Plane

We review in this section the results of the perceptual localization accuracy specific to the rendering system, for stimuli distributed in the median plane. The subjects were played stimuli located in the median plane at elevation angles between 0° and 20° , with a 0° azimuth.

The mean absolute errors show values of 10.7° and 11.0° for free and reverberant field simulations with first-order Ambisonics, and of 12.0° and 11.6° for free and reverberant field simulations with second-order Ambisonics. However, as seen in Table 4.21, there is no correlation between the stimuli locations and the subjects'

responses in the median plane, indicating that the results were primarily based on guessing. Very similar values characterized the localization blur for the subject with best-performance. This indicates that poor localization in the median plane was not due to the lack of training or experience.

In Section 2.5, one notices that the localization blur in the median plane can vary considerably depending on the type of stimulus, reaching values between 4° and 17° . With this in mind, we can possibly assign in part the poor results of our experiment to the human hearing mechanism characteristics. In addition, other reasons that most likely influenced the localization results in the median plane are floor reflections (as tests took place in a hemi-anechoic environment), scattering due to sizable loudspeakers, and vertical asymmetry in the loudspeaker drivers.

6.1.3 System Performance Investigation

Using the Spatial Acoustic Suite software, a series of performance investigations of the Ambisonics sound system were developed, with the scope of showing the accuracy of the wavefronts reproduction over a large listening area. For comparison, ideal simulations of wavefronts generated by a single natural sound source were initially rendered.

For the ideal dodecahedron configuration, the results show that the reproduction using first-order Ambisonics quickly degrades with increasing frequency. The simulations demonstrate, however, a very good behavior with the second-order Ambisonics rendering, showing consistency and correct recreation of the wavefront up to 1000 Hz. This confirms that the practical cut-off frequency is a function of the Ambisonics' order.

On the other hand, one notices a slight difference in the accuracy of the wavefront simulation depending on the virtual source location. One can see that the closer the source position is to any of the loudspeakers, the better the wavefront is reproduced; in contrast, the farther the virtual source moves from all loudspeakers, the more aliasing occurs and the wavefront becomes less accurate. In Chapter 5, one of the virtual sources used in our simulations represented the worst-case scenario in terms of location, and one could see that above 250 Hz, the wavefronts coming

from this direction are less accurately reproduced than in the other cases where the sources were located more in the vicinity of a loudspeaker.

The simulations of the ideal dodecahedron configuration show a better rendering accuracy of a sound field when using second-order Ambisonics as opposed to first-order Ambisonics. In addition, they show the fact that by recreating accurately the sound field over a large area, the system can provide a larger effective listening area, as opposed to a very narrow one. However, one has to note that these simulations ignored certain aspects of a real environment rendering, such as scattering from the loudspeakers or floor reflections specific to a hemi-anechoic environment, causing interference effects.

In addition to the simulations of the ideal dodecahedron configuration described above, we performed simulations of the system in its warped version, as it was used for practical reasons in our perceptual experiments. These simulations proved to be different from the ideal set-up they were designed to model. Due to the phase delays used to compensate for the warped dodecahedron geometry, phase inconsistencies appeared in the total sound field rendered by the twelve decoded signals. The added delays and gain components correct for the level and phase differences only for a limited center listening area. For this reason, the size of the effective listening area is compromised, as the wavefronts are not accurately reconstructed throughout the whole space, and thus the effective listening area is considerably diminished. In this case, the system would not be able to cover a large audience very successfully.

A second series of simulations zooming into a narrower listening area around the center of the system (with a radius of 30 centimeters from the origin) confirmed that the warped dodecahedron configuration was able to render a similar wavefront as in its ideal set-up, in its center area, as it was theoretically assumed. In this area the conclusions drawn previously for the ideal dodecahedron configuration apply, although future investigations are needed to determine the precise size of the listening area and on which Ambisonics parameters it depends (e.g., frequency, Ambisonics order, etc.).

6.2 Final Conclusion and Future Work

This research focuses on studying low-order Ambisonics sound rendering systems, in order to determine how accurately they can convey aural spatial information if implemented in a virtual environment for remote communication. The research was based on perceptual testing using human subjects and on a physical performance investigation, using a twelve-loudspeaker decoding configuration. While the results are not sufficient to determine whether these Ambisonics systems are appropriate for a very large listening area, they can successfully reproduce a sound field and convey localization information that is comparable to localization errors of the auditory system (as seen in [1] and [11]), for a listening position located at the origin of the Ambisonics coordinate system.

Further work should include similar perceptual testing at several locations off-center, as well as similar testing of other types of decoding configurations (loudspeakers numbers and configurations) while still using first- and second-order Ambisonics encoding.

LITERATURE CITED

- [1] J. Blauert. *Spatial Hearing: The Psychophysics of Human Sound Localization*. MIT Press, Cambridge Massachusetts, 1983.
- [2] I. Sutherland. The ultimate display. *Proceedings of the International Federation of Information Processing Congress*, pages 506–508, 1965.
- [3] http://www.holonet.khm.de/visual_alchemy/lumiere_x.html.
- [4] J. Daniel. *Représentation de champs acoustiques, application à la transmission et à la reproduction de scènes sonores complexes dans un contexte multimédia*. PhD thesis, Université Paris 6, 2000.
- [5] <http://history.acusd.edu/gen/recording/notes.html#origins>.
- [6] R. Streicher. The decca tree. *Mix Magazine*, 27(10):50–54, 2003.
- [7] M. Gerzon. Criteria for evaluation surround sound systems. *JAES*, 25(6):400–408, 1977.
- [8] J.S. Bamford and Paul Vanderkooy. Ambisonic sound for us. In *JAES*, presented at the 99th Convention AES, preprint 4138, 1995.
- [9] G. Steinke. Surround sound - the new phase: an overview. *100th Convention of the Audio Engineering Society*, 04 1996.
- [10] Lord Rayleigh. Our perception of the direction of a source of sound. *Nature*, 14(341):32–33, 1876.
- [11] E. Way. Localization of speech and speech-like samples. Master’s thesis, Rensselaer Polytechnic Institute, 2004.
- [12] M. Kleiner. Auralization - an overview. *presented at the 91st Convention AES*, 1991.
- [13] M. Gerzon. Surround sound psychoacoustics. *Wireless World*, 80(1468):483–486, 1974.
- [14] M. Gerzon and G. Barton. Ambisonic decoders for hdtv. *AES preprint 3345, Vienna*, 1992.
- [15] M. Gerzon. Nrdc surround-sound system. *Wireless World*, 83(1496):36–39, 1977.
- [16] www.muse.demon.co.uk/ref/speakers.html.

- [17] Richard J. Shavelson. *Statistical Reasoning for the Behavioral Sciences*. Allyn and Bacon, 1996.
- [18] P. Henderson. Software application spatial acoustics suite tm, Copyright 2000-2004.
- [19] P. Henderson. Wave field synthesis for perceptually accurate aural telepresence. Master's thesis, Rensselaer Polytechnic Institute, December 2003.