SPATIAL SOUND DESIGN AND PERCEPTUAL EVALUATION OF MOVING SOUND SOURCES IN AMPLITUDE PANNING

Author: David Caro Moreno
Supervisor: Alexander Raake
Tutors: Hagen Wierstorf, Climent Nadeu

Master thesis (PFC)
Telecommunication Engineering, ETSETB-UPC (Barcelona)
Carried out at Technische Universität Berlin
October 2015

Abstract

Music is organized sound. Usually, we think on organizing sound in frequency and time, but sound may be also organized in space. Today, musicians and scientists have a new artistic and scientific challenge to face, the creation of spatial music. Spatial music is music which involves the organization of sound in the three-dimensional space.

So far, the availability of sound reproduction systems with height is very limited. Most popularly, we find stereo and surround systems, which use panning techniques to reproduce sound. From stereo to surround, a very important step further is the use of a scene-based paradigm, which opens the possibilities of music composition beyond the classical stereo image.

In this research, a spatial sound system is proposed to create complex virtual sound scenes consisting of several moving sound sources. On the other side, we study the perceptual characteristics of moving sound sources in Vector Based Amplitude Panning, which is one of the most widespread spatial audio technologies for loudspeaker-based setups. Finally, we perform listening tests with a group of subjects to investigate the perceived attributes of moving sound sources reproduced using amplitude panning.

Resum del projecte

La música és so organitzat. En general, l'organització de so és en freqüència i temps, però el so pot ser també organitzat en l'espai. Avui en dia, músics i científics tenen un nou repte artístic i científic, la creació de música espacial. La música espacial és música que involucra l'organització de so en l'espai tridimensional.

Fins al moment, la disponibilitat de sistemes de reproducció de so amb altura és molt limitada. Més popularment, ens trobem amb sistemes estèreo i so envoltant, que utilitzen tècniques de “panning” per reproduir el so. De l'estèreo al so envoltant, un pas endavant molt important és l'ús d'un paradigma basat en escenes, el qual obre les possibilitats de composició musical més enllà de la imatge estèreo clàssica.

En aquesta investigació, es proposa un sistema de so espacial per a la creació d'escenes de so complexes que consten de diverses fonts de so en moviment. D'altra banda, s'estudien les característiques perceptives de fonts de so en moviment per a “Vector Based Amplitud Panning”, que és una de les tecnologies d'àudio espacial més esteses per a configuracions basades en altaveus. Finalment, portem a terme proves d'escolta amb un grup de subjectes per investigar els atributs percebuts de fonts de so en moviment reproduïdes mitjançant “amplitude panning”.

Resumen del proyecto

La música es sonido organizado. En general, la organización de sonido es en frecuencia y tiempo, pero el sonido puede ser también organizado en el espacio. Hoy en día, músicos y científicos tienen un nuevo reto artístico y científico, la creación de música espacial. La música espacial es música que involucra la organización de sonido en el espacio tridimensional.

Hasta el momento, la disponibilidad de sistemas de reproducción de sonido con altura es muy limitada. Más popularmente, nos encontramos con sistemas estéreo y sonido envolvente, que utilizan técnicas de "panning" para reproducir el sonido. Del estéreo al sonido envolvente, un paso adelante muy importante es el uso de un paradigma basado en escenas, que abre las posibilidades de composición musical más allá de la imagen estéreo clásica.

En esta investigación, se propone un sistema de sonido espacial para la creación de escenas de sonido complejas que constan de varias fuentes de sonido en movimiento. Por otra parte, se estudian las características perceptivas de fuentes de sonido en movimiento para “Vector Based Amplitud Panning”, que es una de las tecnologías de audio espacial más extendidas para configuraciones basadas en altavoces. Finalmente, llevamos a cabo pruebas de escucha con un grupo de sujetos para investigar los atributos percibidos de fuentes de sonido en movimiento reproducidas usando "amplitude panning".

Acknowledgments

First of all, I would like to thank Hagen Wierstorf for productive discussions that always stimulated my work on this thesis and Climent Nadeu for his guidance and support before and during the mobility program.

I thank the whole TU staff for the friendly working environment and to Alexander Raake and Jens Ahrens for their inspiring suggestions. Special thanks go to Chris Clock, who contributed to the development of the Soundscape Renderer control interface.

I am very grateful for the possibility of using the extensive facility and equipment at Technische Universität Berlin. The experiments would not have been possible without participants. I therefore thank all members of the listening panels for their participation, as well as the TU Berlin for absorbing the expense allowance.

Finally, I would like to express my gratitude to my parents Antonio and Maria Dolores, as well as my sister Marta, my girlfriend Banu and my friends Pere and Yusuf for their ongoing support and motivation.

Contents

List of Figures

List of Tables

Chapter 1

INTRODUCTION

1.1 Research frame

This project investigates spatial sound generation in the context of music production and its reproduction with loudspeakers for multiple listeners. This musical context not only involves concerts, but also cinema, theater, performance, sound installations and even video-games, virtual reality and other working or commercial environments.

In the majority of these scenarios, audio reproduction systems are mono, stereo or surround systems, and in the case of multi-channel systems, sound is positioned in space using amplitude panning. More advanced spatial audio reproduction techniques are theoretically possible, although all of them require a large number of loudspeakers, turning to be very expensive to built and hence rarely found in practice. This research hereby addresses sound spatialization with amplitude panning, aiming to enhance today's sound experiences. All the same, it is pointed out that the next generation of spatial systems may combine panning with other spatial audio reproduction techniques. Thus, this investigation shall be useful for the future as well.

On the other hand, in order to progress in the challenging problem of spatial audio reproduction, it is essential to dispose of the proper materials to test new reproduction systems. This requires to conceive spatial composition techniques and to design spatial sound scenes with these techniques. The production of convenient research materials presents a problem because most audio production tools are two-channel-based, being hardly suitable for larger setups. Solutions for handling complex auditory scenes are very limited. For this reason, a system to create spatial music using various state-of-art audio tools is devised and proposed as a standard sound scene design methodology.

1.2 Aims of the project

The aim of this study is to serve professionals who in these days are working with spatial sound. An important step forward is the separation of sound composition and sound reproduction. This can be possible using an object-oriented approach. Inevitably, the final reproduction system limits musicians in their process of creation. The object-oriented approach makes possible to separate the production and reproduction of spatial audio into different processes. This opens a lot of possibilities for composers. Ideally, spatial music production tools are object-oriented and include basic functions to control virtual audio scenes in real-time. For instance, functions to create complex movements of sound objects, to assign behavioral roles, to model and modify sources' size, shape,...

After a virtual sound scene is sent to a spatial audio rendering software, it is processed with rendering algorithms for being adapted to a certain speaker layout. In order to support the development of state-of-art sound reproduction systems and novel methods, we need to gain a better understanding on the field of human auditory perception [1]. The focus of this project is then on the psychoacoustic study of the present generation of spatial audio reproduction systems and the perceptual evaluation of complex auditory scenes.

1.3 Facilities and Resources

1.3.1 Laboratory and Equipment

This research has been conducted at the Pinta 3D audio laboratory of Technische Universität Berlin. The reverberation time of the room is nearly frequency independent from 10 kHz to 160 Hz in the range of 0.1 s to 0.15 s. Below 160 Hz, the reverberation time reaches 0.225 s maximum at 63 Hz. The noise level of the room is below GK25 with air condition turned on [2].

The room is equipped with a reference circular loudspeaker array installation consisting of 56 speakers and a subwoofer, suitable for Vector Based Amplitude Panning, Ambisonics, Wave Field Synthesis and some more unusual options. All algorithms are using different subsets of the same speaker setup [2]. The used loudspeaker system is described in detail in [3].

Figure 1. Circular loudspeaker array of the 3D audio lab

1.3.2 Software

The software used for real-time spatial audio reproduction is the Soundscape Renderer (SSR). The SSR is a free software in current development at Technische Universität Berlin, providing the most widespread spatial audio rendering algorithms. For instance, binaural (HRTF-based) simulation, Vector Base Amplitude Panning (VBAP), Ambisonics Amplitude Panning (AAP), Wave Field Synthesis (WFS) and some more peculiar options. According to the SSR manual, "*the SSR is intended as versatile framework for the state of the art implementation of various spatial audio reproduction techniques*" [4]. It only supports two-dimensional reproduction and this is an important drawback, but it is only a matter of time to have a comparable tool for three-dimensional reproduction.

The SSR runs under GNU/Linux and Mac OSX. The compatibility with Mac OSX makes possible to combine it with most of the commercial digital audio workstations used nowadays by music producers. A key feature of the SSR is the IP network interface. It enables remote control with any type of interaction tool via a TCP socket. The messages are sent in XML format and terminated with a binary zero.

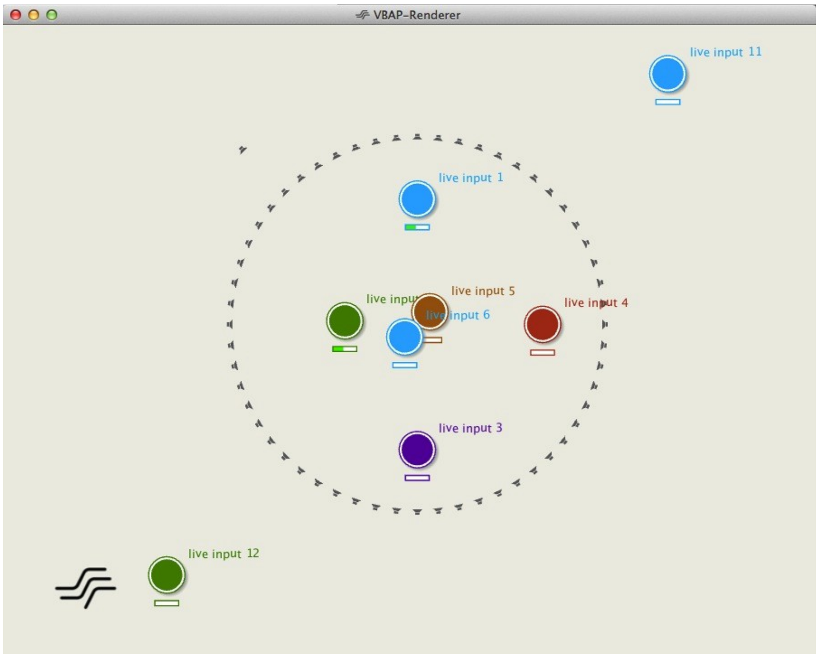


Figure 2. Soundscape Renderer GUI

The SSR can read audio and metadata. The metadata is embodied in the Audio Scene Description Format (ASDF) file. The ASDF standard is being developed at Technische Universität Berlin as well. The description format offers the possibility to specify the speaker layout. So far, it only implements static features and that means in the current state it is not possible to describe e.g. movements of a virtual sound source [5]. However, the IP interface enables control over source positions and thus the description of movements may be done from a remote software.

The tool used for the generation of sound is Ableton Live, one of the most popular digital audio workstations for music production. Nevertheless, any other DAW's are suitable for sound generation. Audio streams are routed to the SoundScape Renderer via the Jack Router tool, which facilitates internal audio routing between any audio applications. The positioning of sound objects and creation of trajectories is done with Max for Live, a version of Max/Msp integrated in Ableton Live. Pure Data is the free alternative to Max/Msp.

1.4 Organization of contents

The next section documents the characteristics and design of a digital musical system for real-time spatial sound composition. The system is object-oriented and enables audio reproduction in different loudspeaker setups, with a variety of rendering algorithms. The section describes how to proceed in order to connect all the tools indicated previously.

Following the design of the system, we review the Vector Based Amplitude Panning method (VBAP) and the current knowledge on the perception of moving sound sources. We study localization and coloration of moving sound sources in VBAP. Finally, we identify the problems of amplitude panning for both properties and describe how these properties are enhanced or deteriorated when varying the number of active loudspeakers.

The last part of this work discusses the choice of the number of loudspeakers for the reproduction of sound sources in movement with VBAP. This choice may be influenced by the trade-off between sound localization and sound coloration properties. In the future of sound reproduction, one easily would tend to anticipate the addition of more speakers. Nevertheless, when using a different number of active speakers, some sound characteristics are improved and others are degraded, as in the case of timbre coloration. The overall quality of a spatial sound reproduction system is not only determined by the quality of the reproduction of spatial attributes, but also the quality of the rest of sound characteristics. Accordingly, the evaluation of the perceptual listening tests looks for possible answers to questions like:

- How many speakers shall we use for a music performance?
- Which speaker layout is better for specific sound choreographies?
- Will the addition of more speakers improve the quality of the listening experience?

Chapter 2

REAL-TIME SPATIAL SOUND SYSTEM DESIGN

2.1 Spatial sound system overview

In a spatial sound system, a virtual audio scene is constructed by distributing a set of sound objects in a certain space. Audio files or audio streams are associated to sound objects. Each sound object represents a one-channel audio signal and it is accompanied by its metadata. The metadata contains information like source position, source size, etc. Then, spatial renderization algorithms transform audio signals using their metadata into signals ready for playback in multi-channel loudspeaker setups. The spatial renderer provides audio signals for every loudspeaker. Figure 3 shows a schematic representation of a spatial audio system as described in [5]:

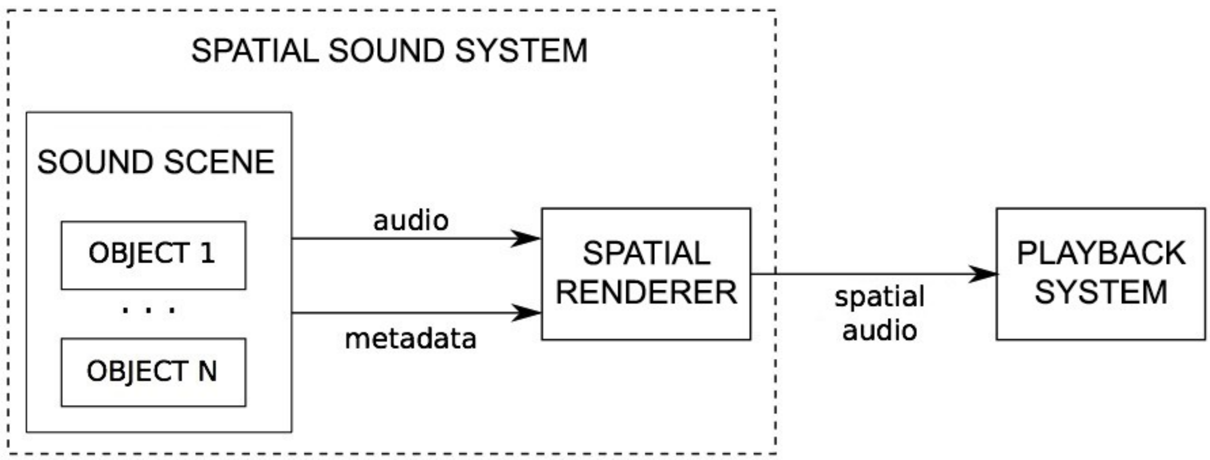


Figure 3. Representation of a spatial sound system [5]

In the system used for the realization of this project, sound scenes, composed of a set of sound objects, are created in Ableton Live. The metadata is constructed with Max for Live, indicating positions and trajectories of each sound object. Objects are distributed in an horizontal plane. Then the SSR renders audio with the rendering algorithm chosen upon execution of the software and sends it to the output for its reproduction in the Pinta 3D audio lab. Figure 4 represents this system:

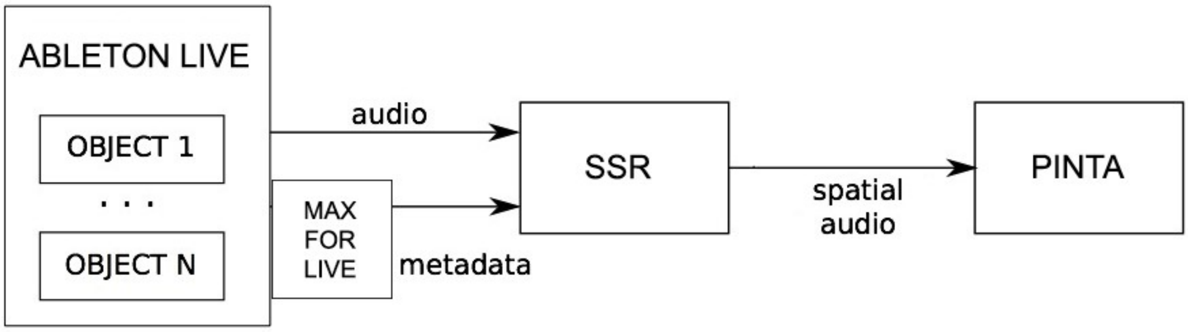


Figure 4. Representation of the spatial sound system used in this work

The set of tools used in this work to experiment with creating spatial music could be used out of the lab by any professional musician without previous experience on spatial audio. The first module of the spatial sound system is Ableton Live, one of the most widespread tools for music production. This music software was introduced with the idea to transform computers in instruments for live music performance on stage. The function of this module is to create the sound elements of the virtual auditory scene. Nonetheless, other programs like Logic Pro, Cubase, etc. could be used as well.

The metadata used to describe the movements of the virtual sound sources is created in Max for Live, which is a version of Max/Msp integrated in Ableton Live (see section 2.3). It is possible to use also Max/Msp or Pure Data together with Ableton Live or other digital audio workstations.

The outputs of any music production software chosen are individual audio streams which can be routed through the Jack Router free tool to a spatial rendering software like the SoundScape Renderer. Jack Router enables virtual buses for communication between audio softwares running in the same computer. The SoundScape Renderer receives audio streams and metadata to position the audio in a space as sound objects.

2.2 Spatial sound system parameters

Next, we present two sets of parameters, corresponding to the modules shown in Figure 3 and 4. One set of parameters is related to the production of a virtual auditory scene and the other set to its reproduction.

2.2.1 Virtual auditory scene control parameters

A fundamental characteristic of spatial sound is the position of sound sources. In a basic implementation of a spatial sound system, we have at least a 2-dimensional position for each sound source. More complex implementations introduce modeling of sound sources and room simulation parameters, increasing dramatically the dimensionality of these systems [5]. The management of such multi-dimensional systems is not an easy task and most commercial softwares available are not adequately prepared. It is necessary to design new spatial sound tools.

Marshall [6] proposed the following classification of possible control parameters:

- Sound position: two or three-dimensional coordinates (cartesian or polar)
- Source characteristics: size, directivity, material...
- Environmental and room model: reverberation, doppler effect, air absorption and distance decay,...

The SSR implemented 2D positioning of sound sources with cartesian coordinates. In case polar coordinates are preferred, it is possible to do the calculations with polar coordinates in Max for Live or the software used to create the metadata and transform them to cartesian coordinates before sending them to the SSR.

The SSR provides virtual point sources and plane waves. It also adds an option for doppler effect. We miss characteristics like source size, source directivity, material and the environmental and room model. However, some characteristics like source material may be emulated in Ableton Live or the corresponding music production software. Also, the room simulation may be produced with a combination of point sources or plane waves and external multi-channel reverberation algorithms.

2.2.2 Spatial sound system features

In addition to parameters related with the spatial nature of sound, there are some relevant system characteristics which are desirable in a spatial renderer software [5]:

- Speaker layout configuration
- Number of sound sources
- Support for various technologies

The SSR has implemented all these features. The speaker layout configuration is limited to 2D speaker arrangements. The number of sound sources is not limited by the SSR itself but by the CPU load. Even so, the SSR is multi-core-able, hence it is possible to increase the number of used CPUs to render lots of sources. The SSR covers the most important spatial audio technologies and additionally some less common options.

2.3 Control interface

The control interface of the real-time spatial audio system used in this work is shown below in Figure 5. The interface enables control over some of the parameters listed in section 2.2.1. It has been designed to be used in live performance. It is possible to configure up to eight different movements and to control up to 8 sound objects. An additional feature is the control of the reference coordinate system, which is equivalent to moving all sound sources at once. Also, it includes a preset system to save movement configurations and trigger them at any time. The use of presets may release part of the cognitive load during live performance.

Figure 5. Interface for controlling the movement of sound objects

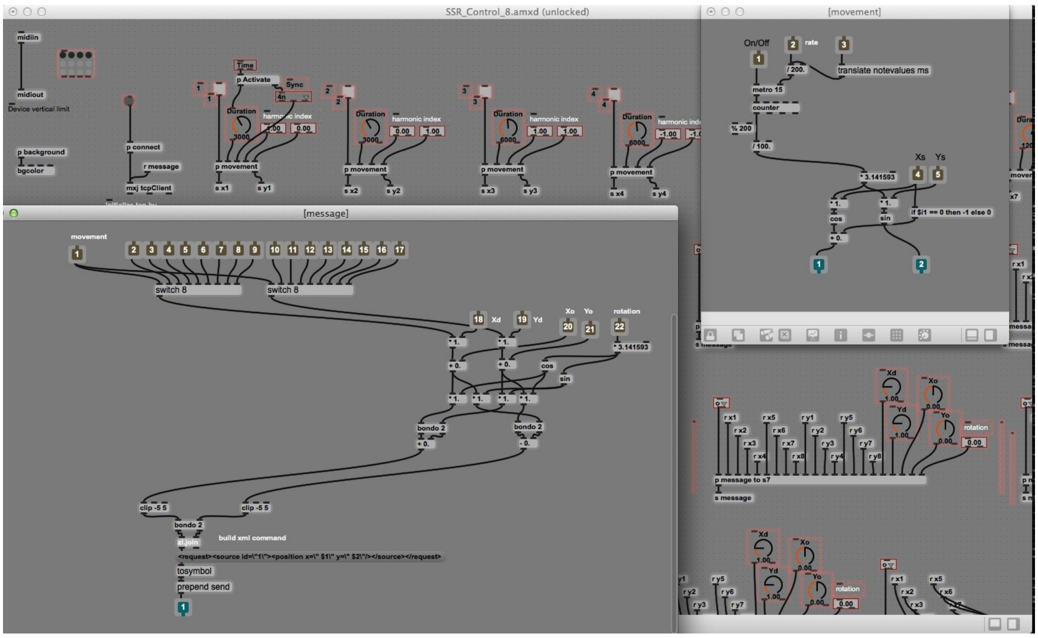


Figure 6. Control interface patch programming in Max for Live

The control interface has been developed with Max for Live. Max for Live is a graphic programming environment integrated inside Ableton Live. With Max for Live is possible to create and send metadata for each sound object. It is a practical choice in our case, although there are other comparable alternatives like Pure Data, which is a free open software.

2.4 Spatial composition techniques

Common spatial composition techniques encompass the creation of trajectories and sound choreographies, the diffusion or distribution of sound energy, room acoustics simulation by adding reverberation and echoes, enhancing acoustics by exciting specific resonances of a space, alluding to specific environments by using certain sounds,... [7]

2.4.1 Creating trajectories and sound choreographies

The development of sound choreographies is certainly an important extension of the compositional task. Spatial music or music done for being played in a spatial sound reproduction system calls for a meaningful use of the space. There is more to say, more space to fill and more space to travel.

Music has been often defined as the organization of sound in time. Some examples found in [8] are:

- Music is the organization of sound and time. Elliott Schwartz (composer)

- Music is nothing else but wild sounds civilized into time and tune. Thomas Fuller, History of the Worthies of England (1662), 'Musicians'

- Geometry in time. Arthur Honegger, I am a Composer (1951)

We can also find this idea in most definitions of music on dictionaries:

- An art of sound in time that expresses ideas and emotions in significant forms through the elements of rhythm, melody, harmony, and color. (http://dictionary.reference.com/)

- The science or art of ordering tones or sounds in succession, in combination, and in temporal relationships to produce a composition having unity and continuity. (http://merriam-webster.com/)

–The art of arranging sounds in time so as to produce a continuous, unified, and evocative composition, as through melody, harmony, rhythm, and timbre. (http://www.thefreedictionary.com/)

Spatial music opens new dimensions in music, where sound is organized not only in time, but also in space. To be properly complete, spatial music is sound organized in time, frequency and space. As suggested by the definition of Arthur Honegger, the organization of sound in time is geometric. It is well known that the organization of sound in frequency is also geometric. Then, it could not be in another way, the organization of sound in space shall be geometric as well.

The principles of the harmonograph [9] are a good option to start, given the structural relation of these movements with the harmonic structure of sound. The organization of sound in frequency is represented with the harmonograph in the form of beautiful spatial movements (see Figure 7). These have been implemented in the control interface developed in this work to create and organize movements. Some of these movements are shown in Figures 7, 8 and 9.

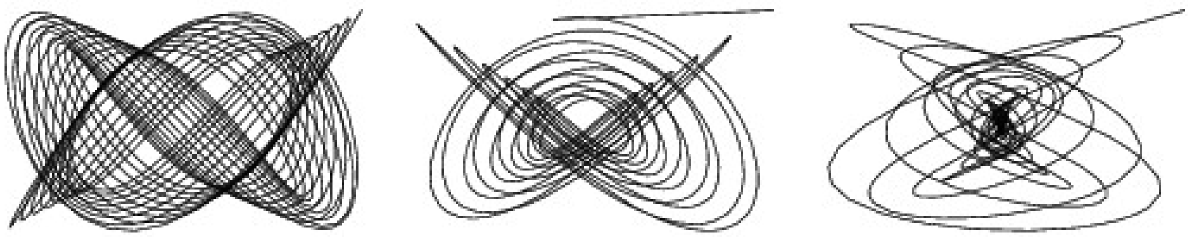


Figure 7. Harmonograph patterns [10]

The movements of the harmonograph are described by the parametric equations of the Lissajous curves:

$$x(t) = A\cos(\omega_x t - \delta_x)$$
$$y(t) = B\cos\left(\omega_y t - \delta_y\right),$$

which are sometimes also written in the form [11]:

$$x(t) = a\sin(\omega t + \delta)$$
$$y(t) = b\sin t.$$

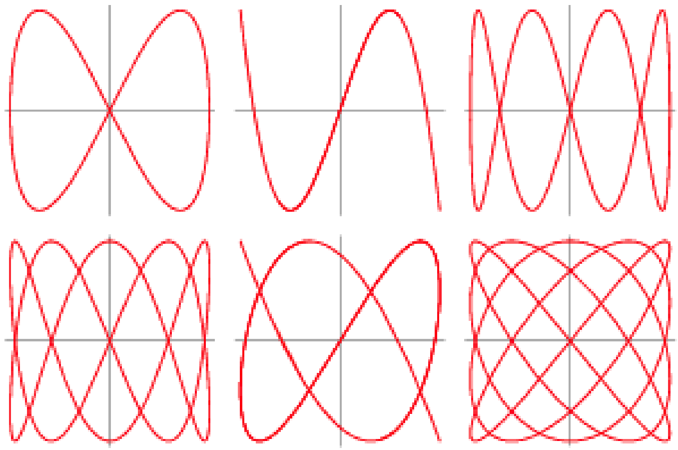


Figure 8. Lissajous Curves [11]

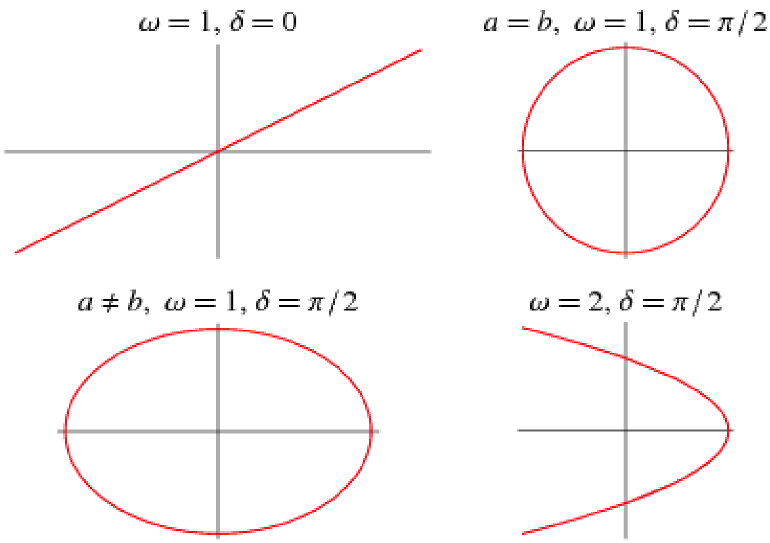


Figure 9. Special Lissajous Curves [11]

These movements may be a good tool not only for composing spatial music but also for designing complex virtual sound scenes for the perceptual evaluation of spatial sound technology.

In addition, basic behaviors are implemented as well in this spatial movement control tool. For instance, any sound source may be set to follow the movement of a leader source, with some possible variations.

2.4.2 Multi-channel reverberation for room simulation

A good reverberation is crucial for creating the feeling of reality, the feeling of space. Feeling the space is a critical part of spatial sound. With a good reverb is possible to transport listeners to an imaginary or a real space. If the reverb is not good, all attempts will lack of credibility. However, how to implement a reverberation in a multi-channel system is a question with no clear answer yet. The emulation of reverberation is computationally expensive. Far from emulating reality, we aim to convince the audience.

Creating a ‘convincing’ spatial reverberation is primarily dependent on the diffusion of the first reflections. From practice, even if this is done more or less random, it will greatly contribute to the experience of the reverberation. To approach a more natural sounding and realistic room simulation, it is needed to process pre-delays, filters and levels for each speaker output that take into account the virtual position of each sound source and the distance of each reflecting surface of a virtual room [12].

The reverberation density is another key parameter for obtaining a convincing spatial reverberation. In case of using existing reverberation algorithms prepared for stereophonic reproduction systems, we have to adjust reverberation density. Compared to a stereophonic reverberation, an artificial reverberation system with more than two channels will be less dense in individual channels, so the total reverberation radiated by the speakers is equally dense.

The reverberation we implemented in the system used in this work is an extension from a stereophonic reverberation to a 4 channel reverberation. The four channels are point sources and they are distributed in the horizontal plane, equidistant from the sweet spot and forming a square. These point sources are more distant from the sweet spot than the rest of sound sources. The distance is long enough so the reverberated sound reaches the listener approximately in the form of wavefronts, hence reverberation is not felt as coming from a focused point.

Chapter 3

MOVING SOUND SOURCES IN AMPLITUDE PANNING

3.1 Fundamentals

Amplitude panning, also known as intensity panning, is the simplest and most popular method to create spatial sound scenes. In this method, the same audio signal is reproduced with different amplitudes in two or more loudspeakers, creating the illusion of a virtual sound source positioned between them. The perceived direction of the virtual sound source depends on the relation of amplitudes radiated by the speakers. This relation is represented in Figure 8.

The panning technique is the most widespread spatial audio technology due to its low requirements in computational effort and its flexibility in terms of number of loudspeakers. [1] However, comparing it to other alternatives, a major drawback is the existence of the sweet spot. For this reason, there is a great interest in developing other technologies which try to reconstruct an entire area of the physical sound field, although they require much higher computational complexity and number of loudspeakers.

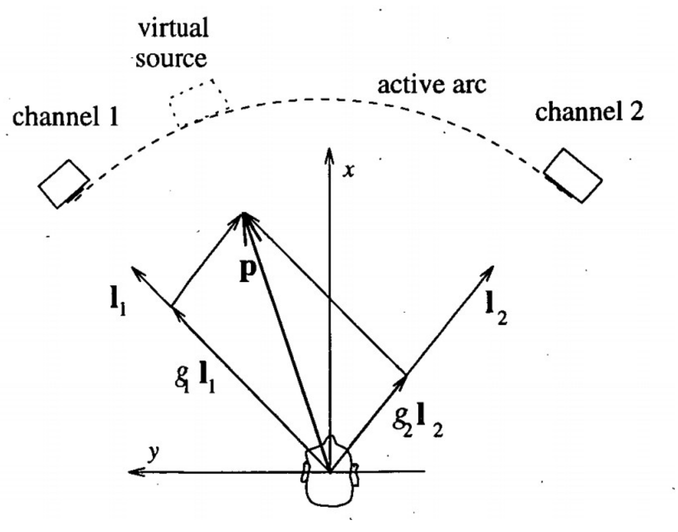


Figure 10. Stereophonic configuration formulated with vectors [13]

In the Vector Base Amplitude Panning formulation, a vector base is defined by the loudspeaker configuration, with the origin in the sweet spot and the vectors pointing toward each speaker. A direction vector ***p*** = *[p1 p2]* pointing to the virtual sound source is derived from the vector base, with the gain factors corresponding to the loudspeakers as vector components. These gain factors of the channel levels have to be adjusted to maintain a constant loudness when moving virtual sources [13]. The panning formula is given by the linear combination of loudspeaker vectors:

$$\mathbf{p} = g_1 l_1 + g_2 l_2$$

Typically, amplitude panning has been applied in 2-channel stereophonic reproduction systems, although it may be applied to multi-channel loudspeaker setups with more than two loudspeakers as a reformulation of the existing pair-wise panning method [14]. The extension of VBAP to a system with more than two loudspeakers in the horizontal plane is realized by dividing the plane in non-overlapping arcs, which are defined by the set of loudspeaker pairs, as shown in Figure 9. Then, the audio signal is reproduced by the pair of speakers that define the active arc inside which the sound source is located.

In such loudspeaker configurations, it is possible to create a spatial sound scene around the listener. Virtual sound sources may be positioned anywhere in the circle formed by the non-overlapping arcs, making possible to locate a source in any horizontal direction. To create depth, we can adjust the level intensity and direct-to-reverberation ratio of sound sources, thus making them to appear closer or farther from the listener [1]. Therefore, virtual sound sources may be arranged anywhere in the horizontal plane.

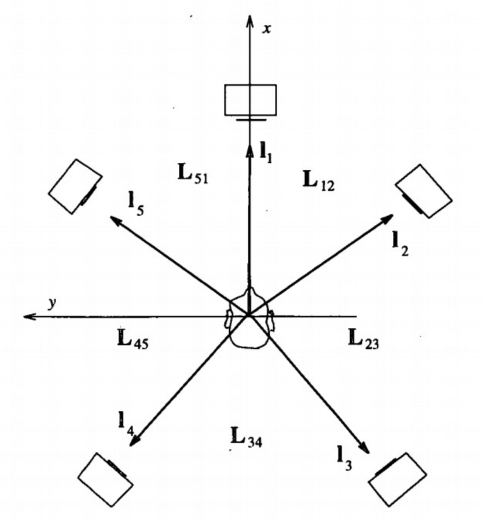


Figure 11. Plane division by non-overlapping arcs [13]

3.2 Localization of moving sound sources

The perceived direction of a sound source is determined by differences between the signals arriving to the ears. These are the so called Interaural Level Difference (ILD) and Interaural Time Difference (ITD). ILDs dominate the localization at high frequencies and ITDs at low frequencies. Still and all, the human hearing is very complex and these are only useful simplifications. Furthermore, the human spatial hearing is individual and varies from person to person [14].

In panning methods, there are no time differences between the signals emanating from the loudspeakers, but only level differences. Nonetheless, at low frequencies the level differences between the loudspeakers create ITDs [15]. Those differences are the basis of the summing localization theory, that explains the creation of phantom sound sources [16]. Phantom sources are virtual sound sources that seem to appear from a direction between the speakers where there is no physical source.

Therefore, we have two possible situations. When a sound source is located in the direction of an existing physical source, this sound will be reproduced by only one speaker. If the direction of the sound source doesn't match with the direction of any speaker, this sound will be reproduced by combining intensities in the two closest loudspeakers. Consequently, during the reproduction of a moving sound source, there may be a change in the number of active loudspeakers used to reproduce it, from one to two loudspeakers and vice versa. The gain factors of the channel levels have to be properly controlled to keep the energy of a sound source constant.

Together with the direction of sound sources, other properties are modified unintentionally, like the sound coloration, which will be discussed in the next section, and the apparent source width. These sound properties appear to modulate for a moving sound source. The apparent width of a phantom source is dependent of the loudspeaker aperture angle and the panning position. When the panning position is closer to the middle, the increment in the perceived source width gets to a maximum [15]. For a moving sound source, the apparent source width may increase and decrease continuously, as the sound source is moving and reproduced by one or two loudspeakers. This affects the perceptual localization of sound source trajectories, as sources with a narrower apparent width seem to be more directional.

Given a surrounding loudspeaker configuration, a challenging problem is the creation of stable lateral phantom images. The loss of symmetry respect to the listener for non-frontal speakers (see Figure 12) blurs localization of phantom sources. It has been suggested that the simulation of early reflection patterns may be a sufficient enhancement for source localization [17], although this is computationally expensive, as commented in section 2.4.2 in the study of multi-channel reverberation for room simulation.

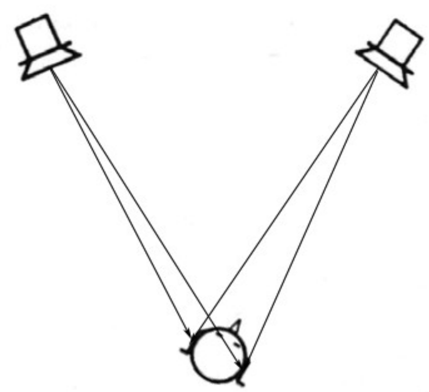


Figure 12. Loss of symmetry for non-frontal phantom sound source positioning

3.3 Coloration of moving sound sources

A side effect of the phantom source technique is the spectral distortion of the source. When two loudspeakers radiate identical signals at different distances from the listener, a delayed version of the signal is superimposed to the first one. This causes a comb filter effect, coloring sound with a series of spectral peaks and valleys.

The spectral distortion resulting from the comb filter effect would not be present in a physical sound source situated at the same position as the phantom source. Even though, the comb filter effect at the ears does not influence the perception of direction nor distance [18]. This does not mean that sound coloration is not relevant. On the contrary, it has been found that timbral fidelity is of high relevance for the overall quality of multi-channel stereophonic reproduction systems using an object-oriented paradigm. According to Rumsey, quality is determined a 70% by timbral fidelity and a 30% by spatial fidelity [19].

In a typical stereophonic configuration, the loudspeakers are equidistant to the listener, although they are not equidistant to the listener's ears. When listening monaurally, i.e. covering one ear, a big change in the color of the phantom source is perceived (see Figure 13). The distortion resulting from monaural hearing disappears when listening binaurally. This phenomenon is called binaural decoloration and it is explained by the association model [15].

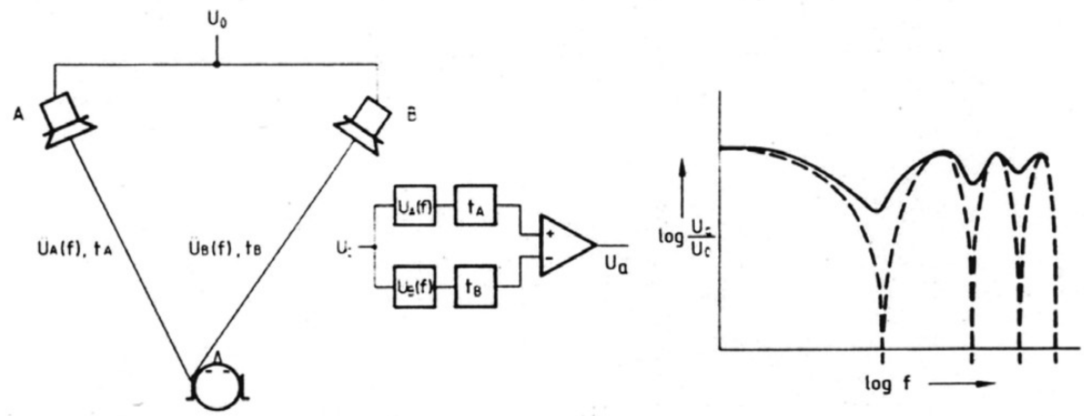


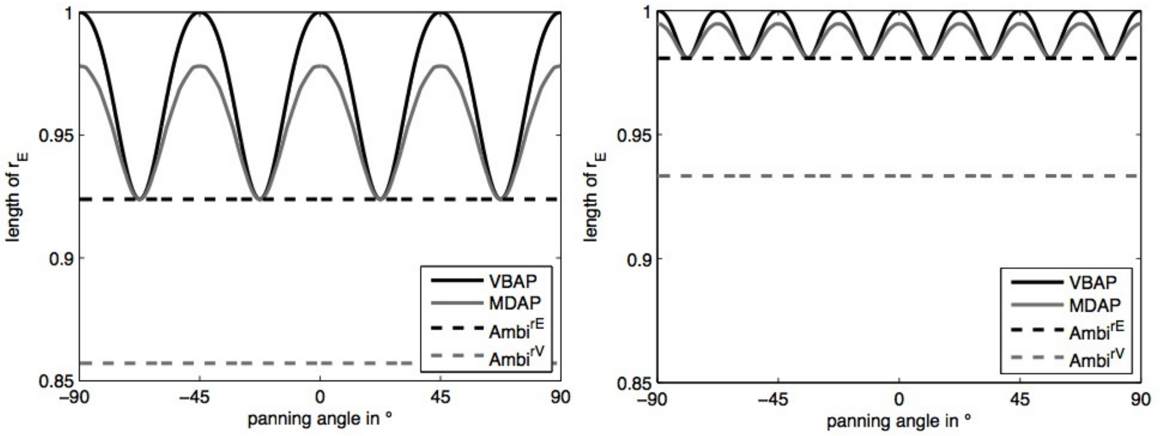
Figure 13. Comb filter effect [18]

An average untrained listener may not detect sound coloration, but will easily be perturbed by a continuous change in sound color. In the case of a moving sound source, the effect of a comb filter is variable and much more disturbing. Sound color may change slow or fast. It has been observed during this research that faster changes in coloration may cause the listener a stronger unnatural feeling, even when they are smaller changes in terms of spectral intensity (see Figure 14 in next section). This phenomenon can not be measured directly with equipment as it belongs to the terrain of human auditory perception. To that end, the perceptual evaluation of moving sound sources through listening tests performed in the next chapter will provide a practical perspective.

3.4 Varying the number of loudspeakers

The extension of two-channel stereophony to systems with more than two channels like surround was the natural path to create larger spatial images, which are not limited to the frontal dimension. The use of a higher number of loudspeakers, though, involuntarily brings more problems due to sound coloration produced by the unwanted comb filter effect. The quality of sound reproduction systems not only depends on spatial fidelity, but timbral fidelity is of high relevance. For a moving sound source, as the number of loudspeakers increases, its sound color changes faster, producing a very disturbing effect on the original sound.

The direction and distribution of energy is observed in the energy vector, which is considered a localization model at higher frequencies [15]. The fluctuations of the energy vector length shown in Figure 14 correspond to the number of active loudspeakers. The energy vector is 1 every time a single loudspeaker is active. Comparing these two cases, we observe the fluctuation of the energy vector to be deeper but slower for the case with 8 speakers and vice versa for 16 speakers. Matthias Frank stated in [15] that the fluctuations of the energy vector length are a predictor for the fluctuations in timbre.



(a) panning with 8 loudspeakers, $\Delta\phi_L = 45°$ (b) panning with 16 loudspeakers, $\Delta\phi_L = 22.5°$

Figure 14. Fluctuation of the energy vector length [15]

Adding more speakers we can broaden sound spatiality at the expense of deteriorating sound timbre. Sound coloration is hard to notice by untrained people, but a change in sound color is much easier to detect. In the same way, trained people may be more disturbed by a faster change, which causes a very unreal, undesired effect. In the experiments conducted during this research, we focus on this aspect of amplitude panning. We analyze how this effect is perceived by the human hearing for different numbers of active loudspeakers. In addition, we study the localization property as well, in order to contrast the data obtained.

Therefore, the choice of the number of loudspeakers for a certain application should take into account this drawback to avoid notorious coloration artifacts and achieve a decent sound quality. This choice should be derived from the balance of localization and coloration properties. Distance cues are another very important spatial sound property, but they will be left apart as they are reproduced with techniques like adjusting sound intensity level and direct-to-reverberation ratio, which are independent of the number of loudspeakers [1].

To sum up, increasing the number of loudspeakers benefits the perceptual localization of sources' trajectories but at some point affects coloration in a very undesirable way. This work is an attempt to defy the intuitive idea of adding more and more speakers to future spatial sound reproduction systems based on amplitude panning.

Chapter 4

PERCEPTUAL EVALUATION OF MOVING SOUND SOURCES IN AMPLITUDE PANNING FOR DIFFERENT NUMBERS OF LOUDSPEAKERS

4.1 Experiment setup

All listening tests of this research have been conducted in the 3D audio lab described in section 1.3.1. The tests are pairwise comparisons on different loudspeaker setups, with different number of active speakers. The conditions examined are setups with 4, 8, 14, 28 and 56 speakers. Figure 15 and Figure 16 show the loudspeaker configuration with 4 active speakers and the loudspeaker configuration with 8 active speakers respectively.

Test subjects sit in a central position. A computer is placed in front of them, showing instructions in the screen to guide them during the experiment. An answer sheet is provided to the subjects for indicating their answers. Subjects are forced to choose one of two conditions presented in each comparison, A or B.

Figure 15. Speaker layout for 4 and 8 active loudspeakers, respectively

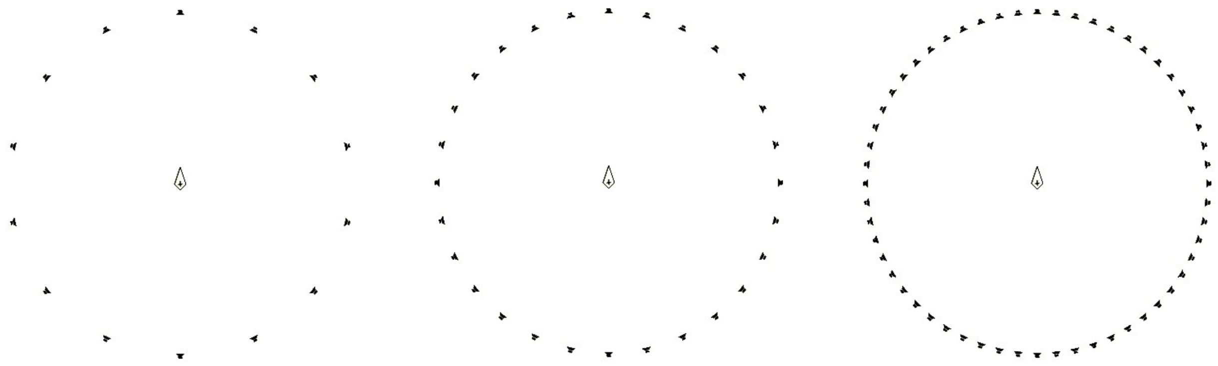


Figure 16. Speaker layout for 14, 28 and 56 active loudspeakers, respectively

4.2 Stimuli and conditions

At first, the experiments were intended to evaluate spatial sound attributes with reproducing complex virtual sound scenes consisting of several synthetic sound sources moving in a horizontal plane. This was not practical because subjects would not be familiar with the sounds, and to evaluate their properties would be a hard task. Also, when including several sources in the virtual scene, we would not know where the subjects would focus their attention.

Finally, it has been decided to use as a stimulus in all experiments one virtual sound source reproducing pink noise at a moderate level and moving in a horizontal plane. The broadband noise ensures good localization performance [20] and it is also good for detecting changes in coloration at any part of the spectrum. The stimulus were rendered for its reproduction with 4, 8, 14, 28 and 56 speakers. The same stimulus were used for localization and coloration comparisons. The order of the presented comparisons was randomized.

Various conditions were compared: a slow circle trajectory, a fast circle trajectory, a slow square trajectory and a fast square trajectory. All trajectories were placed inside and close to the loudspeaker array. With VBAP it is not really possible to place virtual sources inside the listening area. Consequently, as the paths are shorter, the movement of the source appears to be faster. Nonetheless, we can create the effect of a sound source “in your face” by increasing its intensity.

The reason to include two types of trajectories is that a circle is much easier to localize in a circular loudspeaker array than a square. This is because the distance from the virtual source and the speakers respect to the listener varies in a circular array for a square trajectory, contrary to the case of a circle trajectory, where there is no change in distance.

4.3 Test panel of the first experiment

The test panel of the first experiment was formed by 10 subjects. The number of trained and untrained listeners, male and female subjects was balanced (6 male, 4 female, 5 trained and 5 with no experience with spatial audio reproduction systems). The age of the subjects was between 22 and 51, with an average of 30 years.

An interesting detail noticed is that almost all male subjects were trained listeners and female subjects untrained. By trained subjects we understand those who are familiar with localization and coloration attributes of sound and does not need an explanation to clarify the topics of the test.

4.4 Experiment procedure

In the first listening test, subjects were asked to answer to the questions:

- For which condition you can better localize the trajectory?
- For which condition the sound color is less changing during movement?

The following instructions were presented to the subjects previously to the first experiment realization:

In this experiment the abilities of the human ear in localizing the trajectory of moving sound sources and detecting changes in their sound color are examined under different conditions. For this purpose, you will be placed on a chair in the middle of a circular speaker array. Then, we will reproduce noise-like test signals through the system. Sound sources will be moving around you in circle or square trajectories.

The experiment is composed of 56 paired comparisons divided in two parts of 28 comparisons. For each comparison, two conditions will be presented to you, A and B. An answer sheet will be provided to you for marking your answers.

In the first 28 comparisons your task is to decide for which of the two conditions you can better localize the trajectory.

In the second part, you will be presented an example of changes in sound color as a previous training. Next, in the following 28 comparisons, your task is to decide for which of the conditions you perceive less changes in sound coloration.

Thank you for your participation and have fun.

4.5 On the evaluation of the validity of the experiment

We evaluate the consistency of the results by looking to the cumulative distribution function (CDF) of the binomial distribution of the subjects' preferences. The localization and the coloration tests consist in 28 paired-comparisons. First we proceed by calculating the values of the probability mass function, Pr(X=x), for a binomial distribution with n=28 and p=0.5, which are shown in Table 1:

Pr(X=0)	3.7252903E-009	Pr(X=15)	0.1394829154
Pr(X=1)	1.0430813E-007	Pr(X=16)	0.1133298688
Pr(X=2)	1.4081597E-006	Pr(X=17)	0.0799975544
Pr(X=3)	1.2204051E-005	Pr(X=18)	0.0488873944
Pr(X=4)	7.6275319E-005	Pr(X=19)	0.0257302076
Pr(X=5)	0.0003661215	Pr(X=20)	0.0115785934
Pr(X=6)	0.0014034659	Pr(X=21)	0.0044108927
Pr(X=7)	0.0044108927	Pr(X=22)	0.0014034659
Pr(X=8)	0.0115785934	Pr(X=23)	0.0003661215
Pr(X=9)	0.0257302076	Pr(X=24)	7.6275319E-005
Pr(X=10)	0.0488873944	Pr(X=25)	1.2204051E-005
Pr(X=11)	0.0799975544	Pr(X=26)	1.4081597E-006
Pr(X=12)	0.1133298688	Pr(X=27)	1.0430813E-007
Pr(X=13)	0.1394829154	Pr(X=28)	3.7252903E-009
Pr(X=14)	0.1494459808		

Table 1. Probability mass function of binomial distribution

Next, we calculate the values of its cumulative distribution function (CDF) and their complementaries, which are shown in Table 2. Then, we consider that a subject is answering according to the hypothesis if 18 answers are good, and that a subject answers randomly or disagrees with the hypothesis if less than 18 answers are good. Taking as successes those preferences matching the hypothesis, the probability that a subject gets 18 successes or more than 18 successes with answering randomly the 28 comparisons is given by the complementary value of the CDF:

$$1 - CDF(X=18) < 5\%$$

CDF(X=0)	3.7253E-009	1-CDF(X=0)	0.9999999963
CDF(X=1)	0.000000108	1-CDF(X=1)	0.999999892
CDF(X=2)	1.5162E-006	1-CDF(X=2)	0.9999984838
CDF(X=3)	1.3720E-005	1-CDF(X=3)	0.9999862798
CDF(X=4)	8.9996E-005	1-CDF(X=4)	0.9999100044
CDF(X=5)	0.0004561171	1-CDF(X=5)	0.9995438829
CDF(X=6)	0.001859583	1-CDF(X=6)	0.998140417
CDF(X=7)	0.0062704757	1-CDF(X=7)	0.9937295243
CDF(X=8)	0.0178490691	1-CDF(X=8)	0.9821509309
CDF(X=9)	0.0435792767	1-CDF(X=9)	0.9564207233
CDF(X=10)	0.092466671	1-CDF(X=10)	0.907533329
CDF(X=11)	0.1724642254	1-CDF(X=11)	0.8275357746
CDF(X=12)	0.2857940942	1-CDF(X=12)	0.7142059058
CDF(X=13)	0.4252770096	1-CDF(X=13)	0.5747229904
CDF(X=14)	0.5747229904	1-CDF(X=14)	0.4252770096
CDF(X=15)	0.7142059058	1-CDF(X=15)	0.2857940942
CDF(X=16)	0.8275357746	1-CDF(X=16)	0.1724642254
CDF(X=17)	0.907533329	1-CDF(X=17)	0.092466671
CDF(X=18)	0.9564207233	1-CDF(X=18)	0.0435792767
CDF(X=19)	0.9821509309	1-CDF(X=19)	0.0178490691
CDF(X=20)	0.9937295243	1-CDF(X=20)	0.0062704757
CDF(X=21)	0.998140417	1-CDF(X=21)	0.001859583
CDF(X=22)	0.9995438829	1-CDF(X=22)	0.0004561171
CDF(X=23)	0.9999100044	1-CDF(X=23)	8.99956E-005
CDF(X=24)	0.9999862798	1-CDF(X=24)	1.37202E-005
CDF(X=25)	0.9999984838	1-CDF(X=25)	1.51619E-006
CDF(X=26)	0.999999892	1-CDF(X=26)	0.000000108
CDF(X=27)	0.9999999963	1-CDF(X=27)	3.72529E-009
CDF(X=28)	1	1-CDF(X=28)	0

Table 2. Cumulative distribution function and complementary values

4.6 Results of the first experiment

The first experiment was interrupted after 10 subjects performed the test. The results were positive for the coloration hypothesis, but not at all in the case of localization. After realizing the tests, subjects were asked about their impressions and a possible problem was detected in the localization test. As there were circle and square trajectories but it was not specified previously to the subjects which shape they should expect, it could be possible that they think about a square trajectory when a circle was presented with a low number of active speakers. Some subjects reported that the cases of 4 and perhaps 8 active loudspeakers might be confusing due to the geometry of the speakers' position.

The results of the first experiment are presented in Tables 3, 4, 5 and 6. Results that strongly contradict the hypothesis are highlighted in grey:

n speakers	localization		coloration	
A – B	A	B	A	B
4 – 8	1	9	5	5
8 – 14	3	7	8	2
14 – 28	3	7	10	0
28 – 56	8	2	5	5
8 – 28	5	5	8	2
8 – 56	5	5	8	2
4 – 56	1	9	10	0

Table 3. Test summary for a sound source moving slow in circles

n speakers	localization		coloration	
A – B	A	B	A	B
4 – 8	3	7	5	5
8 – 14	6	4	6	4
14 – 28	4	6	8	2
28 – 56	3	7	7	3
8 – 28	5	5	8	2
8 – 56	6	4	7	3
4 – 56	1	9	8	2

Table 4. Test summary for a sound source moving fast in circles

n speakers	localization		coloration	
A – B	A	B	A	B
4 – 8	3	7	7	3
8 – 14	5	5	9	1
14 – 28	5	5	10	0
28 – 56	5	5	7	3
8 – 28	7	3	7	3
8 – 56	3	7	8	2
4 – 56	1	9	8	2

Table 5. Test summary for a sound source moving slow in squares

n speakers	localization		coloration	
A – B	A	B	A	B
4 – 8	2	8	7	3
8 – 14	6	4	8	2
14 – 28	7	3	9	1
28 – 56	5	5	10	0
8 – 28	2	8	9	1
8 – 56	3	7	9	1
4 – 56	2	8	9	1

Table 6. Test summary for a sound source moving fast in squares

Besides the contradictions in the results of the localization test, we observe the preferences on coloration to be more consistent with the hypothesis. Results for the coloration test support the idea that for a higher number of speakers the change in sound color is less preferred. We also notice that when comparing 4 to 8 speakers the preferences of the subjects are less clear in general than for the rest of cases. These makes sense according to the idea that amplitude panning is a technology designed for setups with a low number of loudspeakers.

Next, the method of paired comparisons presented by Kendall and Smith [21] is followed to build tables again with the results of the test. The last column is the mean of the total number of preferences for each number of speakers. The mean is calculated in order to be able to compare the values. These values should be higher for the more preferred conditions.

If results are consistent with the hypothesis, the values in the last column of each table should be decreasing (up to down) for the results on localization and increasing for the results on coloration. This is right for some of the localization tables and for all the coloration tables:

	56	28	14	8	4	mean
56		2		5	9	5.3333
28	8		7	5		6.6667
14		3		7		5
8	5	5	3		9	5.5
4	1			1		1

Table 7. Results for the localization of a sound source moving slow in circles

	56	28	14	8	4	mean
56		7		4	9	6.6667
28	3		6	5		4.6667
14		4		4		4
8	6	5	6		7	6
4	1			3		2

Table 8. Results for the localization of a sound source moving fast in circles

	56	28	14	8	4	mean
56		5		7	9	7
28	5		5	3		4.3333
14		5		5		5
8	3	7	5		7	5.5
4	1			3		2

Table 9. Results for the localization of a sound source moving slow in squares

	56	28	14	8	4	mean
56		5		7	8	6.6667
28	5		3	8		5.3333
14		7		4		5.5
8	3	2	6		8	4.75
4	2			2		2

Table 10. Results for the localization of a sound source moving fast in squares

	56	28	14	8	4	mean
56		5		2	0	2.3333
28	5		0	2		2.3333
14		10		2		6
8	8	8	8		5	7.25
4	10			5		7.5

Table 11. Results for the coloration of a sound source moving slow in circles

	56	28	14	8	4	mean
56		3		3	2	2.6667
28	7		2	2		3.6667
14		8		4		6
8	7	8	6		5	6.5
4	8			5		6.5

Table 12. Results for the coloration of a sound source moving fast in circles

	56	28	14	8	4	mean
56		3		2	2	2.3333
28	7		0	3		3.3333
14		10		1		5.5
8	8	7	9		3	6.75
4	8			7		7.5

Table 13. Results for the coloration of a sound source moving slow in squares

	56	28	14	8	4	mean
56		0		1	1	0.6667
28	10		1	1		4
14		9		2		5.5
8	9	9	8		3	7.25
4	9			7		8

Table 14. Results for the coloration of a sound source moving fast in squares

In the localization hypothesis, we found that only half of the subjects agreed with the hypothesis and the other half answered more or less randomly, perhaps because of the problem in the localization test with identifying the shape of the trajectory.

One of the subjects answered by crossing instead of circling the answers and his answers are mostly opposing to the hypothesis, which might not be a coincidence. Except this subject, in the coloration test, all trained subjects agreed with the hypothesis in 24 of 28 comparisons or more, and one agreed in all comparisons. Untrained subjects agreed with the hypothesis in 18 of 28 comparisons or more and one agreed in 27 of 28 comparisons.

4.7 Test panel of the second experiment

The panel for the second test was formed by 12 subjects. The number of trained and untrained listeners, male and female subjects was balanced (7 male, 5 female, 6 trained and 6 with no experience with spatial audio reproduction systems). The age of the subjects was between 18 and 31, with an average of 26 years.

Again, almost all male subjects were trained and female subjects untrained.

4.8 Modifications on the experiment procedure

In the second listening test, two modifications were introduced. First, the shape of the trajectory was presented to the subjects previously to reproducing each condition in order to avoid possible confusions, as explained in section 4.6.

Second, the question regarding sound coloration was changed because the question of the first test, “For which condition the sound color is less changing during movement?”, is a question about the change in sound color and it was preferred to ask generally about sound color, including “how much sound color” and “how much change in sound color” in one question.

Then, subjects were asked to answer to the questions:

- For which condition you can better localize the trajectory?
- For which condition the sound coloration is less disturbing?

The following instructions were presented to the subjects previously to the second experiment realization:

In this experiment the abilities of the human ear in localizing the trajectory of moving sound sources and detecting changes in their sound color are examined under different conditions. For this purpose, you will be placed on a chair in the middle of a circular speaker array. Then, we will reproduce noise-like test signals through the system. Sound sources will be moving around you in circle or square trajectories.

The experiment is composed of 56 paired comparisons divided in two parts of 28 comparisons. For each comparison, two conditions will be presented to you, A and B. Instructions for following the test will be shown in a screen placed in front of you. An answer sheet will be provided to you for marking your answers.

In the first 28 comparisons your task is to decide for which of the two conditions you can better localize the trajectory. In the screen instructions, you will see before hearing the conditions which trajectory is going to follow the sound source – circle or square. The trajectory will be indicated also in the answer sheet.

In the second part there is a previous training. Before the comparisons, you will be presented an example of pink noise and another 2 examples of pink noise with a small and a large change in sound coloration. This change is an unwanted artifact. Next, in the following 28 comparisons, your task is to decide for which of the conditions the sound coloration is less disturbing.

Thank you for your participation and have fun.

4.9 Results of the second listening test

One subject reported that he was answering which condition is more disturbing instead of less. At first, this would not affect the results as we only had to invert the results of this particular subject.

The results of the second experiment are presented in Tables 15, 16, 17 and 18. Results that strongly contradict the hypothesis are highlighted in grey:

n speakers	localization		coloration	
A – B	A	B	A	B
4 – 8	1	11	7	5
8 – 14	6	6	9	3
14 – 28	6	6	10	2
28 – 56	5	7	8	4
8 – 28	4	8	11	1
8 – 56	2	10	10	2
4 – 56	4	8	11	1

Table 15. Second test summary for a sound source moving slow in circles

n speakers	localization		coloration	
A – B	A	B	A	B
4 – 8	4	8	7	5
8 – 14	3	9	8	4
14 – 28	7	5	10	2
28 – 56	4	8	10	2
8 – 28	3	9	10	2
8 – 56	4	8	10	2
4 – 56	3	9	9	3

Table 16. Second test summary for a sound source moving fast in circles

n speakers	localization		coloration	
A – B	A	B	A	B
4 – 8	3	9	9	3
8 – 14	3	9	11	1
14 – 28	4	8	9	3
28 – 56	5	7	8	4
8 – 28	3	9	10	2
8 – 56	3	9	10	2
4 – 56	1	11	10	2

Table 17. Second test summary for a sound source moving slow in squares

n speakers	localization		coloration	
A – B	A	B	A	B
4 – 8	3	9	6	6
8 – 14	4	8	10	2
14 – 28	3	9	9	3
28 – 56	7	5	9	3
8 – 28	6	6	10	2
8 – 56	5	7	10	2
4 – 56	6	6	10	2

Table 18. Second test summary for a sound source moving fast in squares

Again, the results for the localization test are not clear enough. However, in this second test we observe an interesting detail. It seems that localizing trajectories is not so easy when a sound source is moving fast than when it moves slow. Perhaps this is because we are limited in the processing of aural information. Faster movements require much more attention to be localized. Also, technology is limited in the processing of data and for faster movements is less accurate.

In the coloration test, the results are similar to what we got in the first test. Subjects preferred the conditions with less speakers, in accordance with the hypothesis. In the comparisons between 4 and 8 speakers, the preferences are spread more or less evenly. Again, this suggests that coloration artifacts get worst for setups with high number of loudspeakers.

Next, the following tables show again the results of the second test in the way Kendall and Smith proposed [21]. A quick reminder: if results are consistent with the hypothesis, the values in the last column of each table should be decreasing (up to down) for the results on localization and increasing for the results on coloration.

	56	28	14	8	4	mean
56		8		12	8	9.3333
28	6		8	10		8
14		6		6		6
8	2	4	6		11	5.75
4	4			1		2.5

Table 19. Results for the localization of a sound source moving slow in circles

	56	28	14	8	4	mean
56		8		8	10	8.6667
28	4		5	9		6
14		7		9		8
8	4	3	3		8	4.5
4	4			4		4

Table 20. Results for the localization of a sound source moving fast in circles

	56	28	14	8	4	mean
56		8		11	11	10
28	6		8	10		8
14		4		9		6.5
8	3	4	3		9	4.75
4	1			3		2

Table 21. Results for the localization of a sound source moving slow in squares

	56	28	14	8	4	mean
56		5		7	8	6.6667
28	7		11	6		8
14		3		8		5.5
8	5	6	4		9	6
4	6			3		4.5

Table 22. Results for the localization of a sound source moving fast in squares

	56	28	14	8	4	mean
56		5		2	1	2.6667
28	9		2	2		4.3333
14		12		4		8
8	12	12	10		5	9.75
4	11			7		9

Table 23. Results for the coloration of a sound source moving slow in circles

	56	28	14	8	4	mean
56		2		2	4	2.6667
28	10		2	2		4.6667
14		10		4		7
8	10	10	8		5	8.25
4	10			7		8.5

Table 24. Results for the coloration of a sound source moving fast in circles

	56	28	14	8	4	mean
56		3		2	2	2.3333
28	10		3	2		5
14		9		1		5
8	12	12	11		3	9.5
4	10			9		9.5

Table 25. Results for the coloration of a sound source moving slow in squares

	56	28	14	8	4	mean
56		3		2	3	2.6667
28	9		3	2		4.6667
14		11		2		6.5
8	10	10	10		6	9
4	11			6		8.5

Table 26. Results for the coloration of a sound source moving fast in squares

This time, the results of the localization test are in more support of the hypothesis for slow moving sources as they are for the fast moving sources. The results of the coloration test show once more consistency with the hypothesis.

In Table 26, we observe that the last value of the column for 4 speakers is a bit lower, 8.5, than the value above for 8 speakers, 9. Although this is not a substantial difference, it might come from the fact that in the mean of the preferences, the value of the comparison between 4 and 8 speakers, 6 for both cases, is less dominant in the case of 8 speakers because it is weighted with more comparisons than in the case of 4 speakers.

In the localization hypothesis, 5 subjects agreed with the hypothesis in 18 or more of the 28 comparisons. One subject clearly disagreed with the hypothesis and six answered showing no special agreement. Looking only to the comparisons with slow movements, there is a bit more of agreement with the hypothesis. However, in general the results of the localization test are not convincing enough.

In the coloration test, except one not trained subject, all agreed with the hypothesis in 20 or more of 28 comparisons. If we look only at trained subjects, except one, they all agreed in 24 or more of 28 comparisons and two agreed in 27 of 28 comparisons.

Chapter 5

CONCLUSION

5.1 Conclusion

Spatial audio offers to electroacoustic and electronic music a new field to explore. In addition to playing with instruments, microphones, synthesizers, recordings,... sound artists and musicians are challenged to design sound scenes in 2 or 3-dimensional spaces and to include the configuration of the sound reproduction system as a part of their composition. Thus, a tool to create spatial trajectories in the Soundscape Renderer has been developed in this work. Specific tools like the SSR are available and ready for musicians to use in live performance, making possible to play with movements and to create sound choreographies in real-time.

This research has been focused on amplitude panning for being the most widespread spatial audio technology. Amplitude panning was introduced at first for its use with a low number of loudspeakers. Would the extension of stereo or surround systems to systems with a higher number of loudspeakers suit with panning techniques as well? The most clear idea derived from the results of the listening tests is that if we search for the most realistic spatial sound experience, using a high number of loudspeakers with amplitude panning is not a good idea, because it degrades considerably sound coloration, which is very important for the overall quality of sound reproduction systems. This is specially important in the case we have moving sound sources. Unfortunately, it is not possible to determine a generic optimal solution, because changes in sound coloration not only depend on the distance between speakers but also on the speed of the movement of sound sources.

In general, we may add more speakers to amplitude panning reproduction systems, but in a moderate way. If we know which spatial sound work we need to reproduce and this involves some specific sound source movements, we may place speakers in line with these trajectories and use combinations of these speakers depending on the trajectory we want to reproduce. For instance, if we dispose of a circular speaker array but we want to reproduce a sound which moves doing squares, it is interesting to use only 4 speakers positioned in the corners of the square. The geometry of the sound reproduction will provide an easy association for the brain. In the case of a circle, we would use all the speakers in the array. Each situation requires to be studied individually.

In case we want to improve localization of virtual sound sources using a high number speakers in a setup, it is probably much better to use other technologies than amplitude panning, like Ambisonics or Wave Field Synthesis. As Wave Field Synthesis requires a very high number of loudspeakers, in situations where we can not afford such sound reproduction system, the best option to advance towards a more realistic spatial sound it is possibly Ambisonics.

5.2 Future Work

Following this research, it would be interesting to study other spatial audio technologies to see if sound coloration is degraded by increasing the number of loudspeakers, as in amplitude panning. The case of low order ambisonics is probably similar to amplitude panning. It would be interesting to test low order Ambisonics and higher order Ambisonics as well.

On the other hand, there is a lot of interesting stuff to do in the design of tools for spatial sound design, spatial sound installations and spatial music live performance. For the latter, it may be of special interest to focus on control interfaces for real-time performance and to try different approaches. The development of the control interface used in this work will be continued to cover more aspects of spatial sound design and sound choreographies. Hopefully, the SoundScape Renderer will keep being updated and will offer in the near future audio rendering options for three-dimensional reproduction.

Finally, something that would contribute extensively to spatial audio research would be to launch a crowdsourcing for collecting spatial audio mixes. This would help to discover diverse personal approaches to the creation of spatial sound scenes and would give multiple clues to find the needs of those who will use spatial sound currently and in the near future.

Bibliography

[1] Spors et al. “Spatial Sound With Loudspeakers and Its Perception: A Review of the Current State”, in Proceedings of the IEEE | Vol. 101, No. 9, 2013.

[2] Melchior et al. “Perceptual evaluation of a spatial audio algorithm on wave field synthesis using a reduced number of loudspeakers”, paper presented at the 131th AES Convention, 2011.

[3] Stephan Mauer and Frank Melchior. “A reference class loudspeaker panel for wave field synthesis”, paper presented at the 130th AES Convention, 2011.

[4] Jens Ahrens, Matthias Geier and Sascha Spors. “Introduction to the SoundScape Renderer (SSR)”, SSR manual, 2011.

[5] Andrés Pérez López. “Real-Time 3D Audio Spatialization Tools for Interactive Performance”, Master Thesis, 2014.

[6] Mark T Marshall, Joseph Malloch, and Marcelo M Wanderley. “Gesture Control of Sound Spatialization for Live Musical Performance”, in Gesture-Based Human-Computer Interaction and Simulation, number 1, pp 227–238. 2009.

[7] Marije A.J. Baalman. “Spatial Composition Techniques and Sound Spatialisation Technologies.”, in Organised Sound Vol. 15 (3), pp 209-218, 2010.

[8] http://www.williamepowell.com/pdfs/DefinitionsOfMusic.pdf

[9] Anthony Ashton. “Harmonograph. A visual guide to the mathematics of music.”, Wooden Books, 2003.

[10] http://mathworld.wolfram.com/Harmonograph.html

[11] http://mathworld.wolfram.com/LissajousCurve.html

[12] Knud Bank Christensen and Thomas Lund. “Room Simulation for Multichannel Film and Music”, paper presented at the 107th AES Convention, 1999.

[13] Ville Pulkki. “Virtual Sound Source Positioning Using Vector Base Amplitude Panning”, in Journal of the Audio Engineering Society, Vol. 45 (6), pp. 456-466, June 1997.

[14] Ville Pulkki. “Spatial sound generation and perception by amplitude panning techniques”, PhD Thesis, 2001.

[15] Matthias Frank. “Phantom Sources using Multiple Loudspeakers in the horizontal plane”, PhD Thesis, 2013.

[16] Helmut Wittek. “Perceptual differences between wavefield synthesis and stereophony”, PhD Thesis, 2007.

[17] G. Theile, G. Plenge. “Localisation of lateral phantom sources”, in Journal of the Audio Engineering Society, Vol. 25 (4), pp. 196-200, 1977.

[18] Günther Theile. “On the localization in the superimposed soundfield”, PhD Thesis, 1980.

[19] Alexander Raake, Hagen Wierstorf and Jens Blauert. “A case for TWO!EARS in audio quality assessment”, paper presented at the Forum Acusticum, Krakow, 2014.

[20] Jens Blauert. “Spatial Hearing”, MIT Press, 1983.

[21] M. G. Kendall, B Babington Smith. “On the method of paired comparisons”, in Biometrika Vol. 31, No. 3/4, pp. 324-345, 1940.