SPATIAL AUDIO REPRODUCTION TECHNIQUES AND THEIR APPLICATION TO MUSICAL COMPOSITION: THE ANALYSIS OF “WUNDERKAMMER”, “POINT-INSTANT” AND “HOLLOW”

A Master’s Thesis

by
ENGİN DAĞLIK

Department of
Music
İhsan Doğramacı Bilkent University
Ankara
August 2019

ENGİN DAĞLIK
SPATIAL AUDIO REPRODUCTION TECHNIQUES AND THEIR APPLICATION TO MUSICAL COMPOSITION: THE ANALYSIS OF “WUNDERKAMMER”, “POINT-INSTANT” AND “HOLLOW”
Bilkent University 2019

SPATIAL AUDIO REPRODUCTION TECHNIQUES AND THEIR APPLICATION TO MUSICAL COMPOSITION: THE ANALYSIS OF "WUNDERKAMMER", "POINT-INSTANT" AND "HOLLOW"

The Graduate School of Economics and Social Sciences

of

İhsan Doğramacı Bilkent University

by

ENGİN DAĞLIK

In Partial Fulfillment of the Requirements for the Degree of

MASTER OF ARTS IN MUSIC

THE DEPARTMENT OF

MUSIC

İHSAN DOĞRAMACI BİLKENT UNIVERSITY

August 2019

I certify that we have read this thesis and have found that it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Arts in Music

Dr. Öğr. Üyesi Tolga Yayalar
Advisor

I certify that we have read this thesis and have found that it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Arts in Music

Dr. Öğr. Üyesi Onur Türkmen
Examining Committee Member

I certify that we have read this thesis and have found that it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Arts in Music

Doç. Dr. Can Karadoğan
Examining Committee Member

Approval of the Graduate School of Economics and Social Sciences

Prof. Dr. Halime Demirkan
Director

ABSTRACT

SPATIAL AUDIO REPRODUCTION TECHNIQUES AND THEIR APPLICATION TO MUSICAL COMPOSITION: THE ANALYSIS OF "WUNDERKAMMER", "POINT-INSTANT" AND "HOLLOW"

Dağlık, Engin

MA,in Music

Supervisor: Dr. Öğr. Üyesi Tolga Yayalar

August 2019

The use of space is considered as a structural part of a musical composition. The developments in the spatial audio reproduction technology has changed the way we perceive the sound. So, it is inevitable to put emphasis on "composing space" in parallel with "composing sound".

This thesis is aimed at composers who seek how recent spatial audio reproduction technology contributes to the compositional process.

In order to grasp the idea behind "composing space" by using recent spatial audio reproduction methods; the importance of space in history, pioneering technological innovations in audio technology, brief technical explanations of the spatialisation techniques and analysis of three pieces in the context of spatial audio reproduction technologies will be presented and discussed throughout the text.

Keywords: Spatial Audio, Space, Soundfield Reproduction.

ÖZET

MEKANSAL SES ÜRETİM TEKNİKLERİ VE MÜZİKAL KOMPOZİSYONDAKİ UYGULAMALARI: "WUNDERKAMMER", "POINT-INSTANT" VE "HOLLOW" PARÇALARININ ANALİZİ

Dağlık, Engin
Yüksek Lisans, Müzik
Tez Danışmanı: Dr. Öğr. Üyesi Tolga Yayalar
Temmuz 2019

Müzikte mekan kullanımı, kompozisyon pratiğinin yapısal bir elemanı olarak kabul edilir. Mekansal ses üretimi teknolojisindeki gelişmeler sesi algılama şeklimizi değiştirmiştir. Bu yüzden "sesi bestelmek" ile beraber "mekanı da bestelemek" kaçınılmaz olmuştur.

Bu tez, yakın zamandaki mekansal ses üretim teknolojilerinin kompozisyonel sürece sağladığı katkıları merak eden besteciler için kaynak oluşturmayı hedefler.

Yakın zamandaki mekansal ses üretim teknolojilerini uygulayarak "mekanı besteleme" fikrini kavrayabilmek için metin boyunca; mekanın tarihteki önemi, ses teknolojilerindeki öncü yenilikler, mekansallaştırma tekniklerinin özet açıklamaları ve üç örnek parçanın mekansal ses üretim teknolojileri açısından analizi sunulup ele alınacaktır.

Anahtar sözcükler: Mekansal Ses, Mekan, Ses Alanı Üretimi.

TABLE OF CONTENTS

LIST OF FIGURES

INTRODUCTION

Space is an inseparable part of sound as an entity. If we consider a musical composition as an art form by composing sound, composing space is also a part of the creation process. Throughout the history of music, space has been always shaped the way of perceiving and creating music. With the advent of technology whether in acoustics or computer science, the way of implementing spatial thought into musical composition has been changed.

This thesis aims to represent how recent technology in spatial audio reproduction contributes to the way of composing music by discussing the roots of spatialisation in history, first by discussing the general technical definitions of recent spatial audio reproduction methods with their implementations, advantages and disadvantages, and then by presenting three pieces of mine in the context of spatialisation. The main purpose is to provide a relationship between the physical realities of the spatialisation technology and their creative usage in musical compositions.

In the first chapter, the main purpose is to briefly outline how recent technology has been evolved by showing the roots of spatial compositions. Although the standpoint of this thesis is generating spatial audio digitally, it is inevitable to represent what space means in the history of music. Especially, section 2.2 and 2.3 should be considered as the pioneering events which lead today's spatial audio reproduction methods.

In the second chapter, the most important spatialisation methods are described by their general technical formations. Their usage in musical art and audio industry is discussed and their technical limitations, advantages and disadvantages are compared as well. Instead of reading the technical aspects of the methods, this thesis will mainly focus on their applications in terms of musical composition. In this regard the thesis will be more beneficial to composers and aim to serve as a guide to spatial composition.

In the third chapter, three musical works of mine are presented in the context of the application process of spatialisation. The common focus of all three pieces is that space acts as the main element whether as a technical application or as a semantic approach to convey ideas behind them. It is beneficial to compare the actual facts of different spatialisation methods and their creative usages in the pieces.

The final chapter aims to summarize the overall content of the thesis. The relationships between history, methods and their applications are commented at this chapter against the backdrop of space in music. The common features and differences of the spatial audio reproduction techniques are raked together to provide an overview of the topic.

CHAPTER 1

THE ROOTS OF SPATIALISATION IN MUSIC

In order to understand how today's technology of spatial audio reproduction has been shaped, it is important to be acquainted with the evolution of the usage of space in terms of sound. It is inevitable to mention correlations between architecture and space and the resulting musical applications in order to grasp the starting points of the technology of spatialisation. So, the understanding of space in history and the pioneering technological developments in audio will be discussed in the following sections.

1.1. The Importance of Space Regarding Architecture in History

Spatialisation in music has a long and profound history dates back to the prehistoric era and the periods of Ancient Greece and Rome. One of the most important research about Paleolithic era and acoustics has been conducted by I. Reznikoff. He investigated the relationship between the locations of paintings and acoustic resonances in Paleolithic caves in France and found remarkable outcomes. Based on these acoustics researches, I. Reznikoff came up with the conclusion that "the more resonant the location, the more paintings or signs are situated in this location." (Reznikoff, 2008). The studies show that there was a correlation between the effects of resonant space and the paintings, indicating the ritual celebrations.

Later, as the agriculture invented and nomadic societies turned into settled ones, spaces for ritual and architectural understanding became more advanced. Ancient Greek "odeion", Hellenistic and Roman theater, is a good example of that understanding.

Aristotle and his student Aristoxenus who lived in 4th century BC, were among the earliest philosophers who focused on space, motion and the impact of change of location on sound. In his works such as "De Anima" and "Problemata", Aristotle deals with hearing and sound in detail. He underlines the importance and the impact of transmission of movements, location and distance in terms of the perception of the listener. (Johnstone, 2013).

Aristoxenus wrote two significant treatises entitled "Elementa Harmonica" and "Elementa Rhythmica" on subjects such as musical intervals, sound, modes and systems. In these treatises, Aristoxenus uses space-related terms, such as "upwards" and "downwards" motions, in a metaphorical way to describe the movement of the voice (Andrew, 1989).

To understand the history of spatialisation in music, and the effect of space on music, one should continue with the history of the Christian church, its architectural design and the emergence of polyphony. The use of space can be found in the practice of "antiphonal singing" and "responsorial singing" of Gregorian chants which dates back to the 8th century. Antiphonal refers to alternating choirs and the responsorial applies to a choir responding to a soloist. It is important to note that the history of "antiphons" dates back to Hebrew Psalms, and they are later introduced to the West in the 4th century (Britannica, 2007).

Furthermore, famous British architectural theorist and acoustician Hope Bagenal states that "the medieval church is the antithesis of the classic theatre. It is acoustically an enormous bathroom furnished with a whole system of baths in the shape of lesser cells and chapels which select and reinforce some tones above others." (Bagenal, 1927)

In the 15th and 16th centuries, a compositional style had been developed in Italy called "cori spezzati", practiced by the composers such as Adrian Willaert, Andrea Gabrieli, and Claudio Monteverdi (Howard, Moretti, & Moretti, 2009). Examples of pieces that take advantage of the use of space and the sonic capabilities of combining multiple choruses, include Willaert's Vespers of 1550 and Gabrieli's 3-choir mass for the visit of the ambassadors of Japan in 1585. Braxton Boren gives a good example for an earlier stereo effect dates back to Venetian Renaissance:

> ...Because the ruler of Venice, Doge Andrea Gritti, became too fat to reach his old elevated throne in the Basilica San Marco, in 1530 he moved his seat into the chancel, where previously the high priest had resided. After this move, Jacopo Sansovino, the chief architect of the church, constructed two pergoli, or raised singing galleries, on either side of the doge's new throne because former galleries lower down had become obstructed by wooden seating in the chancel. Moretti argues that these galleries were used to give a stereo effect for the performance of split-choir music to the doge's position. (Boren, 2017)

It is important to underline that church interiors have their own unique acoustic features, and these special acoustic characters resulted in the birth of some specific compositional styles. After the German reformation, churches were re-designed according to the needs of new preaching and singing in the native language. Hope Bagenal analyzed St. Thomas church at Leipzig and came up with very significant conclusions. After the reformation, wooden areas which are resonant were added to naked stones, and this resulted in the reduction of reverberation. Bach composed much of his cantatas for this church from 1723 until his death in 1750. Hope Bagenal claims that all these interior wooden additions made cantata and passion possible. Comparing the reverberation at a medieval church (from 6 to 8 seconds), present reverberation at Leipzig is around 2.5 seconds. (Rasmussen, 1964).

During the classical period, echo effects became popular among composers such as W.A. Mozart and Joseph Haydn. Haydn's symphony no:38 and his string sextet called "das echo", for two trios separated in two different rooms are some examples of alternating the performance space (Boren, 2017).

In the nineteenth century, many opera houses and concert halls continued to be designed and built. Hector Berlioz covered the subjects in his book entitled "Instrumentation" such as the placement of the orchestra members, concert hall acoustic and theater acoustic. The book, later, augmented by Richard Strauss (HERR & SIEBEIN, 1998).

Orchestras and halls evolved and grew in the nineteenth century. The architecture of concert halls became more advanced and scientific. Composers like G. Mahler, H. Berlioz and R. Wagner took advantage of these developments and composed new pieces which put a good deal of emphasis in tone color and orchestral effect. R. Wagner supervised the construction of Bayreuth Festspielhaus for the performance of his stage works which used the effect of space regarding architecture as shown in the Figure 1.1.

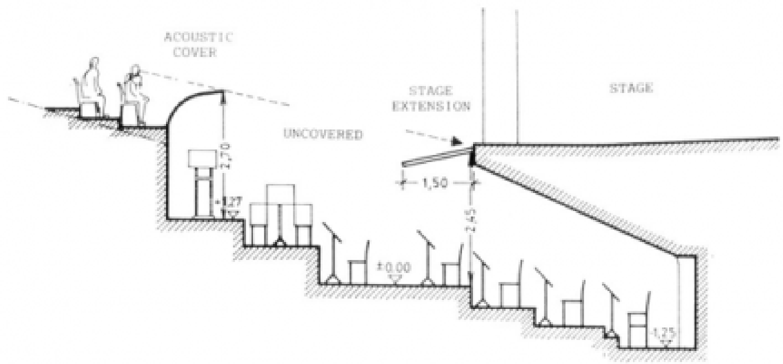

Figure 1.1: Layout of the orchestra pit of Bayreuth Festspielhaus

1.2. Early Technological Developments in Audio

In fact, it is the twentieth century that many artists and composers took advantage of the music technology, including recording technology and microphone technologies, invention of tape recorder and the loudspeaker etc... In the 20th century composers of both acoustic and electronic music found new ways of incorporating spatialisation in their compositions, which I will mention after giving a brief history of the developments of music technology now.

One of the most important developments in sound technology which gave way to audio spatialisation is the invention of the telephone. G.Bell was studying human hearing and speech which influenced his work in the invention of telephone. As M.F Davis states that Graham Bell did some experiments with two telephone receivers and transmitters (Davis, 2003). Another milestone marked the beginning of stereophonic sound reproduction had been realized by a French engineer, Clément Ader in 1881. Ader's idea was to put telephones on stage, so that people can listen the performances at home. As Ader himself describes in his patent file:

> The transmitters (i.e., telephone mouthpieces) are distributed in two groups on the stage - a left and a right one. The subscriber has likewise two receivers, one of them connected to the left ground of transmitters,

the other to the right one... This double listening to sound, received and transmitted by two different sets of appartus, produces the same effect on the ear that the stereoscope produces on the eye. (Fantel, 1981)

In 1881, C. Ader broadcasted the presentations in the Paris Opera during the Paris exposition. Later, the théâtrophone, a distribution system, evolved from his invention. With this advancement, the subscribers were able to listen to opera and theatre performances over the telephone lines.

Figure 1.2: Le Théâtrophone, an 1896 lithograph from the Les Maitre de L'Affiches series by Jules Chéret

As Braxton Boren states that:

> ... because it conveyed interaural differences to be conveyed to both ears of the listener, some have classified this as the earliest example of binaural sound. It is important to note that at the time, binaural was principally used to mean hearing with two ears, rather than recording with a real or synthetic human head. The modern distinction between binaural and

> stereo was not even suggested until the 1930s, and widespread usage of these separate definitions would not follow until the 1970s. (Boren, 2017)

Later, Harvey Fletcher developed "Oscar", a binaural recording device in 1931. Around 1930s, Alan Blumlein was working on capturing directional information on two channels of audio in EMI in Britain. It is possible to say that Blumlein's work is one of the early steps of ambisonic 3-D surround sound system. Walt Disney and engineers of RCA developed a special recording device for the animated film "Fantasia".

Stereo became commercial in 1950s. By the end of 1960s, quadrophonic system had been developed by Peter Scheiber, and in 1970s, Michael Gerzon, pioneering British engineer, has developed the idea of ambisonics system. In these years, hardware had been developed but they were not in use commercially until 1990s. Moreover, Dolby Laboratories contributed to the development of surround format and multichannel audio.

In recent years, from 1990s to today, diverse spatial audio technologies have been developed thanks to the application of the computers, including high order ambisonics (HOA), wave field synthesis (WFS), distance based amplitude panning (DBAP), vector based amplitude panning (VBAP) and then some.

1.3. The Applications of Spatialisation in Musical Composition in 20th Century

Spatial applications in music composition continued to evolve as the modern musical technology grew and developed in a more advanced level. American composer Charles Ives, in the beginning of twentieth century, applied some spatial ideas in his music such as placing particular instrument families in different positions on stage and having a different sonic results. It is the 1950s and 1960s that we can also observe some spatial applications in the instrumental works of I. Xenakis and H. Brant. H.Brant's work entitled 'Antiphony 1' (1953) is written for five spatially-separated orchestras. In 1965/6, I. Xenakis composed Terretektorh, based on a spatial idea creating a special sonic image in the minds of the listeners by placing them inside the orchestra in order to compose surrounding sound as shown in Figure 1.3.

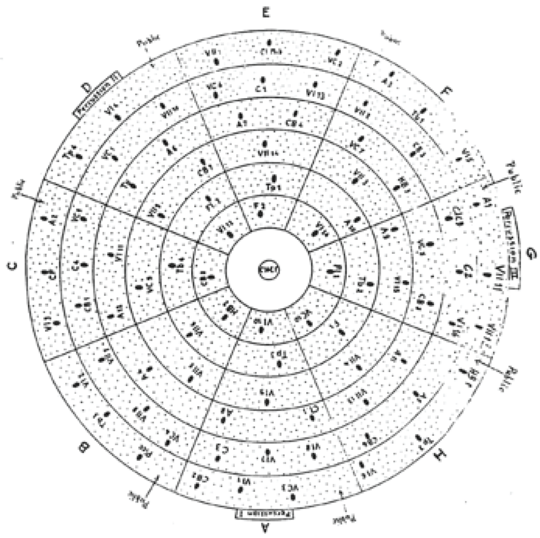

Figure 1.3: Arrangement of orchestras in Xenakis' Terretektorh

Another example for spatial applications in the instrumental music of twentieth century is I. Xenakis' Persephassa, composed in 1969 for six percussionists placed around the listeners, as shown in Figure 1.4. This piece is at the same time an attempt to create and organize different sonic points regarding the space. I. Xenakis makes the audience understand and feel the space that they are positioned in order to create more dynamic space. In his piece Eonta, composed in 1964, the brass players walk around on stage and change the position of their instruments, and this creates a different sonic impact by moving sounds through trajectory composition. Seating arrangements in his compositions in order to create spatial effect are considered visionary.

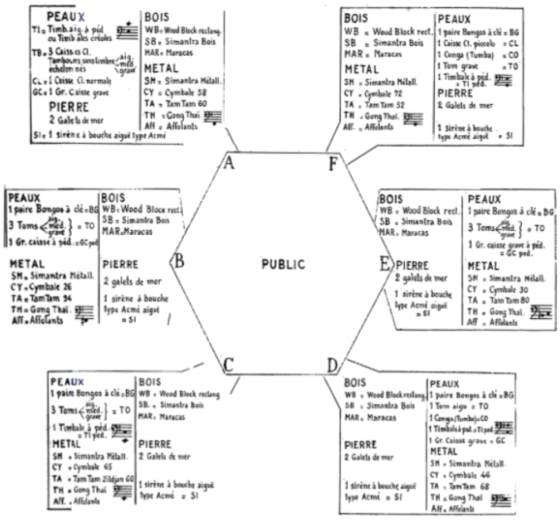

Figure 1.4: The Placement of Percussions in "Persephassa"

Invention of new electronic instruments such as Thaddeus Cahill's Telharmonium(1900-1906), and phonographs definitely affected the way the composers work in terms of spatial understanding. Another turning point was the researches done by P. Shaeffer in search of a link between modern musical technology and composition:

> Multitrack recorders were not yet available, so the RTF composers often used multiple mono tape decks, with up to five tape signals routed to a 4-channel speaker system. The speakers were arranged in a tetrahedral configuration, with Front Left and Right, Back, and Overhead. To facilitate distribution of the sound Schaeffer conceived of a mechanism called the "potentiomètre d'espace" (1951) which used induction coils to control the signal routing. The user interface was highly theatrical, consisting of four large hoops surrounding the performer, whose arm movements regulated the spatialization. (Zvonar, 1999)

In 1952, John Cage composed a piece entitled "Williams Mix" for eight mono tapes each playing through its own loudspeaker that positioned in the performance space.

Later in 1956, K.Stockhausen composed "Gesang der Jünglinge" for electronic sounds and the recorded voice of a boy soprano. This piece is generally considered as the first piece for multi-track tape. "Kontakte" (1960) by K.Stockhausen is also considered as the first quadraphonic composition. In his other pieces such as "Gruppen" and "Carre", K.Stockhausen searched for the spatial movement of sound extensively (Zvonar, 1999).

Another milestone in spatial composition is the installation of the tape composition entitled "Poème Électronique" (1958) by Edgard Varèse. The famous Phillips pavilion was designed by the architect and composer I. Xenakis, and E. Varèse's composition was installed to create a spatial multimedia environment. Each track in the piece was distributed to 425 speakers via an 11-channel sound system in the pavilion.

It is important to note that modular analog synthesizers and quadraphonic array of speakers also came available in the beginning of 1960s. Composer Morton Subotnick composed "Sidewinder" in 1971 and released a quadraphonic version of the album.

On the other hand, at the end of 1950s, Max V. Mathews invented computer sound synthesis at Bell laboratories and he completed his first program called "Music I" around the end of 50s. Later, John Chowning developed famous FM (frequency modulation) synthesis technique and worked on spatial sound.As a result, he composed his famous pieces entitled "Sabelith" (1970) and "Turenas" (1972) for quadraphonic system. He underlines:

> The culmination of two research paths, Turenas (1972), embodies both sound localization (begun in 1964) and Frequency Modulation (FM) synthesis (begun in 1967). Illusory motion of sound in space was a musical goal from the very beginning of my work with computers, a goal that led to a number of perceptual insights that not only set up my discovery of FM

> synthesis but insights that have enriched my thinking about music ever since. In tracing the long compositional trajectory of Turenas, I note the connections between the work of Mathews, Schroeder, and Risset, at Bell Telephone Laboratories, and the development of musical tools by David Poole, Leland Smith and myself within the research community at the Stanford Artificial Intelligence Laboratory (A.I. Lab). (Chowning, 2011)

Apart from these pioneering musical compositions, composers and sound designers has begun to develop a great variety of techniques in terms of spatial composition in parallel with the development of new spatial audio reproduction techniques. It is important to underline that cognizance of musical composition has changed dramatically when the spatialisation techniques began to contribute to the process of creating music. Composing sound movement and trajectory as well as installing sound in space changed the way the music is composed and perceived.

CHAPTER 2

SPATIAL AUDIO REPRODUCTION TECHNIQUES

2.1. Stereophony

Stereophonic or stereo sound is the most prevalent multichannel recording and reproduction technique of the audio technology for today's market since 1950s. Stereophony can basically be defined as constructing the illusion of an immersive sound environment by localizing the directional sound sources between two or more loudspeakers settled in front of the listener or inside the headphone system. When the headphone system is mentioned, it should be clarified that although binaural reproduction and recording technique is related with two-channel audio system, it should be approached as a diverse technique due to some reasons. Snow clearly stated that "the binaural system transports the listener to the original scene, whereas the stereophonic system transports the sound source to the listener's room." (Snow, 1953)

Panorama can be defined as continuous narrative scene to conform to a flat or curved background, which surrounds before the observer in visual arts (Britannica, 2018).

When it is aimed at generating illusion of sounds containing direction and depth, we are talking about creating a panorama inside the area of loudspeakers and listener. The field formed by panorama with two audio channels is called as stereo image. In order to perceive an accurate stereo image, the orientation and placement of the speakers is adjusted to a point specific location which is the center of the listening area. This point inside the listening area often referred as the sweet spot. The fact that the sound field can only be perceived accurately at the sweet spot, is one of the main disadvantages of most multichannel reproduction systems as they all rely on the presence of a sweet spot.

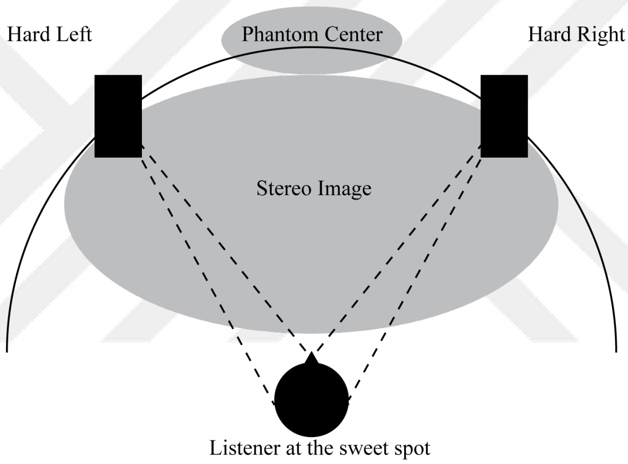

Figure 2.1: Representation of the stereo image

In stereophonic audio, the panorama of the sound field can be generated by either stereo recording with two or more microphones or panning the mono source into stereo with various techniques. In the following sections, the fundamental techniques of stereo panning will be represented in order to understand reproduction of the stereo image in a deeper manner.

2.1.1. Stereo Panning Types

If a sound source is aimed to spatialise in the stereo panorama, there are two possible techniques: adjusting the relative amplitudes for the left and right channels or delaying one channel respect to the other. In this section, the former approach will be discussed because the most common way of panning in the design of mixers and DAWs is distributing the loudness (power) of an input source between left and right channels of a stereo output in order to control the stereo panorama. However, there are various laws of panning based on mathematical equations which should be discussed in order to understand how they handle the distribution of the signal and their advantages and disadvantages concerning the stereo spatial field.

2.1.1.1. Simple Linear Panning

The simplest form of panning to localize a sound source in the stereo panorama is changing the amplitude levels of left and right channels equally. The implementation of this direct panning method is to increase the amplitude of one channel linearly from 0 to 1 while decreasing the other channel's amplitude linearly from 1 to 0 as shown in Figure 2.2.

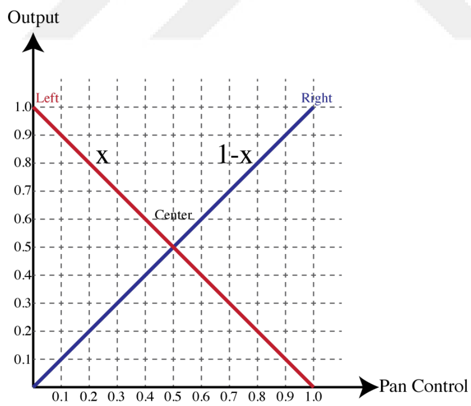

Figure 2.2: Linear amplitude panning

As it can be observed from Figure 2.2, the power of the source is distributed equally between two loudspeakers. The disadvantage of simple linear panning method is that in the center position, the amplitude of the signal is divided in half, shared by two loudspeakers, so each receives the half of the total amplitude. Thus, the resulting sound is weaker than the amplitude of the original signal. Although the method of simple linear panning works well when the source is closed to the right or left channel, the sound source loose its energy in the center of the panorama.(Farnell, 2010)

2.1.1.2. Sine-Cosine Panning

If it is desired to keep the energy of the input source constant at the center of the stereo image, the method of sine-cosine panning is the most efficient way. The mathematical representation of the signal's power is like that:

$$left.amp = \cos(a) \times input$$

$$right.amp = \sin(a) \times input$$

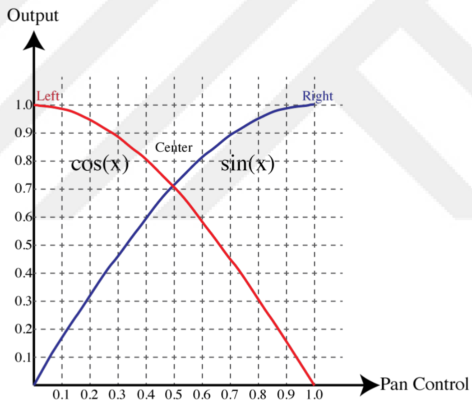

Figure 2.3: Sine-cosine amplitude panning

In sine-cosine panning, two loudspeakers are separated by +-45 degrees. Thus, the location of the source is at the left when $a = 0$, at center when $a = 45$ and at the right when $a = 90$ degrees. The exact left, right and center positions in the panorama are accurately predicted. However in this panning law, the stability of the position of the source is deviated when the sound panning is done in the area between center and left or between center and right (Griesinger, 2002).

The sine-cosine panning models the placement of the input source on a circle around the listener position. It works well for the classical music because an orchestra is generally arranged in a semicircle. However there is another panning law which has a better response around the center position if there is no need to pan hard left and hard right. That is the square-root panning (Farnell, 2010).

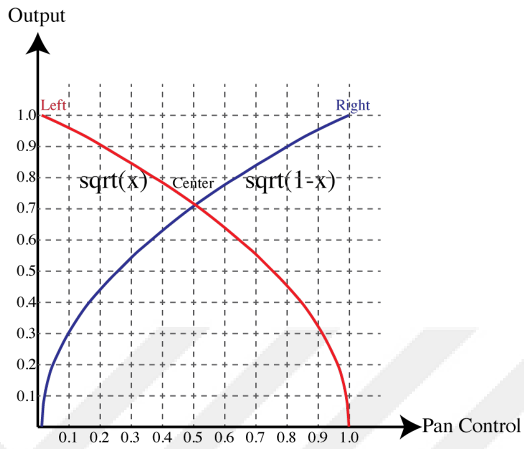

Figure 2.4: Square-root amplitude panning

2.1.1.3. Square-Root Panning

If the intensity of the loudness is desired to be same at any point of the stereo panorama, the square-root panning is the most convenient approach to panning. Like sine-cosine pan law, square root panning compensate the disadvantage of the linear panning which creates "hole in the middle" effect. The amplitudes of the left and right output of the sound source is basically calculated by this equation:

$$left.amp = \sqrt{\frac{1 - panSet}{2}}$$

$$right.amp = \sqrt{\frac{1 + panSet}{2}}$$

The resulting amplitude distribution between loudspeakers are slightly different from sine-cosine panning law, but one that also produces a nearly equal sound intensity as the pan setting is changed as shown in Figure 2.4 (Mitchell, 2009).

2.2. Binaural Audio

Etymologically the term of “binaural” refers to something that is related to “two ears”. So, binaural sound can be defined as two-channel audio which arrives at the right and left ears of the listener. Although general agreement about all stereo sound being binaural, the notion should be accepted in spatial audio terminology as “sound where the two-channel sound entering a listener’s ears has been filtered by a combination of time, intensity, and spectral cues intended to mimic human localization cues”. (Roginska & Geluso, 2017)

Mimicking two ears of a human is the key element of the binaural audio. The main idea behind binaural recording, synthesizing and reproducing is related to how each of the eardrum reacts to sound pressures. If this reaction is recorded or synthesized and after reproduced exactly as they were, the entire auditory sensation is accomplished with its features including timbre and spatial aspects. (Møller, 1992)

In the following sections localization process, binaural recording, synthesizing and reproduction technology and how this technology contribute to the field of spatial audio will be discussed.

2.2.1. Physical Cues To Localize Sound

The hearing needs some cues in order to determine where the sound source is coming from and how far it is located. The research of which physical factors affect the localization of a sound is based upon the “duplex theory” of Lord Rayleigh (1907) which still remains accurate with some extensions. The four fundamental cues are interaural time differences (ITD), interaural phase differences (IPD), interaural level differences (ILD) and coloration. Sound pressures are resulted from a sound wave which obtain a certain direction and distance according to listener. These pressures vibrate each eardrum in a slightly different way. This variation between left and right eardrum creates linear distortion such as coloration and interaural time and spectral differences. If these tiny differences can be correctly reproduced by recording or synthesizing, it is possible to reproduce the signals filtered by eardrums. (Møller, 1992)

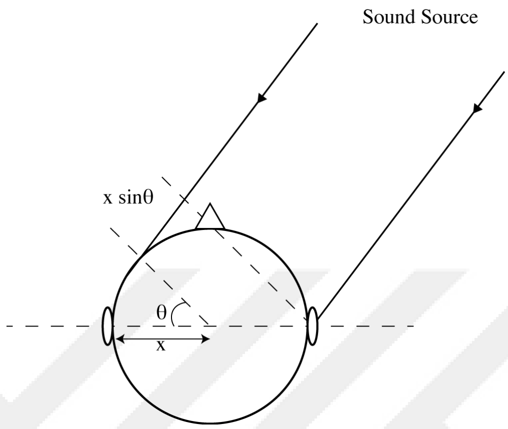

Figure 2.5: ITD and ILD caused by the shape and the size of the head

The most determinant cues to localize sound are “interaural time difference” and “interaural level difference” because the sound inherently arrives slightly earlier or later at one ear comparing the other when one of them is closer to the source as shown in the Figure 2.5. The ear which is farther from the sound source is vibrated later comparing the other ear. As a result of this time delay, the ”interaural time difference” (ITD) occurs. “interaural phase difference” (IPD) is clearly related to the ITD however this localization cue is mostly prominent in pure tones. The difference of the phase in the pure tones also creates a time difference which is the reason to distinguish the content of the signal in the time domain. On the other hand, human head blocks some energy of the sound due to the angle difference between sound source-left ear and sound source-right ear. This physical interference creates the “shadowing effect” and results in “interaural level difference” (ILD). Because of the physical shape and the size of the head, ILDs are most noticeable at higher frequencies, especially above 1.5 kHz. The size of the head of a human is big enough to produce reflection of the incoming signals when compared to the wavelength of the sound. However, “interaural time difference” occurs at all frequencies. These intensity and delay differences are the most effective cues to localize a sound source for a human. (Stern, 2006)

The other physical factor that constitute a localization cue is the spectral coloration. The fissures of the “pinnae” (outer ear) ,as shown in Figure 2.6, reshape the spectral content of original signals arriving at eardrums. This spectral cues are especially essential in the localization of the vertical plane and the decision of front-back directionality in sound sources. It is difficult to state a unique explanation how this physical cue process in a physical and mathematical manner due to the fact that each human has a unique shape of outer ears. Even the asymmetry of two ears affect the localization process. However the researches about how these spectral filtering process creates the transfer function resulted in a concrete conclusion which is frequently referred as the “head-related transfer function” (HRTF) and the representation in the time domain called “head-related impulse response” (HRIR). These transfer functions will be discussed for further information in the following sections.

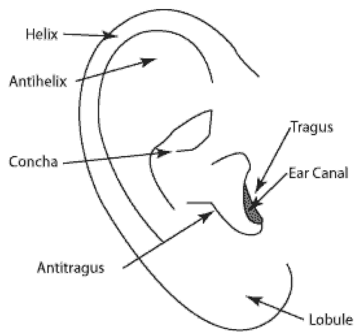

Figure 2.6: Pinnae (Outer Ear)

2.2.2. Binaural Recording

The most direct reproduction technique for the binaural audio is directly recording the audio signals from the perspective of the ears of a human. The details about how we record the sound source effects the overall definition of the final product. The recording can be done by putting small microphones into the ear canals of a listener or a model of a human head may be used. We can define the process of placing two microphones either at two ears on a human head as the "listening subject recording"; or recording inside the ears of an artificial head as "dummy head recording". The latter method constitute more controllable and consistent environment compared to former. Still, both of them have certain advantages and disadvantages. (Zhang, Samarasinghe, Chen, & Abhayapala, 2017)

In order to create filtering which result from the physical features of the head creating localization cues, a good representation of the human head is needed. "Dummy head" is an artificial design of a human's head which represent the anatomical features of an adult head. The size and the shape of the head, location of ears, pinnae, shoulders and torso are the most key elements of a dummy head. Moreover, the quality of the little microphones and their placement should not be underestimated regarding the final result of the recording. It should be stated that an average design of a dummy head that creates high fidelity recording for everyone, is almost impossible due to the fact that each human has a unique anatomy which creates a unique binaural hearing. So, theoretically if one wants to enhance their own definition of the reproduction of binaural audio, she/he must create an exact model of her/his head.

There are different types of dummy heads available in the market to capture sound. These may be categorized into three types: dummy head including just the head, dummy head including shoulder and torso and binaural microphones placed in artificial ears without a head as shown in Figure 2.7. (Roginska & Geluso, 2017)

Copying a human head accurately is not enough to generate binaural reproduction. It is important how we record the signals and where to place the microphones whether into a dummy head or a human head. The placement of the microphones can be performed at three points to get the full spatial information. These three different methods can be done by recording at the eardrum, entrance to the open ear canal or

a b c

Figure 2.7: (a)dummy head including just the head, (b)dummy head including shoulder and torso, (c)binaural microphones placed in artificial ears without a head

entrance to the blocked ear canal (Møller, 1992). It is hard to specify which method is superior because other physical factors such as the type of the dummy head, intended result *etc*... also should be considered. Although it is not clear which method is the best, headphone calibration and equalization correction should be implemented according to recording types in order to compensate the spectral and spatial distortions caused by microphone types, microphone placement, recording room effect.

2.2.3. Synthesizing Binaural Audio

If one desires to localize a virtual sound and decode in binaural sound field, she/he needs a sophisticated algorithm that mimics all natural localization cues resulted from the nature of human hearing. To simulate a binaural signal which reflects a sound source that appears at a place around the listener, the virtual source should be convolved with head-related transfer function (HRTF) which is a similar process to impulse response (IR).

The measurement of HRTFs are done by placing microphones in the ears of a human or a dummy head, like binaural recording, and test sound is recorded from different locations including median, frontal, and horizontal planes. The purpose is generating a detailed localization map in order to localize virtual sound without recording. The data generated by HRTF measurement contains all localization cues which is used for the convolution process. It can be stated that the synthesizing the binaural sound field is the absolute reverse process of the natural hearing of a human as shown in Figure 2.8. As stated earlier, there are physical differences between individuals so HRTFs diverge greatly regarding the general shape and details of the resulting transfer functions. Correspondingly, perceptual distortions are occurred. However, calibrating the localization distortions, equalizing the HRTF recordings after constituting the HRTFs minimize these distortions -especially elevation problems and frontal localization distortions- seriously.

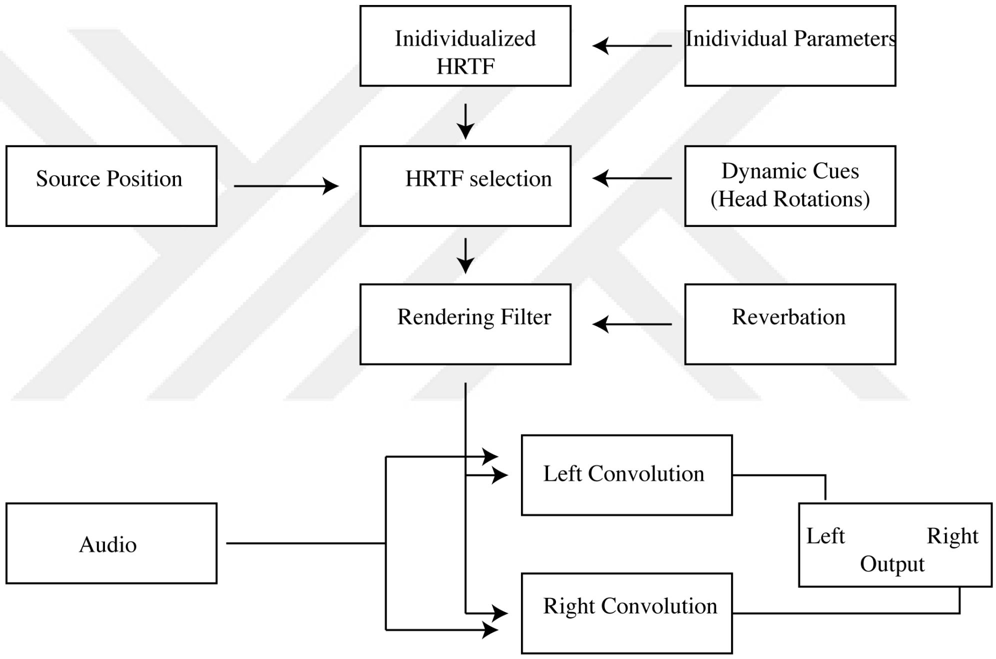

Figure 2.8: Complete signal flow of a binaural rendering system

HRTFs are typically recorded in an anechoic room which do not include any room information. It is also possible to combine HRTFs with room impulse response measurements which is called binaural room impulse responses with the abbreviation BRIR. This type of reverberation helps to decrease frontal and elevation distortions to a large extent. (Zhang et al., 2017)

2.2.4. Usage of Binaural Audio

The most appropriate tool for the reproduction of binaural audio is to output the final result through headphones. Using stereo speakers create a distortion due to the phenomenon called "cross-talk" which is defined as how much the audio signal of the left speaker leaks into the right ear or vice verse. Although there are systems called "cross-talk cancellation" in order to correct the binaural sound field through speakers, accuracy of the spatial auditory image is much more greater by using headphone system. Furthermore, headphones provide not only a controlled listening environment free from the "sweet spot" notion but also acoustic isolation from ambient sounds which creates outside distortion.

There are variety of headphone types which change the experience of listening binaural audio. Common types of headphones can be listed as over the ear headphones, in-ear monitors, ear bud headphones, multi driver headphones, ear speakers. The type, event the brand, of headphone may also distort the binaural sound field. In order to decrease the distortion created by the headphones, equalization and calibration should be realized carefully. The unnatural spectral coloration is occurred because of the frequency response that is introduced by the acoustic emitter of the headphones. In order to improve the externalization with non-individualized HRTFs and increase the localization definition, equalization and calibration should be implied.

It is essential to remark that binaural audio reproduction over headphone is the most practical way to deliver spatial audio regarding today's market. There are numerous applications ranging from Virtual Reality (VR), Augmented Reality (AR), telepresence, music reproduction, virtual acoustics, and simulation, to mission-critical applications.(Roginska & Geluso, 2017)

2.3. Vector Base Amplitude Panning (VBAP)

Vector base amplitude panning (VBAP) is a method of spatialisation based on amplitude panning method in arbitrary 2-D or 3-D loudspeaker setups. The technique which was developed by Ville Pulkki in 1997, is an extension of stereophonic panning method by implying vectorial reformulation and extrapolation of the pair-wise amplitude panning. VBAP aims to create a spatial sound field around a listener by using unlimited number of loudspeakers which can be placed randomly with the restriction of being equidistant to the listener. The reformulation of pair-wise panning method is in use in 2-D setup while a triplet-wise panning method is implied in 3-D setup. (Pulkki, 2001)

It has been known since 1973 that head rotations create more distortion in stereophonic image in the implementation of tangent panning than in the usage of sine panning in the regular stereophonic loudspeaker setup. However Pulkki proposed that tangent panning can be efficient by using an equipollent, vector-based system not only in the horizontal plane but also in the three-dimensional extension to pair-wise level panning, allows to render elevated sound sources in a malleable loudspeaker setup. (Hacihabiboglu, De Sena, Cvetkovic, Johnston, & Smith III, 2017)

Basically, the algorithm of VBAP determines the nearest two or three speakers (two in 2-D setup, three in 3-D setup) in order to compute the amplitude distribution depending on the locations of the chosen speakers. (Pérez-López, 2014)

2.3.1. Technical Aspects

All technical features of the VBAP will be explained in the situation of 3-D speaker layout. All of the formulations may be reduced into 2-D setup without any difference except the reduction of the chosen speakers as explained in the previous section.

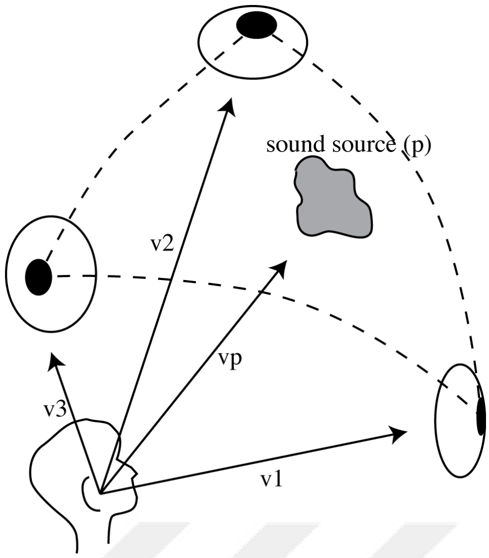

Figure 2.9: Triplet-wise speaker placement in VBAP

When VBAP is implemented on the three-dimensional speaker setup with more than three speakers, only three speakers are chosen to output the audio signal at a single time. The loudspeakers should constitute triangular composition and must be equidistant to the center of the listening area as shown in the Figure 2.8. For a specific sound source position p and the center of the listening position O, just the specific triplet of speakers is determined. As a result, the direction (d) vector between the sound source and the sweet spot can be specified as:

$$d = \frac{p - O}{\mid p - O \mid}$$

If we define gains of each loudspeaker as a vector G = [g1 g2 g3] and unit directions as L = [l1, l2, l3]; the vector of gains can therefore be stated as:

$$G = dL$$

In addition to the calculation of the gains depending of their directionality, final calculated gains of the loudspeakers are normalized to keep the total intensity constant.

It is important to state how the loudspeaker triplets are determined by the algorithm of VBAP. The following conditions should be satisfied for the decision:

1- The loudspeakers and the listener should not be in the same plane.

2- Triangles should not be overlapped.

3- Triangles should have the shortest sides if possible.

In order to calculate gain factors in 3-D VBAP, the first condition should be satisfied. If the listener and loudspeakers will be in the same plane, there will not be 3-D spatial data in the vector base.

The second condition, the triangles can not be overlapped because the sound source which crosses a side of the changing triangle should not be affected the immediate changes of the gains of the loudspeakers. If overlapping is arisen, there would be rapid gain transitions.

The third condition emphasizes that the lengths of the triangle sides should be as small as possible. In order to keep definition of the localization of the sound sources, selected loudspeakers should be as close as possible. (Pulkki, 2001)

As it can be understood that the final output signal is generated with just three speakers in VBAP. This situation may create some problems if the speaker layout includes a large quantity of speakers. The main problem would be the lack of spread of the virtual source in the sound field. To overcome this complication, Pulkki extended VBAP to multiple-direction amplitude panning (MDAP). With the help of this algorithm extension, it is possible to control the number of activated speakers fed by the virtual sound source. Under favor of MDAP, one can hold not only the perceived width of the sound source but also the overall coloration of the output. (Hacihabiboglu et al., 2017)

2.3.2. Evaluation of VBAP

As it was stated earlier; to put it all in simple terms, VBAP is an extension to the tangent panning law which is based on amplitude variation between loudspeakers. The technique is widely used because of its simplicity and effectiveness. However there are also some limitations and drawbacks of the system as well.

In larger speaker setups, the limitation of choosing triplet-wise loudspeaker cause narrowed spread of the source. As explained in previous section, MDAP is used to overcome this drawback. Borß also propose a system which generate symmetric panning gains for symmetric number of loudspeaker layout by using N-wise panning technique by using polygons instead of triangles. Although there are a lot of similarities between VBAP, MDAP and the N-wise panning technique (Borß calls it "Edge Fading Amplitude Panning"), MDAP and EFAP result in a greater boost in the lower frequencies which may be considered as a disadvantage in certain conditions. (Roginska & Geluso, 2017)

When the loudspeaker setup is near the median plane, it is obvious that VBAP predicts the localization of the sound source quite accurately. However, the perception of the direction is corrupted to some extent if the loudspeaker setup is moved to lateral plane. (Pulkki, 2001)

As it is the case for the most spatialisation technique, VBAP also requires the need of the "sweet spot". Although the sweet spot in VBAP is slightly larger compared to ambisonics, which will be explained later, it does not include any distance data in its algorithm.

Furthermore, when the position of a loudspeaker and the sound source coincides exactly, that speaker will output the complete signal. This situation leads to reduced width of the perceived sound source. (Pérez-López, 2014)

It is important to mention that the sound sources at different locations inside the same direction vector make no difference in the output resulted from VBAP algorithm. In order to solve this problem, one should blend distance-based data into the vector base panning technique. The transitions between vector base and distance base algorithms utilize the overall accuracy of the spatial field reproduction.

Although there are huge drawbacks in the implementation of VBAP, it still provide the most simple and effective panning technique compared to other sound-field reproduction technique because of the efficiency of CPU load and the simplicity of the algorithm.

2.4. Distance Based Amplitude Panning (DBAP)

Distance based amplitude panning (DBAP) was proposed by not only Lossius and Pascal Baltazar at 2009 but also Kostadinov and Reiss in 2010, independently. As it was stated in the previous sections, most of the spatialisation techniques require a sweet spot in order to provide accurate spatial information. This condition is not ideal especially in some concert situations or sound installations. By contrast with those techniques that require a limited listening area, the algorithm of DBAP provides an alternative spatialisation technique which does not depend on a specific speaker layout and the position of the listener. In other words, DBAP is a convenient approach for not only irregular loudspeaker layouts but also for the situations where a sweet spot is not possible.

The idea behind DBAP is that a matrix based algorithm is applied in order to take the exact positions of the loudspeakers into consideration for distributing the virtual sources. So, the main parameter in the algorithm is the distance between the sources and loudspeakers.

DBAP is similar to VBAP where speakers' position determine the gain factors. However, DBAP uses the distance data in calculating the gains rather than implying the directional components into algorithm. At the end, the most important feature of DBAP is that the gains of each loudspeaker are independent of the location of the listener. The only essential factor is the distances to the sound source.

2.4.1. Technical Aspects and Formulation

The technique of distance based amplitude panning is very similar to simple linear panning method (in other words, equal intensity panning). DBAP just extends the quantity of loudspeakers from being pairwise into a loudspeaker layout of any size without any assumptions about their locations in the space. Although there are some extensions in the algorithm of DBAP, it is important to understand the simplest formulation of the method in order to understand the spatialisation process. The formulation will be discussed in two dimensional space, however it is easy to extend the model to three dimensions.

If we assign the coordinates of a sound source as (x_s, y_s) for a setup include N loudspeakers. We can calculate the distance d_i between the source and the i*th* speaker which has the coordinate of (x_i, y_i) as:

$$d_i = \sqrt{(x_i - x_s)^2 + (y_i - y_s)^2} \quad \text{for} \quad 1 \leq i \leq N$$

If the amplitude of the sound source is fixed with unity, the amplitude of the i*th* speaker (v_i) is calculated as:

$$I = \sum_{i=1}^{N} v_i^2 = 1$$

Furthermore; if it is assumed that all speakers output signal at all times, relative amplitude of the i*th* speaker according to the distance from the source is:

$$v_i = \frac{k}{2d_i a}$$

k is the coefficient of the location between all speakers and the source while a is the coefficient of roll-off R in decibels per doubling the distance.

$$a = 10^{\frac{-R}{20}}$$

This equation shows the relation between distance and the attenuation of the level and generally R = 6db. When all of the equations are combined, the simplest representation of the DBAP is shown as:

$$k = \frac{2a}{\sqrt{\sum_{i=1}^{N} \frac{1}{d_i^2}}}$$

(Lossius, Baltazar, & de la Hogue, 2009)

2.4.2. Drawbacks and Extensions of DBAP

If the sound source coincide with the exact location of a loudspeaker, the division by zero will occur in the equation because of the zero distance between the loudspeaker and the source. In that case, the output gain occurs at only that speaker. As it is the same case with the VBAP; the spatial spread and coloration problem arises. In order to prevent this problem, the normalization should be applied by using a scaling factor to the distance parameter. As a result, spatial blur will be less sensitive to variations in the size of the loudspeaker setup. (Lossius et al., 2009)

The algorithm of the DBAP is not able to calculate the gains of the sources which are located outside the field of the loudspeaker setup. To atone this calculation failure, the embracing field which is called "convex hull" should firstly be calculated in order to determine whether the sound source is inside or outside the convex hull. If the position of the sound source is outside the convex hull, the shortest distance between the boundary of the convex hull and the location of the source is calculated in order to convert to subsequent calculations. Thus, it is possible to process gain attenuation. (Lossius et al., 2009)

DBAP also includes the limitation of distribution of the source into all of the loudspeakers all of the time. If the restriction to distribute to a subset of loudspeakers are needed, then an extended feature is also needed. The most appropriate solution to this restriction is provided by the technique called KNN panning technique which is available IRCAM's Spat engine. KNN panning technique is completely based on DBAP with the extra feature which lets to choose how many loudspeakers will output the final signal. The algorithm is based on "k-nearest neighbors" algorithm mostly used in statistics and data science. According to algorithm, the engine choose the nearest loudspeakers to the sound source in order to distribute the spatialized output. This method provides to create restricted spatial spread for the localization process which is effective especially creating spatial scenarios when it is needed.

2.4.3. Evaluation of DBAP

DBAP provides a great flexibility to compose spatial scenes in an arbitrary number of loudspeakers using asymmetrical speaker arrangements. The easiness of the gain calculations ensures the simplicity and results in low CPU load.

The system is not dependent on the listener position like VBAP, stereo or ambisonics. The method is suitable for both two and three dimensional setups. It is effective to implement in special installation projects. The idea behind this spatialisation method is universal and can be used for home movie and game systems. (Kostadinov, Reiss, & Mladenov, 2010)

2.5. Ambisonics

Ambisonics is a sophisticated spatialisation method developed by Michael A. Gerzon from the University of Oxford in the 1970s. It can be implemented in two or three dimensional space by deconstructing of a given sound field using spherical harmonics. Then, the method allows to reconstruct the sound field in various loudspeaker layouts by generating appropriate decoding coefficients in matrix.

Gerzon firstly introduced the "First-Order Ambisonic" also called B-Format, which contains the information of directionality of a given three-dimensional sound field into four channels symbolized by (W, X, Y, Z). Channel "W" represents the omnidirectional signal of the encoded source while X, Y, Z serve as front-back, left-right and up-down respectively.

The conventional directions of ambisonic format is as stated and generally called as Furse-Malham order. However some applications like Spat by Ircam use different conventions. They may denote X as left-right and Y as front-back. It is very important to check the order of the channels within the application in order to implement the spatialisation accurately. The conventional Furse-Malham order will be used throughout this text.

2.5.1. The Roots of Generating Sound Field: Mid-Side Recording

When the ambisonic system was introduced as a representation of a sound field, it is inevitable to touch on the starting points that inspired Gerzon to create the algorithm of B-Format. The first attempts to develop sound field approach can be found at the initial works of Blumlein on X/Y technique and most importantly the M/S technique for stereo recording.

Two directional microphones which are perpendicular to each other constitute the

X/Y pair. On the other hand, M/S pair includes one omni-directional (or cardioid) microphone facing forward (Mid) and one figure of eight microphone along the left/right axis (Sides) as shown in Figure 3.10.

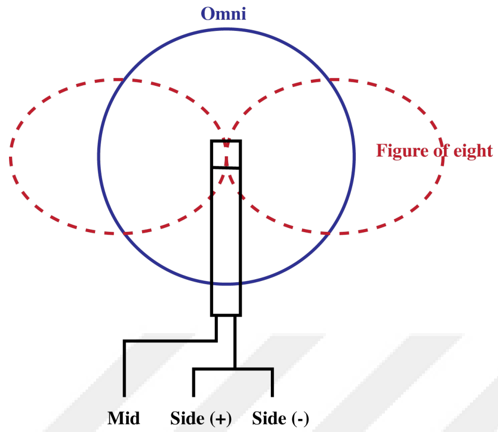

Figure 2.10: Illustration of Mid-Side Recording

Especially the idea behind the M/S recording technique is very similar to ambisonic encoding algorithm. The Mid microphone contains the frontal signal of the sound field. In addition, it represents the overall power of the sound field because of the feature of recording pattern of omni-directional signal. Mid channel resembles with W channel of the B-format in the aspect of including the overall propagation of the source in the sound field.

The side microphone captures the lateral components of the sound field. By convention, the left side of the recorded signal is represented by the positive phase of the channel, and the right side can be extracted by the negative phase of the same channel.

It can be concluded that recorded channels are not sufficient to output the intended result without implying some further process. Recording process just contains the audio data of the sound field and is called as encoded signal. If the difference of Mid and Side channels are computed, the right channel of the output can be achieved. On the other hand, the sum of Mid and Side channels gives the left channel of the final output. This simple algorithm, with the help of phase differences of the side recording, provides to distribute spatialized output to the loudspeakers.

$$L = M + S$$

$$R = M - S$$

It is important to understand this simple calculation in order to grasp the idea of the ambisonics. Both methods have the intention of composing directionality in a cartesian plane to constitute the final sound field. While the method of M/S creates just left-right directionality of the sound field, the technique of ambisonics aims to create full spherical components of the given sound field.

2.5.2. First-Order Ambisonic Recording

After providing a basis with M/S recording technique, it is beneficial to grasp how to record full spherical sound field in order to understand the formulation of the ambisonics used in digital encoding process.

Unlike conventional multichannel surround reproduction techniques, sound field systems consider all directions equally. So, in order to record all directions of the sound field, at least three microphones with the feature of figure of eight are needed including all directional axes (X, Y, Z). In addition, we need one omni-directional microphone to capture the overall pressure of the field (W). However this arrangement is not practical to implement (Craven & Gerzon, 1977). The intended result can also be achieved with four cardioid microphones arranged in tetrahedron as shown in the Figure 2.11.

The tetrahedral microphone including four cardioid microphones does not directly provide the components of W, X, Y, Z. The signals recorded by tetrahedral microphones provide the fields of left-front (LF), right-front (RF), left-back (LB) and

Figure 2.11: Illustration of Tetrahedral Microphone and its directions

right-back (RB). These recorded components constitute the A-Format ambisonic. In order to achieve B-format components (W, X, Y, Z) as shown in Figure 2.12, an encoding process should be applied as a linear sum of the signals (LF, RF, LB, RB):

$$W = LF + LB + RF + RB$$

$$X = LF - LB + RF - RB$$

$$Y = LF + LB - RF - RB$$

$$Z = LF - LB - RF + RB$$

It can be concluded that tetrahedral microphone is a three dimensional extension of M/S microphone technique. The differences of the four recorded signals give the encoded output which represents the spherical sound field. It is obvious that if we extracted the Z component of the B-Format, two dimensional representation of the sound field can be achieved. The flexibility of the ambisonic system provides to imply the encoded signals into different loudspeaker layout which will be covered in the following sections.

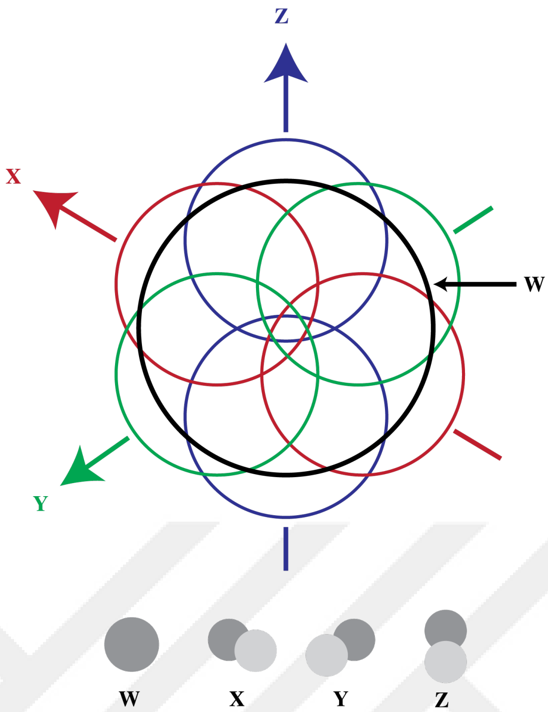

Figure 2.12: Directional Components of B-Format Ambisonic Field

2.5.3. The Formulation of Encoding

In B-Format ambisonic, the positions of the sound sources are accepted either on the surface of a circle in two dimensional space or inside unit sphere. Unit sphere means the sphere contains radius with one unit. If the position of a sound source is located outside the unit sphere, the decoding process do not work properly because the algorithm calculates its location at the nearest location of the sphere. Thus, the first rule of the algorithm stipulates the following condition:

$$(x^2 + y^2 + z^2) \leq 1$$

It is always important to remember that x denotes the distance along front-back axis , y is the distance along left-right axis and z is the up-down axis. When it is assumed that the source is located within the sphere and has the angles according to the center with A between horizontal plane and B between vertical plane (in the case of counter-clockwise), it is possible to define the position of a sound source by trigonometric functions:

$$x = \cos A \times \cos B$$

$$y = \sin A \times \cos B$$

$$z = \sin B$$

It is essential to underline that there is not any distance data in determination of the position of a source. These coordinates are used as multipliers in the calculation of components of the B-format:

$$X = inputsignal \times cosA \times cosB$$

$$Y = inputsignal \times sinA \times cosB$$

$$Z = inputsignal \times sinB$$

$$W = inputsignal \times \frac{\sqrt{2}}{2}$$

The multiplier to calculate the component W provides a more even distribution in all of the four channels. Through the instrument of these equations, it is possible to position the monophonic sounds anywhere in the spatial sound field. (Malham & Myatt, 1995)

That representation of the algorithm of ambisonics is in the simplest form. There are other extensions which denote the rotational principles, higher order implementations, room acoustics, zoom-like operations *etc...* It is intended to grasp the idea behind the method with this simplest form of ambisonic algorithm.

2.5.4. Higher-Order Ambisonics

Ambisonic method can be extended to higher orders which have lots of contributions such as increasing the area of the sweet spot, augmenting the definition of the localization *etc...* Before explaining the work flow of the higher order ambisonics, it is important to know what ”order” means in ambisonics.

If the W component of the B-format is considered as a single channel, it represents the power of the signal without any generated direction. This single channel is considered as the 0th order contains no direction. When we divide the three dimensional space into three harmonic fields which represent the all directions (X, Y, Z) equally, we obtain the 1st order of a given sphere. This logic can be extended to infinite symmetric divisions of the spatial field. So, dividing the three dimensional spatial field into higher orders than the 1st order are called higher order ambisonics. The symmetric divisions of the sphere can be represented as shown in Figure 2.13.

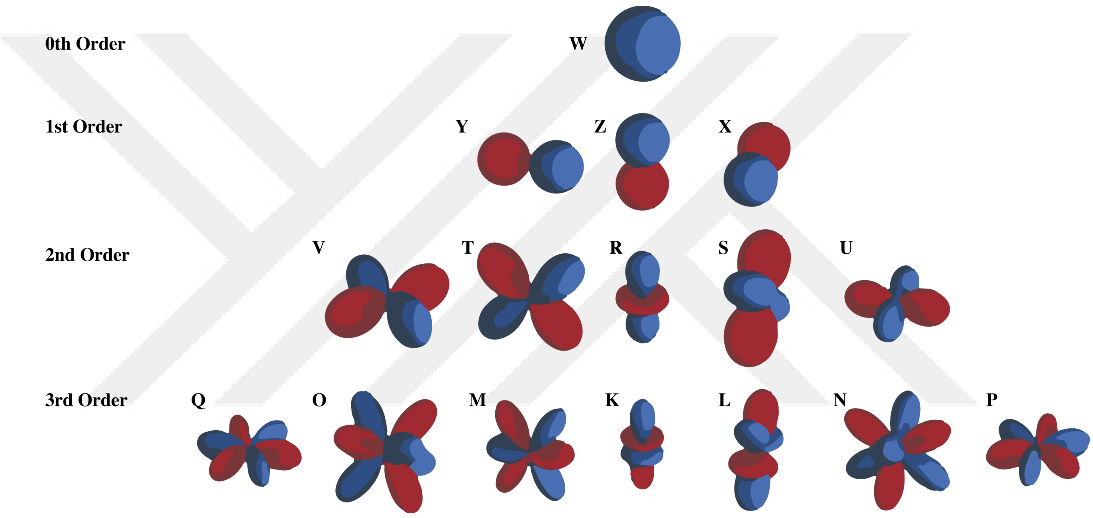

Figure 2.13: Representation of the harmonic fields up to 3rd order

Original definition of the ambisonic denotes that it is limited to 0th and 1st order components. If the components of order exceeds 1 (m ¿ 1), the technique is called higher order ambisonics. There is a simple calculation of finding the number of components respect to the number of order. If we assign the order number as m, the number of the components in two dimensional space is calculated as:

$$Component \quad number \quad in \quad 2D = (2m+1)$$

On the other hand, the equation to find the number of components in 3D is denoted as:

$$Component \quad number \quad in \quad 3D = (m+1)^2$$

These components of HOA represents the spatial information as a function of azimuth and elevation angles. It is very essential to state that the minimum number of the loudspeakers which are needed to decode the signals are the number of components of the HOA. More loudspeakers than the number of the components of HOA may be used to output the final audio respecting the dimensions of the system.

The order number of HOA is related with the definition and the accuracy of the spatial information. Furthermore, when the order number increases, the area of the sweet spot may be increased to some extent. The reason of the increase in the spatial accuracy mostly depends on the definition of the spatialisation of high frequencies. To give an example, the 4th order system can convey spatial information of 250 Hz in a very accurate manner. On the other hand, in order to create high definition for 1 kHZ, we need to encode the sound field in 19th order.

2.5.5. Decoding

The encoded signals of the ambisonics are not feeding the loudspeakers themselves, they just convey the directional information of the spatial field. So, it means that encoded signals are completely independent from the loudspeaker layout. Furthermore, the system does not stipulate a certain number of loudspeakers for the reproduction. As it is stated earlier, the only rule regarding the number of the loudspeaker setup is that the loudspeaker quantity must be at least equal to the number of channels which result from encoding process.

Another important point to consider is that the loudspeaker layout should be as regular as possible in order to provide accurate spatial information. This limitation occurs because the reproduction is done for a limited area, sweet spot. Even, the movements of the head can distract the accuracy of the resulted reproduction. Furthermore, the most problematic thing that should be handled carefully is low frequencies which may distract the localization seriously. However there are some optimization procedures within the decoding stage that can minimize these disadvantages of ambisonic system. The most prevalent optimization techniques are basic, inphase, maxre and their hybrid implementations. The basic optimization provides a correct reproduction of the wavefront at the sweet spot, for low frequency range. The maxre technique is considered to maximize the energy vector. Thus, it concentrates the energy in the direction of the sound sources. At the sweet spot it handles the localization of high frequencies more accurately than the low frequencies. Inphase optimization provides more accurate localization for off-center listening areas. This can be occurred by fading out the distributed signals of the virtual sources to loudspeakers when the sources get further from the loudspeakers. (Carpentier, Noisternig, & Warusfel, 2015)

In addition, it is possible to implement two optimization methods at the same time for different frequency ranges. This hybrid optimization technique provides to handle low and high frequencies with different optimization methods at the same time. To give an example, it is possible to use basic optimization for the frequencies under 300 Hz while implementing maxre optimization for the frequencies above 300 Hz. This feature enables to augment the definition of the reproduced spatial field.

2.5.6. Evaluation of Ambisonics

The most advantageous feature of the ambisonics is being isotropic which means the system treats all directions equally. The aim of the system is to create the sound field by using spherical harmonics. This is completely different from the other techniques which are based on amplitude distribution to the loudspeakers. In contrast, ambisonics is completely free from the loudspeakers. Thus it is a technique that conveys spatial information through the encoded signals. (Jaroszewicz, 2015)

There are lots of researches and applications about how to decode ambisonics signals into stereo formats. The methods of virtual loudspeaker decoding, converting ambisonics to binaural, UFH stereo technique open up the possibilities of reproducing encoded ambisonics signals in commercial audio technology.

It can be assumed that the need of sweet spot and the regular loudspeaker layout are the disadvantages of the system. However the extensions of the system enables to minimize these drawbacks in the decoding process. Speaker alignment algorithms and different optimization tools create flexible work flow for the application of the ambisonics.

Furthermore, the flexibility of implementing head rotations in ambisonics create a base for virtual reality technology and various installation projects. It is possible to compose dynamic spatial scenarios by the extended use of the ambisonic technology.

2.6. Wave Field Synthesis (WFS)

Wave field synthesis (WFS) is the newest sound spatialisation method and it depends on large number of speakers to create a virtual sound field. The technique has lots of advantages when compared to the limitations of the other spatialisation techniques.

By the utilization of WFS, it is possible to recreate an accurate replication of a sound field in two dimensions, by using the principles of Kirchhoff-Helmholtz integrals. The researchers from Delft technical university have developed the basic principles of the method since 1988. The technique is the most accurate application regarding direction and distance of virtual sound objects, without the need of the sweet spot. (Brandenburg, Brix, & Sporer, 2004)

The technique demands big quantity of loudspeakers which bound an area in which sound sources can localized. WFS generates sound fields which represent the actual temporal and spatial properties of the sound sources. Because of the realism of the reproduction, it can be considered as a holophonic method of sound reproduction method. (Roginska & Geluso, 2017)

2.6.1. Basic Theory of WFS

The principle of Huygens states that every point on a wave front is the source of each spherical wavelet.(Blok, Ferwerda, & Kuiken, 1992) Thus, the wave front can be composed of the sum of these spherical wavelets. The explanation to wave propagation in optics can be applied to sound waves as well.

In other words, it was claimed that a propagating wave front of a primary source α can be synthesized by an infinite number of secondary sources, placed on the primary source's wave front as shown in Figure 2.14. (Roginska & Geluso, 2017)

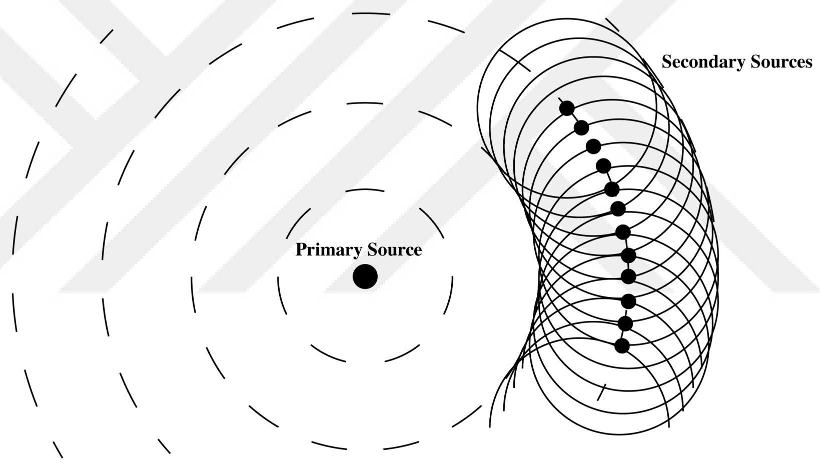

Figure 2.14: Primary source consists of secondary sources

In a mathematical sense the principle states that at time $t + \delta t$, a wave field can be synthesized by replacing the wave front at time t by an infinite number of secondary sources with a small distance δx from each other. At the end, the algorithm is constructed by Rayleigh's representation of the Kirchhoff-Helmholtz integrals and formulated mathematically as:

$$P_A = \frac{1}{4\pi} \int_S \left[\left(P \frac{1 + jkr}{r} \cos\theta \frac{exp(-jkr)}{r} \right) + \left(jwp_0 V_n \frac{exp(-jkr)}{r} \right) \right] dS$$

(Brandenburg et al., 2004)

The method of WFS allows to create a point source at any intended position. The point source is at the location where the propagation start in all directions like any real world situation. However, WFS is implemented just in two dimensional system. It uses the plane waves which are at the intersection of the wave field. So the simplification is occurred at the the field of two dimension (x,y) and the first waves which arrive at the listener point create the localization cues.

As shown in Figure 2.15, the speaker array recreates the virtual wavelets originating from the primary source, resulted in generating the "real" waves. Each speaker is responsible from a portion of the source. Thus, it is vital that all speakers must work together in a synchronized manner.

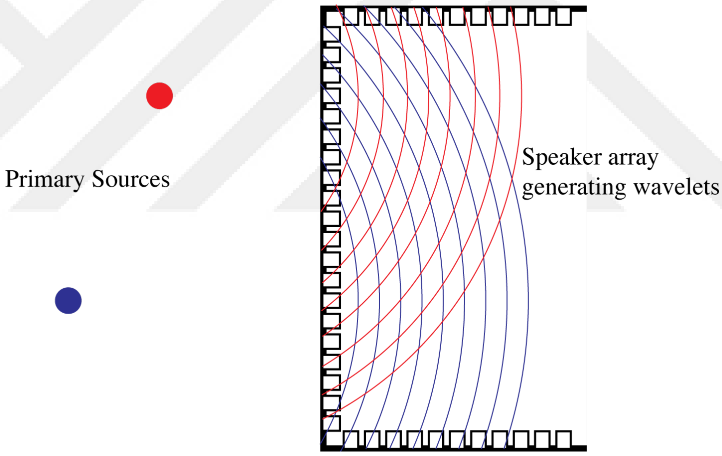

Figure 2.15: Speaker array generating the wavelets

As opposed to Ambisonics, WFS does not stipulate sweet spot. Each listener hears the source from its exact position where radiation begins and observes the physical propagation like occurs in real world.

In order to generate complete real wave field, the distance between speakers should be infinite small according to algorithm of WFS. However, this is impossible to implement and create a limitation for the system. So, there will always be an approximation for the spatialisation. So this approximation is accepted as spatial sampling, like in the case of digital sampling of the acoustic signals. If it is desired to spatialize frequencies between the range of human hearing in a complete accuracy, the distance between two adjacent speakers should be less than 1 centimeter. This situation is not practical because of the need of lots of speakers and the physical constraints of today's speaker technology.

The frequencies which are out of the range of the system's capability generate spatial distortions. Although these frequencies do not contribute to localization process, they still create coloration from the listener's perspective. If the distance between speakers will be 17 centimeters approximately, the spatial distortion begins at around 1000 Hz. The layout containing loudspeakers with 17 centimeters apart is enough to create accurate spatialisation because frequencies between 800-1600 Hz are the most responsible range for the localization cues for a human. (Jaroszewicz, 2015)

2.6.2. The Features of WFS Regarding Applications

The method of WFS allows to localize the sources in an absolute position. Even if the listener moves inside the room, the source is still heard from its absolute position. So, it is possible to create a sense of perspective with the usage of multiple virtual sources.

Furthermore, it is possible to create plane waves which obtain same angle for each listener. This type of localization allows to follow moving listener. This can be used for the room effects.

The sources may be located both inside and outside the speaker array. If the sound sources are inside the speaker array, only the listeners who are in front of the source can perceive the localization of that source accurately.

In addition to flexibility of locating the sources, listeners also can observe the sound field both inside and outside the speaker setup.

As it can be understood, WFS is a very flexible system which can fit in variety of applications such as installations, large concert situations, cinema sound *etc*... The most important disadvantage of the system is the need of large amount of loudspeakers. In addition to this drawback, the system needs high CPU power to implement the algorithm into large quantity of outputs. However, wave field synthesis can be considered as the most sophisticated spatial audio reproduction technique among others. In parallel with the improvement of computer and loudspeaker technology, it is inevitable that WFS will be used in the audio industry more than today.

CHAPTER 3

THE ANALYSIS OF ”WUNDERKAMMER”, ”POINT-INSTANT” AND ”HOLLOW” IN THE CONTEXT OF SPATIAL AUDIO REPRODUCTION TECHNIQUES

Until this chapter, the technical features of some spatial audio reproduction techniques have been presented with their strengths and shortcomings. It is important to choose the right technique to be implemented in order to serve the desired result for the work. In this chapter, I will demonstrate the usage of some spatial audio reproduction techniques implemented in three compositions which I produced between 2018 and 2019 by evaluating their contributions to the compositions.

Although there are some advantages and limitations of each spatial audio reproduction technique, It can be observed from the following pieces that the nature of a particular production and the needs of a artwork may result in some unexpected and creative outputs even if the application seems not logical or wrong. It is important to observe how compositional process and the spatialisation approach can change and shape each other from the following examples.

3.1. "Wunderkammer"

Wunderkammer is a project (ca. 40') which I produced for a specific site (second floor of the music building at Bilkent University) at the end of 2018.

This site-specific sound and light installation/performance is based on my memory of the acoustics of the site perceived as a non-place which refers to transient or anthropological spaces, coined by Marc Augé (Augé, 1992). Since I have been thinking about this project, the most lasting idea is that I should create such a space that I can feel free while I am giving attention to the boundaries.

The piece is written for 16+1 loudspeaker layout, flute, bassoon, trombone and viola. Before starting to work on the piece, it was essential to analyze the features of the site regarding both acoustics and architecture. Natural acoustic features of the site dramatically changes the characteristics of the sound sources. In fact, the shape of the structure, the materials used in the construction and the architectural plan of the site change the perception of localization to some extent. As shown in Figure 3.1, the site includes lots of columns (illustrated with black circles) and open spaces which connect other floors between each other. All these factors constitute an authentic acoustic footprint of the site.

I aimed to create both binaural version of the composition to provide stereo output for commercial release and ambisonics version to provide 16+1 channels output for the performance. In the following sections the process of those two productions will be represented with their interdependence to each other and their contributions to the composition.

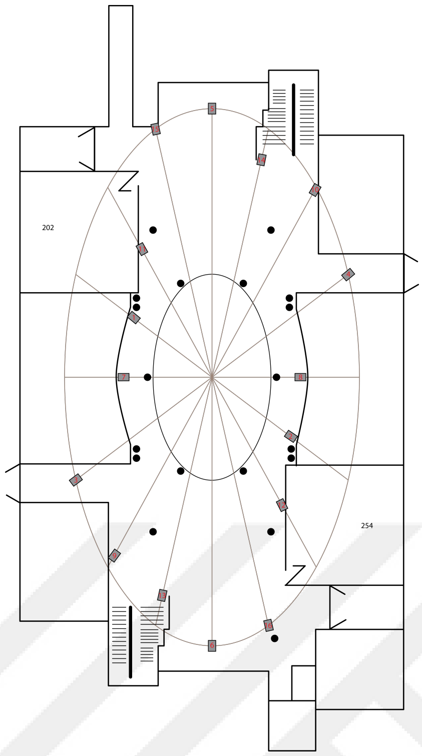

Figure 3.1: Site plan including the positions of the speakers

3.1.1. Generating the Room

In order to perceive the effect of the site's acoustic character which includes early reflections, reverberation, coloration; the aural map of the site should be generated digitally. As mentioned in section 2.2, the definition of the perceived localization increases when the room effects are implied into spatialisation process.

In "Wunderkammer" the effect of the room parameters does not just provide increment in the localization process but also they are used as compositional parameters which act like filtering devices which were applied to sound sources.

It is important to state that the room effect was used just for the binaural version of the production. In the performance, it is not necessary to imply room effects because the site already creates all of the room parameters acoustically.

The idea behind the binaural audio reproduction is that all localization cues are generated for "two big ears" at the center of the sound field. Based on this general definition, the site's response to sound sources should be generated for that "two big ears" in order to use in the composition. In fact, this approach is done literally by creating binaural room impulse response (BRIR) as mentioned in the section 2.2.3. Due to the technical limitations of this particular production, it was not possible to record BRIR of the site in a scientific manner. However, another solution was implemented in order to create three dimensional impulse response of the site digitally.

First of all, the distribution of the loudspeaker layout was decided according to the physical features of the site as shown in Figure 3.1. In the decision process, it was aimed to distribute the loudspeakers as regular as possible in two dimensional space in order to create circle like shape. This approach is important to increase the definition of the spatialisation because of the requirements of the ambisonics implementation which will be presented in the following section. After the placement of the loudspeakers, I recorded one minute sine wave sweep and a sample audio from locations of loudspeakers with a M/S microphone setup, instead of dummy head, located at the center of the site. The recorded sine wave sweeps are used to create impulse responses of the sixteen points of the site. The sample audio was used for the equalization process after the impulse responses created. The process is similar to generate head related transfer function in a different manner. At the end, I ended up with sixteen response impulses at their exact positions in order to generate overall room effect in two dimensional space. The implementation of reverberation effects into the composition will be discussed in the following sections.

3.1.2. Implementation of HOA: Encoding and Decoding Process

It is necessary to present the digital I/O setup of the system in order to understand the encoding process of the piece. Variety of software, which interact with each other, are used in the piece:

1- Ableton Live

2- Max/MSP/Jitter

3- ICST (as a package in Max/MSP/Jitter)

4- HOA (as a package in Max/MSP/Jitter)

5- Altiverb by AudioEase (as a plugin in Ableton Live)

I was able to encode virtual sources into the sound field with 7th order in two dimensional encoding based on the equation presented in the section 2.5.4. According to 7th order in two dimensional space, the system outputs 15 encoded components; so it is applicable with 16 speakers. The sound sources are mixed in Ableton Live and sent to the Max/MSP with the help of Jack Audio software which creates virtual inputs and outputs to communicate between audio software. In Max/MSP first encoding process is done by HOA library [1]. After the decoding operation, decoded signals are sent to the ICST engine [2] in order to apply room effects which are created by impulse responses. The logic behind the application of room effects is that each decoded signal is sent to specific stereo channel containing the reverberation plugin at its corresponding place. In other means, each decoded signal are fed with the reverberation which is generated at the position of that decoded signal. Then, these reverberated signals, obtained from the encoding process of ICST engine, are sent to HOA encoder in order to sum up with the dry signals. At the end, I am able to feed headphones with spatialised output including room effects. At the same time, I am able to decode ambisonics components without room effects. Figure 3.2 and 3.3

[1] HOA is an open-source and free and offers a set of C++ and FAUST classes and implementation for Max, PureData and Unity." (*High Order Ambisonics Library*, n.d.)

[2] ICST is a set of externals for MaxMSP for Ambisonics surround sound processing and source-control in three dimensions

summarizes the all I/O process.

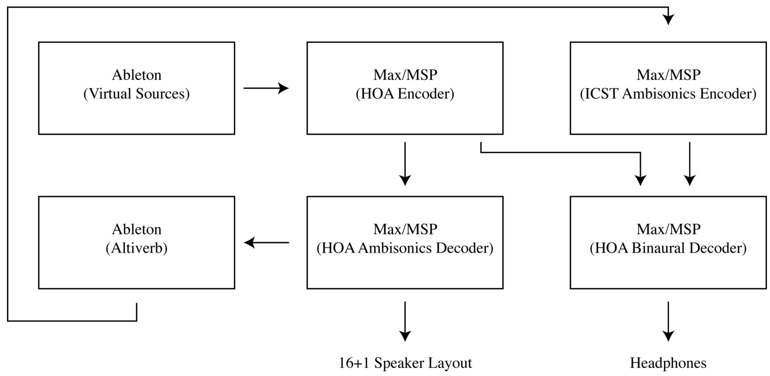

Figure 3.2: I/O diagram of "Wunderkammer"

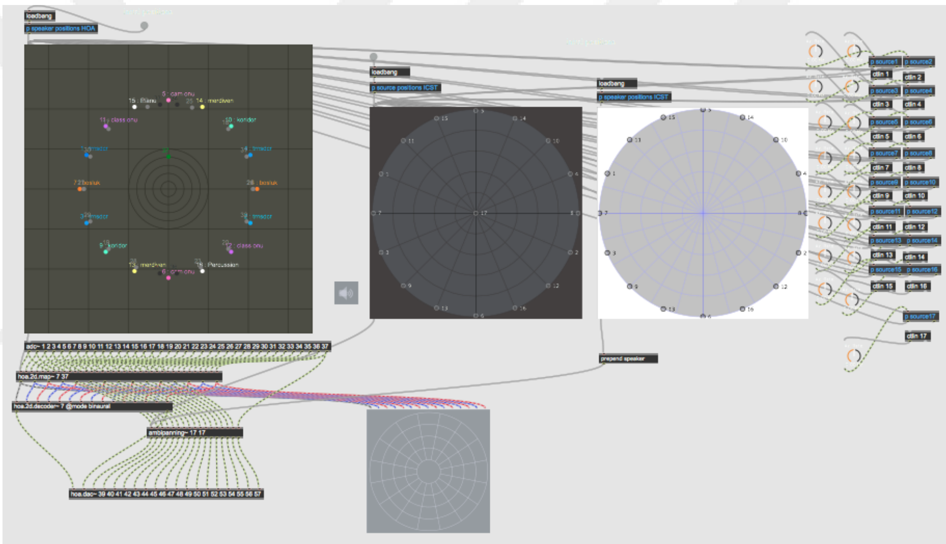

Figure 3.3: Max/MSP Patch of "Wunderkammer"

3.1.3. Evaluation of the Ambisonics Implementation

The most important feature of the ambisonics technique which led me to release two version of the composition is that encoding and decoding process is independent from each other. This allows me to decode both performance version without room effects and binaural version including room effects.

Generating impulse responses with M/S recording method created some disadvantages. It was necessary to make some corrections because of the lack of a dummy head, coloration effects of the altiverb plugin and the quality of the microphones which distorted the reality of early reflections to some extent. In order to compensate this distortion, I tried to implement equalization and time corrections with the help of testing sample.

The optimization procedures could be implemented in order to enlarge the sweet spot and handle low frequencies in a more controlled manner. However, using two different spatialisation engines did not allow me to apply an optimization procedure. In addition to this drawback, the time delay between ICST encoder and the HOA encoder creates an extra pre-delay between dry and wet signals. Still the overall result was satisfying regarding the definition of localization.

3.2. "Point-Instant"

"Point-Instant" is a sound installation piece in collaboration with seven sound artists gathered under the roof of a site specific sound project which is called "Mahal", realized in Müze Evliyagil, Ankara. The performance note of "Mahal" gave an idea of the overall sense of the project:

> ... Even though it contains diverse elements such as musical performance and sound installation, it aims to provide the audience with a collective sound experience. Works, that are formed upon a conceptual basis thought out by seven composers, occupy their own spaces simultaneously in a similar fashion that art objects are placed in a museum. Each work penetrates into the other by interfering with the space of the other. The impossibility of containing sound in a closed space, became a common starting point rather than an obstacle for the project. For a good hour, this collective sound world that penetrates every corner inside and spills outside, aims to add a new layer to the memory that the space has created

> up to now. Unlike a musical concert, Mahal does not require any ideal position to experience. The impossibility of perceiving sounds of Mahal as a whole, encourages listeners to draw their own individual paths, ultimately offering them a unique and subjective experience. (Mahal, 2019)

The project "Point-Instant" was produced for ten transducer speakers attached to the glass located in the stairway of the museum. The physical realities of the place shaped the choice of the spatial audio reproduction technique.

First of all, the stairway does not include any regular shape, so it was not possible to implement any spatialisation method which stipulates a sweet spot. Other than that, the project did not impose an ordinary listening experience which includes stabilized listeners. In addition, the stairway naturally includes an elevation regarding the sound field. Finally, I tried to use material of the construction as a sound producing medium by using transducer speakers. Because of these reasons, I decided to use distance based amplitude panning in order to keep the spatialisation very simple and effective.

3.2.1. Implementation of DBAP in "Point-Instant"

The most substantial aim of "Point-Instant" was to provide a connection between upper and lower floor not just in a morphological manner but also in semantic way. The daily function of the stairway is to create a passing place which function as connecting the floors. DBAP provides a simple panning method which lets me to use the natural elevation of the space.

As it was mentioned in section 2.4 DBAP is a spatialisation method based on the distance data between sound sources and loudspeakers. The absence of the listener in the algorithm of DBAP is cut-out for the implementation of that project because the listeners pass through the stairway. So, the position of the listener should not be in the algorithm.

I attached five transducer speakers at both left and right side of the stairway. At the end, five stereo pairs of transducer speakers were aligned between first floor and the basement as shown in Figure 3.4.

Figure 3.4: 10 transducer speakers attached to the stairway

The spatialisation of the virtual sound sources was applied in Max/MSP/Jitter with an external package, Spat engine which is a real-time spatial audio processor made by IRCAM engineers. After constructing the digital placement of the transducer speakers with their exact distances between each other as shown in Figure 3.5, I applied some extended features to DBAP as mentioned in section 2.4.2. The most important extension is adding "k nearest neighbor" algorithm into DBAP patch. With the help of this feature, I grouped the sound sources into three categories: the focused sounds, spread sounds and the room effects.

While focused sound sources are fed by just three speakers at the same time in order to increase the perception of localization, the spread sounds are panned among five speakers in order to enclose half of the space. In addition, room effects and ambience sounds are fed by all speakers in order to enclose all space. This extension is achieved by a patch as shown in Figure 3.6 which creates an algorithm based on k nearest neighbor theorem.

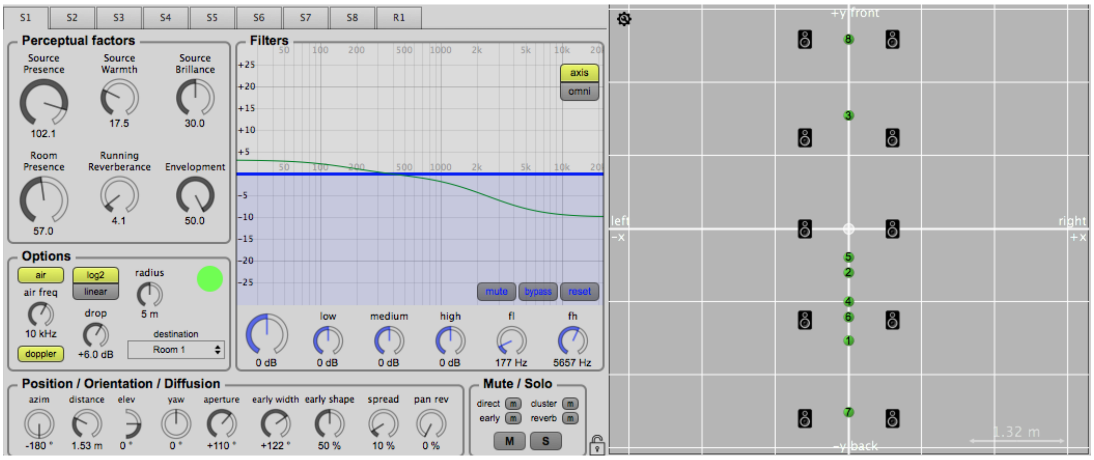

Figure 3.5: The placement of outputs in Spat-Max/MSP

3.2.2. Evaluation of DBAP Implementation

In such a project in which it is impossible to create a regular speaker layout, and there are no stable listeners; DBAP is the most appropriate spatial audio reproduction technique. The simplicity of the algorithm provides a clarity in installation projects like "Point-Instant".

By the help of putting the distance data into localization and the natural spatial elevation of the stairway, the composition find its own way naturally. The most crucial fact about this project is that spatialisation technique is the most core element behind the idea of the composition. Although there are some drawbacks of DBAP such as inability to calculate the gains of the sources which are located outside the field of the loudspeaker setup as discussed in section 3.2.3, it is negligible by creating boundaries within the spatialisation patch.

Furthermore, the feature of "k nearest neighbor" algorithm intrinsically results in effective compositional decisions. This is an evidence that space is a crucial parameter of a musical composition.

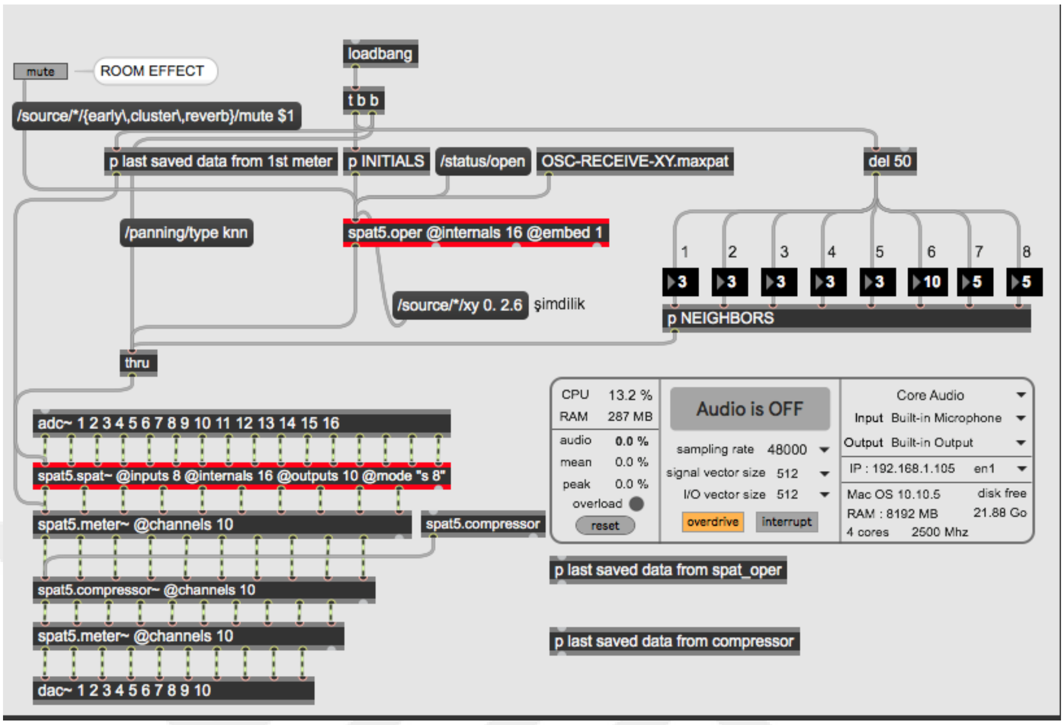

Figure 3.6: Spatialisation patch of "Point-Instant"

3.3. "Hollow"

"Hollow" is a light and sound sculpture which is firstly exhibited at Çağdaş Sanatlar Merkezi as part of an curated exhibition called "Convex". The project is about questioning the dichotomy of "inside-outside" situation by physically creating a boundary between the inner environment of the sculpture and the space which surrounds it.

The binary of inside and outside can not be only perceived as a spatial notion but also should be considered by its semantic aspects. The complexity of this dichotomy is often correlated with the thematization of the body and the world as inner and outer respectively. The questioning of inside/outside, inner/outer, in/out has always created an ambiguity in philosophy as polar elements. By setting apart the philosophical aspects of this topic, I concentrated on the physical perspectives of these polar elements while considering to create an intrinsic interaction between the sculpture and observers by putting the "body" (observers) outside while creating a inside world (sculpture).

Within the process of establishing the concept, the physical elements of the design began to arise. As a result, I built a dodecahedron shaped structure consists of 2nd order three dimensional ambisonic sound design and lighting design by installing one-way mirror, transducer speakers, led, el wire and luminescent fishing line.

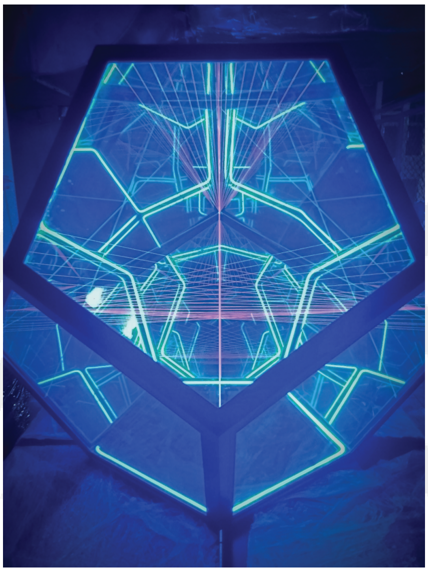

Figure 3.7: "Hollow" - light and sound sculpture

3.3.1. Designing the Sound

The sound design is the single element which creates the motion in the sculpture. I especially gave an attention to balance dynamism and stability among the elements of the design. To create motion in sound composition, I created a three dimensional second order ambisonic spatialisation with ten channel transducer speakers attached to each surface of the structure, ignoring the top and bottom surfaces. The shape let me to use both elevation and azimuth parameters in a perfect symmetry. While constructing the spatial parameters of the sound sources, I purposely used sound materials that correspond the sharp composition of the lighting design. Besides, the nature of the surface speakers creates a sort of filtering effect because of the material and shape of the structure which constitute a correlation between the final sound output and the sculpture itself, like a vocal cords of a human.

In order to construct an immersive environment, I worked on Max/MSP/Jitter with the spatialisation processor of IRCAM, Spat5. In order to create automated trajectories to provide motion in the space; azimuth and elevation control patches are created as shown in Figure 3.8

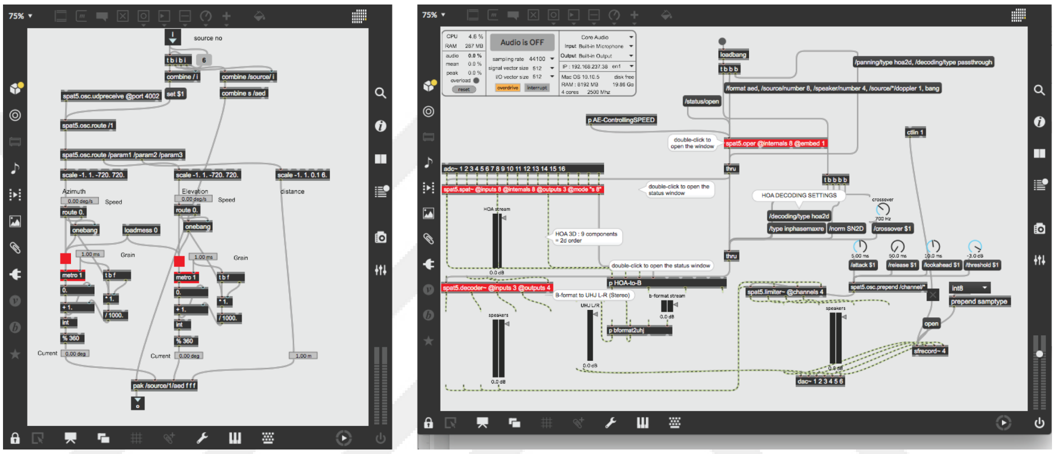

Figure 3.8: Spatialisation patches of "Hollow"

3.3.2. Implementation of HOA in "Hollow"

The layout of ten transducer speakers led me to encode the sound sources in three dimensional space with second order ambisonics. Because of the low order encoding, it is inevitable that the definition of the localization is somewhat blurry. In order to compensate this disadvantage, I applied hybrid optimization which is called inphase-maxre which was discussed in section 2.5.5. This feature of Spat.5 allows to decide a cross-over frequency to handle low and high frequencies in a different manner. With some experimentation and calculation, I decided to decode the frequencies below 700Hz with inphase optimization, the frequencies above 700Hz with maxre optimization procedure. As a result, the blurriness of the localization decreased to some extent.

The Spat.5 has a lot of advantages when compared the other ambisonics tools such as HOA library and ICST. The most prominent feature of Spat.5 is letting the user to decode the encoded signals into several formats at the same time. In the light of this feature I was able to decode the sound field into not just ten channels but also UHJ which is a stereo format derived from ambisonic components. This contributes the piece by means of presenting the inside sound design with stereo format while observing the three dimensional field from outside. It is crucial to state that some flexible features of the spatial sound reproduction technology cause new ideas regarding the design.

Another benefit of Spat.5 is that it has variety of multichannel mixing objects which provide to implement limiter, compression, equalizer within the patch. By the help of these features, it is possible to adjust the sound design anywhere in the process of composing. In order to compensate the sound quality variation between the transducer speakers, I greatly used these mixing devices in order to correct and equalize the final output.

CONCLUSION

As the text proposed, space is an intrinsic part of in the creation of musical artwork. Therefore, the space should be approached carefully more or less within a composition. This thesis discussed recent spatial audio reproduction techniques and their possible roles in a composition.

The most important aspect of spatial audio reproduction methods is that the construction of techniques naturally propose some strong features which should be approached by the demands of a composition. In other words, none of the methods is superior than others. When a composer attempts to implement a digital spatialisation technique, she/he has to evaluate the demands of the composition carefully and try to find the best method that fits into the composition. In fact, it should be stated that the spatialisation process should not be an eclectic attempt in the overall composition. It needs to be handled as one of the principle musical parameters.

The common starting point of all spatialisation methods is to create all localization cues which are the main parameters of human hearing. Furthermore, how we hear and localize the sources is not just a technical issue but it also conveys a meaning through perception. Thus, this semantic asset of spatialisation is a powerful tool regarding composition.

The biggest distinction among spatial audio reproduction techniques is how the localization of the sound sources is approached through the algorithm of the method. Stereo panning types and VBAP propose an amplitude based algorithm depends on the vector geometry. VBAP is even considered as an extension of stereo panning types. DBAP is the single technique which implies distance parameter into its algorithm, among others presented in this thesis. Ambisonics and WFS are considered as sound field generating algorithms. While ambisonics implements the spherical harmonic fields into its algorithm, WFS aims to create each wavelet of a wave. Binaural is a special method which aims to generate the hearing of a head by generating HRTF which is a impulse response applied to the signals in order to localize the sound sources. All reproduction methods stipulate a sweet spot which is an area of listening position in order to perceive the sound field accurately, except DBAP and WFS. In the algorithm of DBAP, there is not a parameter of listener's position. In other words, it is technique which is independent from the location of listener. WFS also does not need a position of listener because it provides an exact simulation of a sound. All techniques provide a different approach to localize sound. It is inevitable that each one has strong and weak features regarding different aspects.

Furthermore, I presented three pieces I produced in order to show how the ideas of these compositions demand different approaches regarding the spatial parameter. In "Wunderkammer", high order ambisonics and binaural method is the most convenient method because of the needs of the project, limitations in technical equipment and the nature of the architecture which are the most important aspects of the production as a site-specific work. In "Point-Instant", I proposed how distance amplitude panning has simple and powerful panning properties when it is not just impossible to create a fixed listening position, but it is also difficult to establish a regular speaker layout. Finally in "Hollow", I showed that how structural design can shape the idea of musical space and its technical implementation through the meaning, although the usage of the spatialisation method seems unpractical.

REFERENCES

Andrew, B. (1989). *Greek musical writings, ii. harmonic and acoustic theory.* Cambridge University Press Cambridge, UK.

Augé, M. (1992). *995 non-places: An introduction to an anthropology of supermodernity.* London: Verso Books.

Bagenal, H. (1927). Influence of buildings on musical tone. *Music & Letters*, *8*(4), 437–447.

Blok, H., Ferwerda, H., & Kuiken, H. K. (1992). Huygen's principle 1690-1990: theory and applications.

Boren, B. (2017). History of 3d sound. In *Immersive sound* (pp. 54–76). Focal Press.

Brandenburg, K., Brix, S., & Sporer, T. (2004). Wave field synthesis: New possibilities for large-scale immersive sound reinforcement. *Fraunhofer IDMT & Ilmenau Tech University, Mo5 E*, *1*.

Britannica, T. E. o. E. (2007, Apr). *Antiphon.* Encyclopædia Britannica, inc. Retrieved from `https://www.britannica.com/art/antiphon-music`

Britannica, T. E. o. E. (2018, Mar). *Panorama.* Encyclopædia Britannica, inc. Retrieved from `https://www.britannica.com/art/panorama-visual-arts`

Carpentier, T., Noisternig, M., & Warusfel, O. (2015). Twenty years of ircam spat: looking back, looking forward..

Chowning, J. (2011). Turenas: the realization of a dream. *Proc. of the 17es Journées d'Informatique Musicale, Saint-Etienne, France*.

Craven, P. G., & Gerzon, M. A. (1977, August 16). *Coincident microphone simulation covering three dimensional space and yielding various directional outputs.* Google Patents. (US Patent 4,042,779)

Davis, M. F. (2003). History of spatial coding. *Journal of the Audio Engineering Society*, *51*(6), 554–569.

Fantel, H. (1981, Jan). *Sound; taking note of the 100th anniversary of stereo.* The New York Times. Retrieved from `https://www.nytimes.com/1981/01/04/arts/sound-taking-note-of-the-100th-anniversary-of-stereo.html`

Farnell, A. (2010). *Designing sound.* Mit Press.

Griesinger, D. (2002). Stereo and surround panning in practice. In *Audio engineering society convention 112.*

Hacihabiboglu, H., De Sena, E., Cvetkovic, Z., Johnston, J., & Smith III, J. O. (2017). Perceptual spatial audio recording, simulation, and rendering: An overview of spatial-audio techniques based on psychoacoustics. *IEEE Signal Processing Magazine*, *34*(3), 36–54.

HERR, C. R., & SIEBEIN, G. W. (1998). 146" souped-up" and" unplugged. In *Proceedings of the association of collegiate schools of architecture annual meeting* (p. 146).

High order ambisonics library. (n.d.). Retrieved from `http://hoalibrary.mshparisnord.fr/en`

Howard, D., Moretti, L., & Moretti, L. (2009). *Sound and space in renaissance venice: architecture, music, acoustics.* Yale University Press New HavenLondres.

Jaroszewicz, M. (2015). *Compositional strategies in spectral spatialization* (Unpublished doctoral dissertation). UC Riverside.

Johnstone, M. A. (2013). Aristotle on sounds. *British Journal for the History of Philosophy*, *21*(4), 631–648.

Kostadinov, D., Reiss, J. D., & Mladenov, V. (2010). Evaluation of distance based amplitude panning for spatial audio. In *Icassp* (pp. 285–288).

Lossius, T., Baltazar, P., & de la Hogue, T. (2009). Dbap–distance-based amplitude panning. In *Icmc.*

Mahal. (2019). *Mahal - sound installation.*

Malham, D. G., & Myatt, A. (1995). 3-d sound spatialization using ambisonic techniques. *Computer music journal*, *19*(4), 58–70.

Mitchell, D. (2009). *Basicsynth*. Lulu. com.

Møller, H. (1992). Fundamentals of binaural technology. *Applied acoustics*, *36*(3-4), 171–218.

Pérez-López, A. (2014). Real-time 3d audio spatialization tools for interactive performance. *Master's thesis, Universitat Pompeu Fabra, Barcelona, Spain.*

Pulkki, V. (2001). *Spatial sound generation and perception by amplitude panning techniques*. Helsinki University of Technology.

Rasmussen, S. E. (1964). *Experiencing architecture* (Vol. 2). MIT press.

Reznikoff, I. (2008). Sound resonance in prehistoric times: A study of paleolithic painted caves and rocks. *Journal of the Acoustical Society of America*, *123*(5), 3603.

Roginska, A., & Geluso, P. (2017). *Immersive sound: The art and science of binaural and multi-channel audio*. Taylor & Francis.

Snow, W. B. (1953). Basic principles of stereophonic sound. *Journal of the Society of Motion Picture and Television Engineers*, *61*(5), 567–589.

Stern, R. (2006). *Binaural sound localization, chapter in computational auditory scene analysis, g. brown and del. wang, eds.* New York: Wiley/IEEE Press.

Zhang, W., Samarasinghe, P., Chen, H., & Abhayapala, T. (2017). Surround by sound: A review of spatial audio recording and reproduction. *Applied Sciences*, *7*(5), 532.

Zvonar, R. (1999). *A history of spatial music. richard zvonar.*