Sweet [re]production:

Developing sound spatialization tools for musical applications with emphasis on sweet spot and off-center perception

Nils Peters

Music Technology Area
Department of Music Research
Schulich School of Music
McGill University
Montreal, QC, Canada

October 2010

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

2010/10/26

Abstract

This dissertation investigates spatial sound production and reproduction technology as a mediator between music creator and listener. Listening experiments investigate the perception of spatialized music as a function of the listening position in surround-sound loudspeaker setups.

Over the last 50 years, many spatial sound rendering applications have been developed and proposed to artists. Unfortunately, the literature suggests that artists hardly exploit the possibilities offered by novel spatial sound technologies. Another typical drawback of many sound rendering techniques in the context of larger audiences is that most listeners perceive a degraded sound image: spatial sound reproduction is best at a particular listening position, also known as the *sweet spot*.

Structured in three parts, this dissertation systematically investigates both problems with the objective of making spatial audio technology more applicable for artistic purposes and proposing technical solutions for spatial sound reproductions for larger audiences.

The first part investigates the relationship between composers and spatial audio technology through a survey on the compositional use of spatialization, seeking to understand how composers use spatialization, what spatial aspects are essential and what functionalities spatial audio systems should strive to include.

The second part describes the development process of spatializaton tools for musical applications and presents a technical concept. The Virtual Microphone Control (ViMiC) system is an auditory virtual environment that recreates a recording situation through virtual sound sources, virtual room properties and virtual microphones. A technical concept is presented to facilitate artistic work with spatial audio systems and to allow the combination of different spatialization tools.

The third part investigates the perception of spatialized sounds as a function of the listening positions in multichannel sound systems. Perceptual experiments were designed to

understand the multidimensional nature of an off-center sound degradation and to propose concepts to improve the listening conditions for larger audiences.

This research extends our understanding of spatial audio perception and has potential value to all those interested in spatial audio quality, including designers, creators and specialists in the fields of acoustics, music, technology and auditory perception.

Abrégé

Cette thèse a pour objet l'étude des dispositifs de production et de reproduction des sons spatialisés. Plus particulièrement, ce travail comporte plusieurs expériences évaluant l'influence de la position d'un auditeur au sein d'un dispositif de haut-parleur de type "surround" sur la perception d'oeuvres musicales.

Au cours des 50 dernières années, de nombreuses applications destinées à la mise en espace de sources sonores ont vu le jour. Cependant, deux problèmes principaux persistent. D'une part, il semble que les artistes n'exploitent que très superficiellement le potentiel de ces nouveaux outils de création. D'autre part, dans la majeure partie des cas, lorsque les dispositifs de reproduction de sons spatialité sont utilisés pour un grand nombre de spectateurs, seulement une fraction du public peut vraiment percevoir les nuances et subtilités de la spatialisation; ces privilégiés se situent dans une zone d'écoute plus connue sous le nom de "sweet spot", caractéristique des systèmes de reproduction du son dans l'espace.

La recherche systématique de solutions à ces deux problèmes constituent le coeur de ce travail. Cette thèse est divisée en trois parties.

Dans la première, grâce à un sondage effectué auprès de compositeurs, nous avons essayé de mieux cerner comment la spatialisation s'intègre dans leur processus de composition, et aussi quels sont leurs exigences vis-a-vis des systèmes de spatialisation afin de faire l'inventaire des fonctionnalités essentielles d'un tel système.

Dans la seconde partie de cette thèse, le développement d'un nouvel outil pour la mise en espace de sources sonores est présenté: le système Virtual Microphone Control (ViMiC). Cette application est un environnement auditif virtuel (EAV) qui permet de recréer des conditions d'enregistrement particulières grâce à la modélisation de salles, de microphones et de sources sonores. Une approche technique permettant l'utilisation facile de plusieurs outils de spatialisation à des fins artistiques est aussi présentée.

Enfin, dans la troisième partie de cet ouvrage, les résultats de différentes études per-

ceptives sont détaillés pour identifier l'influence de la position des auditeurs lors de performances en son “surround”. Ces études ont été mises en place afin de caractériser de manière approfondie la dégradation de la perception sonore en fonction de la position d'un auditeur. Des solutions pratiques permettant de minimiser ces effets négatifs sont aussi proposés.

Ce travail de recherche permet d'étendre notre compréhension de la perception des sons spatialisés. D'autre part, cette étude est certainement précieuse pour ceux et celles travaillant à l'amélioration de la qualité des sons spatialisé, qu'ils soient techniciens, compositeurs, spécialistes en acoustique, en musique ou encore en perception auditive.

Acknowledgments

A number of people has been of great importance to this dissertation. First and foremost I would like to thank my advisors Stephen McAdams and Jonas Braasch for bringing me to McGill and for invaluable guidance over the course of this dissertation.

I fell lucky to had the opportunity in learning from Marcelo M. Wanderley, Philippe Depalle, Gary Scavone, Vincent Verfaille, Sean Ferguson, Fabrice Marandola, William M. Martens, Martha De Francisco, Catherine Guastavino, Evan Balaban and Yoshio Takane.

Of great importance were the members of the Music Perception and Cognition Lab, in particular Bruno L. Giordano, Georgios Marentakis, Bennett Smith and Finn Upham. I am thankful for all the discussions with students from Sound Recording and Music Technology, in particular Sungyoung Kim, Joe Malloch, Mark Marshall, Steve Sinclair, Doug van Nort, Avrum Hollinger, Carmine Casciato, Andrey da Silva, Mark Zadel, Sven-Amin Lembke and Bertrand Scherrer, who also helped me with the Abrégé. Special thanks to Tristan Matthews for supporting me on the ViMiC development.

I should not forget to thank Darryl Cameron and the CIRMMT crew Harold Kilianski, Yves Methot, Julien Boissinot, Rosemary Marland and Sara Gomez for technical and administrative support.

I also want to mention the close collaboration over the last years with Tim Place, Trond Lossius, Jan Schacher, Marlon Schumacher, Alexander Refsum Jensenius, Pascal Baltazar and Théo de la Hogue in the Jamoma project and the SpatDIF initiative.

I should not forget to acknowledge the Canada Council for the Arts (CAA), the Natural Science and Engineering Research Council of Canada (NSERC) and the Centre for Interdisciplinary Research in Music Media and Technology (CIRMMT) for financial support. I also thank Theresa Leonard from the Banff Centre for a research residency in 2007.

Finally, above all, thanks to my family and Sandra for all the support and encouragements over the years.

Contribution of Authors

The document is formatted as a manuscript dissertation and includes the following publications.

- Chapter 2: Peters N., Marentakis G., McAdams S.: Current technologies and compositional practices for spatialization: A qualitative and quantitative analysis. to appear in *Computer Music Journal (CMJ)*, Spring 2011.
- Chapter 3: Peters N., Lossius T., Schacher J., Baltazar P., Bascou C., Place T.: A Stratified Approach for Sound Spatialization. In *Proc. of the 6th Sound and Music Computing Conference (SMC)*, pages 219–224, Porto, Portugal, 2009.
- Chapter 4: Peters N., Matthews T., Braasch J., McAdams S.: Spatial Sound Rendering in Max/MSP with ViMiC. In *Proc. of the International Computer Music Conference (ICMC)*, pages 755–758, Belfast, Northern Ireland, 2008.
- Chapter 5: Peters N., Braasch J., McAdams S.: Learning from Users - Case Studies for Real-time Spatialization with ViMiC. Submitted.
- Chapter 6: Peters N., Braasch J., McAdams S.: Auditory perception at off-center listening positions in surround sound environments. Submitted to *Journal of the Acoustical Society of America*, 2009. currently under revision
- Chapter 7: Peters N., McAdams S.: Off-center Sound Degradation Based on Geometrical Properties in 5.1 loudspeaker setups. Manuscript prepared for submission.

I was responsible for designing and carrying out all experiments, including the data collection, conduction the data analysis, and preparing the manuscripts for all of the above listed publications. As my primary supervisor, Stephen McAdams provided necessary funding, laboratory space and equipment. Further, he generally contributed guidance in the conception and interpretation of the experimental data. Jonas Braasch, my secondary advisor, guided me primarily on technical aspects throughout the dissertation and provided the initial programming code of the Virtual Microphone Control (ViMiC) system described in Chapter 4. Under my supervision, research assistant Tristan Matthews and I improved ViMiC in various areas. We ported ViMiC from Pure Data to MaxMSP and CoreAudio, debugged and optimized the code, and added multiple features, i.e., a sound source directivity model, a rendering method which compresses the undesired Doppler effect, a more realistic virtual room model and Open Sound Control (OSC) support. I also wrote the user manual and maintain the software. Georgios Marentakis made suggestions to the design of the questionnaire in Chapter 2. For the publication of Chapter 3, I initiated and outlined the problematic and structure for the manuscript. Trond Lossius, Jan Schacher, Pacal Baltazar, Charles Bascou and Tim Place contributed ideas and provided portions to the manuscript of Chapter 3 edited by me.

Contents

List of Figures

List of Tables

List of Acronyms

CLP	Central Listening Position
DAW	Digital Audio Workstation
DBAP	Distance Based Amplitude Panning
DSP	Digital Signal Processing/Processor
EXC	Musical Excerpt
HCI	Human Computer Interaction
HOA	Higher Order Ambisonics
HRTF	Head-Related Transfer Function
ICLD	Inter-Channel Level Differences
ICTD	Inter-Channel Time Differences
MAA	Minimum Audible Angle
MAMA	Minimum Audible Moving Angle
OCP	Off-center Listening Position
OSC	Open Sound Control
SUG	Space Unit Generator
SSP	Sound Surface Panning
SpatDIF	Spatial Sound Description Interchange Format
SPL	Sound Pressure Level
VBAP	Vector Base Amplitude Panning
ViMiC	Virtual Microphone Control
WFS	Wave Field Synthesis

Chapter 1

Introduction

Spatial sound perception is an important process in how we experience sounds in our environment. This process has been attributed to survival and evolutionary purposes (Blesser and Salter 2006) and is generally studied in the fields of otology, audiology, psychology, neuroscience and acoustical engineering (Blauert 1997). The practical implications of spatial sound perception are notably found in communications (Blauert 2005), architectural acoustics and urban planning (Beranek 2003; Kang 2006), film (Chion 2009), media art (Bishop 2005) and music (Harley 1994).

For musical appreciation, it has been suggested that spatial sound perception plays a critical role. For instance, neurologist Oliver Sacks (2007, 143) reported on a patient whose musical appreciation had been altered suddenly since he lost all his hearing in one ear: "The perception of the specific qualities of music—pitch, timbre— did not change," he told Sacks. "However, my emotional reception of music was impaired. It was curiously flat and two-dimensional [...] and lifeless."

In composition, perceptual effects of spatial sound segregation, fusion and divided attention are explored artistically (see Harley 1998; Smalley 2007). Spatial aspects were first introduced as a musical element through static placement and separation of musicians in the concert space, for example in the antiphonal music of Gabrieli in the mid-16th century; later, progressive modifications of sound source position (e.g., by Charles Ives' father George) enabled more dynamic spatialization (Zvonar 1999).

Electro-acoustic inventions in the 19th and 20th century, such as microphones, amplifier and loudspeakers, and the recent increase in computer resources created new possibilities but also complexities and challenges for the artistic use of spatialization. For instance, due to the increasing variety and complexity of spatialization tools and the possibilities of spatializing an ever-growing number of sounds in real time, the development of appropriate control strategies and musically meaningful representations are some of those challenges. Furthermore, improvements in technology also provide opportunities for research on spatial sound perception, which must guide the development of better tools. The interest in this topic builds upon my previous studies in sound engineering at the University of Technology and at the University for Music and Dramatic Arts in Graz (Austria).

Aims of this Dissertation

The principle aim of this dissertation is to inform and contribute to the research and development of spatial audio technology through a systematic development of sound spatialization tools for musical applications and perceptual studies with regard to loudspeaker reproduction for a larger audience.

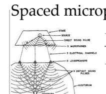
To that effect, a history of spatial audio technology, as it pertains to this principle aim is reviewed in Section 1.1 and general research challenges in the field of spatial audio reproduction (Section 1.2) are discussed. A theoretical model, intended to function as an exploratory paradigm for this research, is proposed in Section 1.3. Based on this paradigm, Section 1.4 outlines the following parts of this dissertation.

1.1 History of Spatial Audio Technology

Beginning with Blumlein (1931), the history of spatial audio technology is illustrated in a timeline at the bottom of each page in this introduction. The timeline contains important achievements in the development of spatial audio technology and is enriched with relevant artistic examples to illustrate the creative application of these inventions. Using primary sources, which are indicated, and relevant secondary sources (i.e., Davis 2003; Thompson 2004; Blesser and Salter 2006; Weinzierl and Tazelaar 2006) reviewed in the course of this research, this timeline provides the historical context for this dissertation. Remarkably, it suggests that the 70s and the time around the millennium have been the most eventful periods in the history of spatial audio technology. However this story probably begins earlier in 1881 with C. Ader's *Théâtrophone*, a telephonic distribution system that allowed music listening over the telephone lines using a headset with loudspeakers for each ear.

After the invention of condenser microphones in 1916 by E.C. Wente, the two major classes of multichannel microphone techniques were almost simultaneously defined by Blumlein (1931) (coincident microphone techniques) and Steinberg and Snow (1934) (spaced microphone techniques); both are still relevant today. Further, they can be seen as the foundation for modern rendering techniques such as Ambisonics (Blumlein) and Wave Field Synthesis (Steinberg and Snow). Steinberg and Snow's work had a direct impact on the development of the first multichannel audio experience in cinemas, know as Fantasound for Disney's movie *Fantasia* in 1940 (Klapholz 1991). Also, their conclusion that three

Spaced microphone techniques



J.C. Steinberg & W.B. Snow:
Auditory Perspective-Physical Factors

Alan D. Blumlein's Stereo Patent

Coincident microphone techniques

AEG Magnetophon

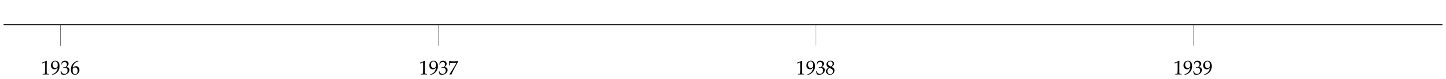
First tape recorder

1931 1932 1933 1934 1935

loudspeaker channels in the front provide adequate spatial accuracy is still embedded in today's surround loudspeaker setups (e.g., ITU 1992). Even more than 80 years after their inventions, coincident and spaced microphone techniques still divide the audio engineering community into two belief systems (e.g., Lipshitz 1986).

Spatial sound synthesis and sound diffusion became possible and popular in contemporary music (e.g., Maconie 2005; Varèse 1936) primarily due to the integration of microphones, tape recorder and loudspeakers into musical performance. One of the early devices that enabled spatial sound positioning was the *Rotation table* created by Karlheinz Stockhausen for his piece *Kontakte* (1958-60) and later extended for *Oktophonie* (1990/91). By using the *Rotation table* in a production studio, further explained in Section 5.3.5, sounds coming from a manually rotated loudspeaker were captured via four microphones and recorded on synchronized tape machines. In a concert, these tapes were reproduced on four loudspeakers surrounding the audience. Further, *Kontakte* was also probably the first composition made for a quadrophonic reproduction setup. This reproduction setup, most popular in the 70s, was used for live-controlled spatialization in a concert for the first time by Pink Floyd in 1967. Around the same time (1957) and already used in *Kontakte*, the company EMT presented the first artificial reverberator EMT-140, a 200-kg plate reverb that enabled users to create reverberation effects for the first time without using reverberation chambers or other physical spaces. Artificial reverberators and their underlying algorithms have evolved over the years (see timeline) and are one of the most important devices in music production.

Sound diffusion, a live performance strategy in which few channels of prepared audio are mapped to a potentially large number of speakers in a given space, has been used



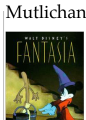
for decades. For instance, for Edgar Varèse's famous *Poème électronique* at the 1958 World Fair in Brussels, three channels of prepared electronic music were mapped from the tape to about 350 loudspeakers by sound projectionists with a series of rotary telephone dials. Later, large sound-diffusion systems using a variety of loudspeaker models (a.k.a. *loudspeaker orchestras*) were created beginning in the 70s (the *Acousmonium* by the *Groupe de recherches musicales* in Paris (Bayle 2007); *GMEBaphone*, now entitled Cybernéphone by the *Institut International de Musique Electroacoustique de Bourges* (Clozier 2001), and the BEAST by the Electroacoustic Music Studios at the University of Birmingham).

Another major concept in sound spatialization is to create trajectories of sounds, the origin of which can be found in the work of Chowning (1971) and in the Halaphone (Haller 1995). For the first time in history, Chowning created spatio-perceptual cues such as Doppler shift, distance attenuation, and air absorption via a computer synthesis algorithm in his piece *Turenas*. Another novelty in *Turenas* was the use of mathematical equations (Lissajous function) to create the sound trajectories around the listeners. Chowning's pioneering work on the simulation of moving sound sources can be found in many other spatial audio applications such as IRCAM's *Spatialisateur* (Jot 1997). The Halaphone, developed by Hans-Peter Haller and Peter Lawo in the early 1970s and refined throughout the years, became famous through its use by Luigi Nono (e.g., in *Prometeo*, 1984) and Pierre Boulez (e.g., in early versions of *Répons*[1]).

Recently, the increase in computer resources has progressively helped to conceptualize,

[1] According to Garavaglia (2009), Boulez revised the piece in the 1980s with the newest developments at IRCAM and replaced the Halaphone with the computerized mixing board MATRIX 32. The latest version of *Répons* from the early 1990s used then the software Spatialisateur.

Mutlichannel cinema sound "Fantasound"



W. Disney: Fantasia

3 channels (L/C/R) + maximum of 65 additional distributed loudspeakers

First stereo tape recordings

Helmut Krüger at the German Radio

1940 1941 1942 1943 1944

develop and refine spatial sound positioning techniques (e.g., Moore 1983; Pulkki 2001a; Daniel 2000; Rabenstein et al. 2004), as well as virtual room environments and auralization methods (e.g., Allen and Berkley 1979; Moorer 1979; Huopaniemi et al. 1996).

Technologies from the field of Human Computer Interaction (HCI) enable the expressive manipulation in real time of spatial sound synthesis, e.g., via graphical user interfaces (Pottier 1998; Delerue 2006), tangible tables (Bredies et al. 2008; Gasteiger 2009) or via gestural control (Pottier and Stalla 2000; Marshall et al. 2006; Schacher 2007). The reference to gestural control of spatialization, which dates back to the *Potentiomètre d'espace* in 1951, is highlighted because this dissertation was funded through a three-year grant[2] from NSERC/CCA entitled *Compositional Applications of Auditory Scene Synthesis in Concert Spaces via Gestural Control*.

A branch in spatial audio technology that directly informs this research is binaural reproduction. Binaural reproduction aims to recreate a listener's ear-signals through headphones or loudspeakers (i.e., transaural), and is elaborated in Section 6.2.5. In one of many use cases, for the purpose of preservation, recent advances in auralization and binaural reproduction technology allow a listener to experience the acoustics of destroyed venues based on architectural simulation models. Examples for such destroyed venues are the old Leipzig Gewandhaus (Weinzierl et al. 2010), used by W. Sabine in 1900 as a reference in designing the Boston Symphony Hall, or the Philips Pavilion, constructed for the 1958 World Fair in Brussels, which hosted Varèse's *Poème électronique* and Xenakis' *Concret PH* and was shortly thereafter dismantled (Lombardo et al. 2005; Zouhar et al. 2005). Binaural technology has been used for this dissertation in perceptual experiments described in

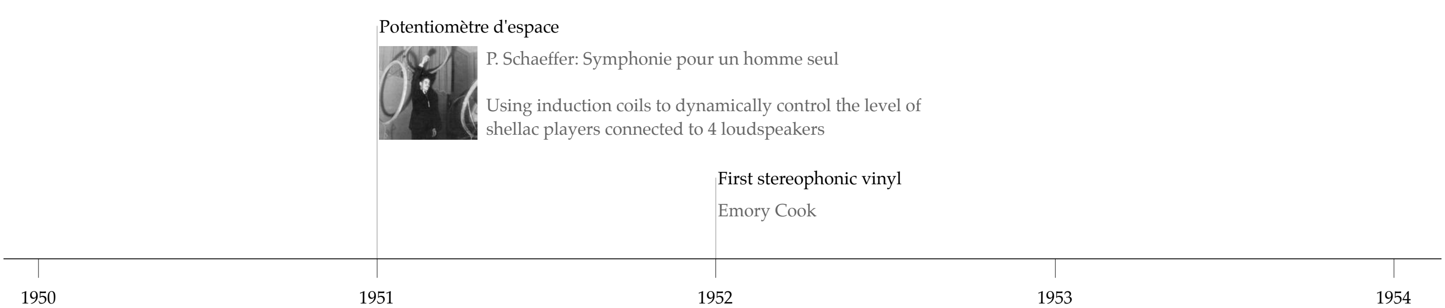
[2] http://www.cirmmt.mcgill.ca/research/projects/Gestural%20spatialization, accessed July 2010

1945 1946 1947 1948 1949

Section 6 and in Marentakis et al. (2008).

Today, many spatial sound rendering applications exist for musical applications, based on different rendering concepts. Because of the close relation between spatial sound recording and spatial sound synthesis, spatial rendering concepts can be associated with microphone techniques as illustrated in Figure 1.1. Four different categories are notable: 1) transaural and binaural reproduction, 2) techniques based on Inter-Channel Level Differences (ICLD), 3) techniques based on Inter-Channel Time Differences (ICTD), and 4) techniques that aim to recreate a soundfield. New concepts, such as Multipole-Matched Rendering (MMR, Hannemann 2009) or novel robust panning algorithms (Poletti 2007; Batke and Keiler 2010), are not illustrated in Figure 1.1, because these rendering techniques are currently under development and are not (yet) available for real-world applications.

In the last decades, due both to the spreading interest in the artistic use of spatialization and to improving audio technology, many venues have been created that are specialized for spatial audio reproduction, e.g., IRCAM's *Espace de projection* (Peutz 1978; Vinet 2007), IEM Cube (Zmöling et al. 2003), ZKM Klangdome (Ramakrishnan et al. 2006), UCSB AlloSphere (Amatriain et al. 2007), MUMUTH Graz. Also, conventional concert halls are increasingly being equipped with multichannel loudspeaker systems (see e.g., Casdorff et al. 2008), allowing composers and performers to experiment with and to employ spatialization techniques in their music. Consequently, such venues double as artistic and acoustic research facilities. Remarkably, whereas large scale systems often employed Ambisonics in the 90s and at the beginning of the millenium (e.g., Malham 1992), it seems that venues



now invest either in more resource-demanding Wave Field Synthesis (WFS) systems, or in a combination of both Ambisonics and WFS.

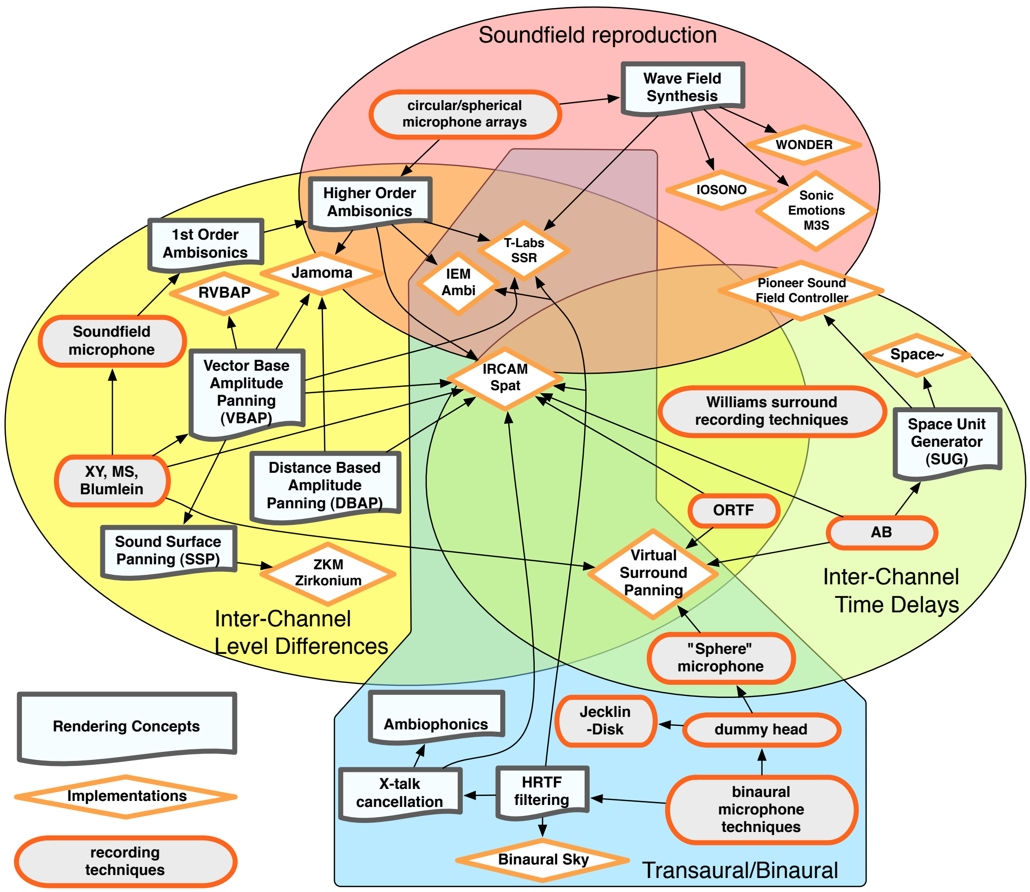


Figure 1.1: Overview and relationships of current spatial rendering concepts, available software applications and microphone recording techniques.

1.2 Research Challenges in Spatial Sound Reproduction

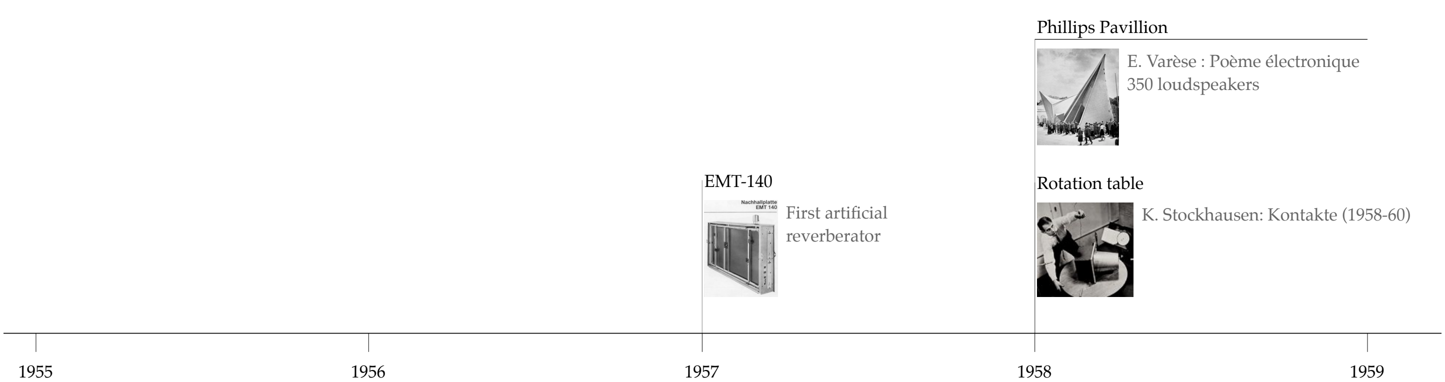
Critical to the success of the composer's work is the faithful delivery of the intended perceptual cues to the audience by way of spatialization. In the specialized and extended concert venues just mentioned, this success depends on several interrelated factors such as sound material, spatialization technique, loudspeaker setup, venue acoustics and the listening position. Especially for these performance venues[3], it is hard to predict *a priori* the quality of the spatial projection, because the understanding of the complex interaction between these factors is under-investigated and still little understood. Consequently, research in spatial sound perception in the context of contemporary music technology derives the primary challenge and inspiration from its interdisciplinary nature; the knowledge of many different research fields needs to be applied, specifically room acoustics, digital signal processing, human-computer interaction, psychoacoustics and music perception and cognition. More recently, neuroscience is also increasingly a contributing factor (e.g. Zatorre et al. 2002).

A fundamental barrier to research in this field is that spatial sound properties are often considered to be a secondary parameter in music research[4], and the vocabulary to describe spatial sound properties is still under development (e.g., Letowski 1989; Rumsey 2002; Kendall and Ardila 2008); it is often challenging for listeners to perceive and communicate these properties correctly (Neher 2004, Introduction chapter).

Another challenge is the use of meaningful sound material for perceptual research. Artificial, yet fully controlled stimuli, such as pure-tones or noise bursts, most often used in psychophysical experiments (e.g., Blauert 1997), do not reveal all perceptually relevant dimensions as embedded in music and therefore are not ecologically valid for the perceptual

[3]In the context of this research, performance venues refers to a large space where spatial audio is a featured or integral component of the artistic practice.

[4]For instance the exclusionary definition of timbre by the American National Standards Institute (1973), which enables the listener to "differentiate two sounds presented under the same conditions with the same pitch, loudness, and duration," seems to neglect human abilities to distinguish sound by spatial properties.



experiments of this dissertation. On the other hand, by using musical examples as stimuli, it is difficult to control all factors in the experimental design, requiring the researcher to trade off between ecological validity and controllability. Olive (2008, 200) writes that "the selection of programs for [the] evaluation of multichannel audio systems is particularly challenging because there are considerable variations among the recordings in their use of the front, center, and surround channels to create the auditory imagery."

1.3 Theoretical Model

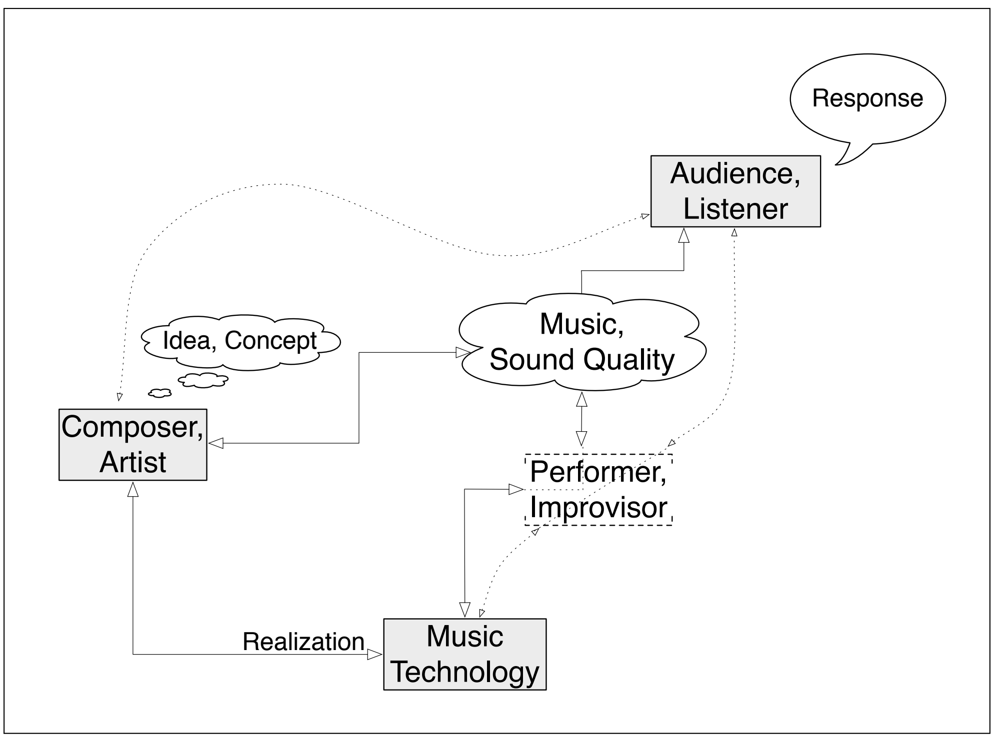


Figure 1.2: From musical ideas to listeners' responses - A exploratory paradigm illustrating the salient relationships and elements investigated in this dissertation.

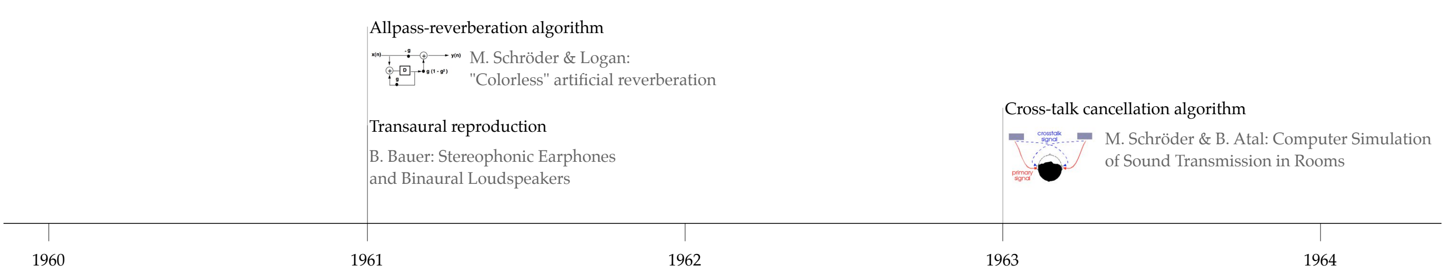
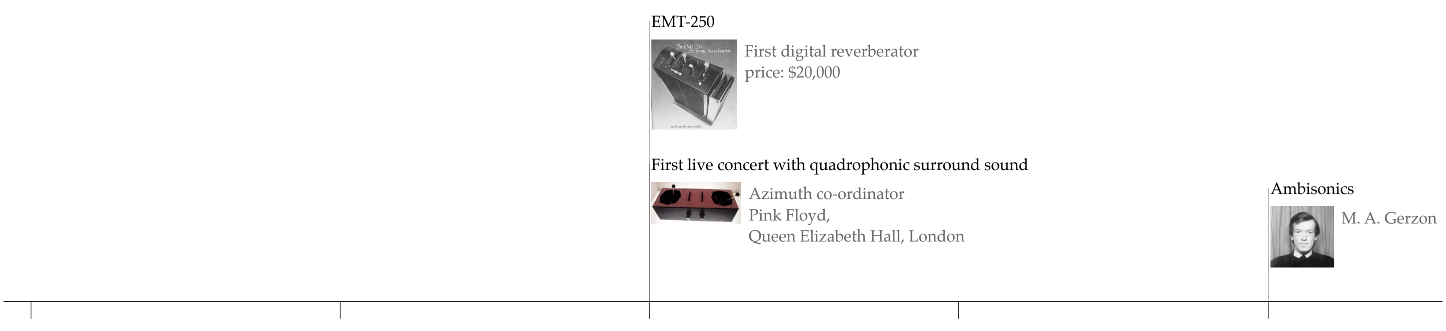


Figure 1.2 presents the author's hypothetical relationship among composer, music technology and the listener. In this paradigm, a composer/artist has an initial musical idea or concept in mind that (s)he wants to transform into music with a certain aesthetic and sound quality through the use of appropriate tools. For thousands of years[5], music technology has been extending the expressivity and acoustical range (e.g., loudness, spectrum) in order to enable the realization of musical ideas. Of course, the selection and usage of those tools (e.g., musical instruments, music software) differs greatly across artists – the reason why music technology has even been considered as an extension of an artist's personality. In this dissertation, we focus on tools for the synthesis and reproduction of spatial sounds.

In the model, through music technology, the initial musical concept is transformed into a composition that can then be presented to an audience. When listening to music, the audience will have reactions, such as satisfaction, arousal, or dislike, potentially driven through experiences, expectations, preferences, mood, and biases toward, for example, music technology, performer, sound quality, or musical structures.

All components in this model are clearly relational, indicated in Figure 1.2 by solid-cornered lines for formal relationships and dotted-curved lines for more loose connections. Scenarios including a live performer or improvisor will add more complexity to the model as illustrated. For instance, a performer could be directly influenced by the audience response, or the experience of using musical tools can alter the composer's musical ideas. Considering

[5]Archaeologists have found bone flutes dating back to 30000 to 37000 years, commonly accepted as the oldest known musical instruments (CBC News 2004).

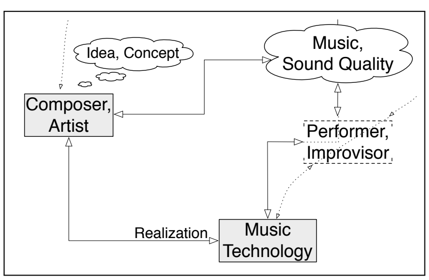


how music technology changes music consumption patterns and how listening habits can be changed by technology (e.g., the Sony Walkman enabled mobile and private listening, Hosokawa 2008), the link from listener to music technology can be strong and important. Because of these dependencies, this model could also be described as an ecosystem.

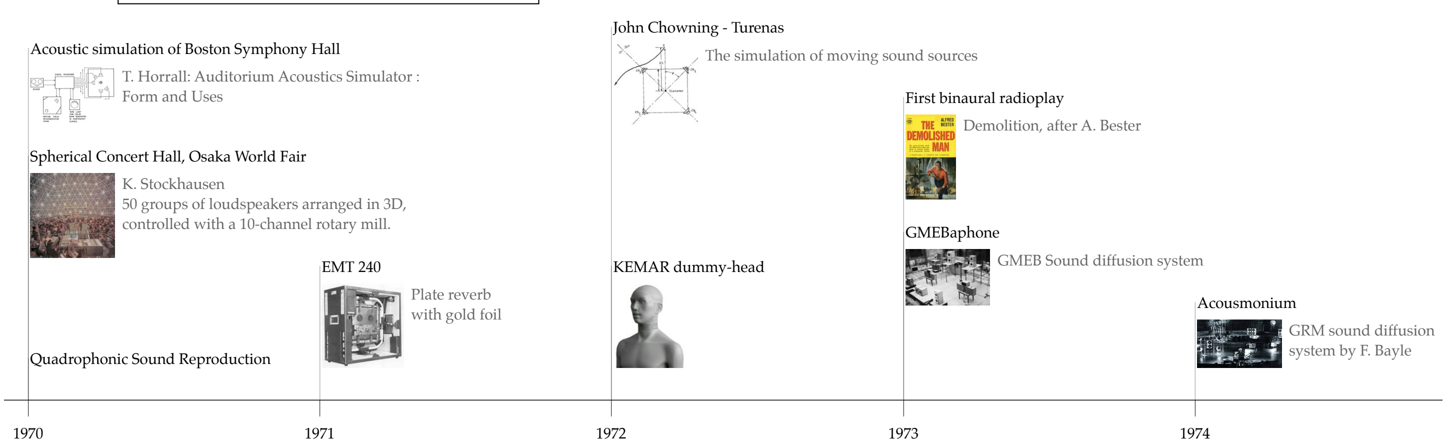
1.4 Dissertation Structure

In three parts, the dissertation investigates the different aspects and relationships illustrated in Figure 1.2. Using qualitative and quantitative methods for the first part, relationships between composers and spatial audio technology are elaborated. The second part describes the development process of spatialization tools for musical applications. The third part relies primarily on quantitative methods to investigate the perception of spatially reproduced sounds as a function of the listening positions.

Part 1 - Exploring the Relationship between Composers and Spatial Audio Technology



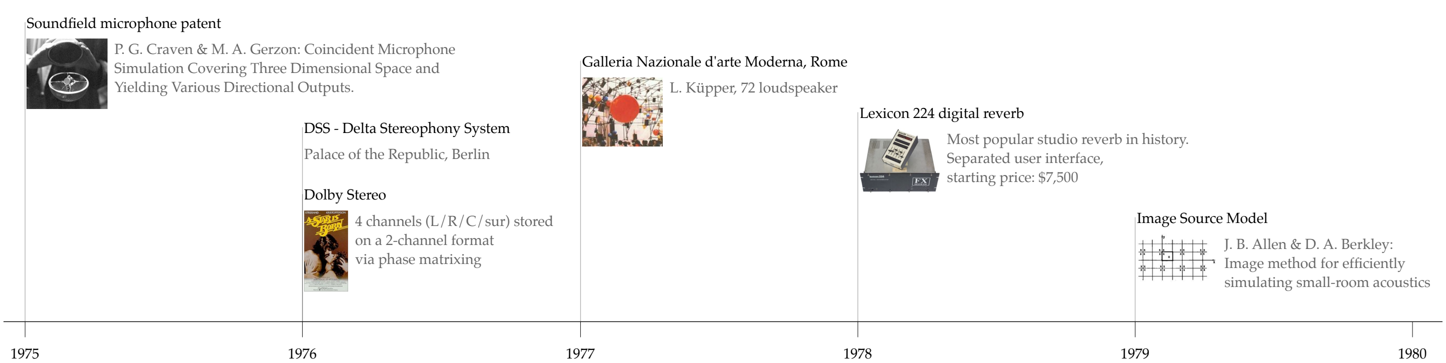
As outlined in Section 1.1, a great deal of progress in creating accessible spatial audio technology has been made. However, according to Blesser and Salter (2006, 204), "Either from lack of scientific interest or from insufficient ed-



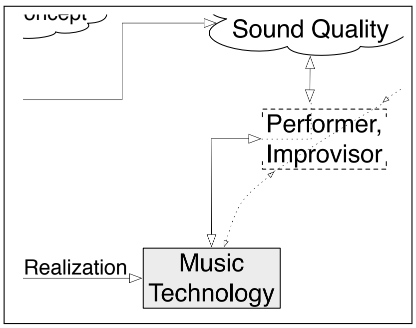
ucation, most artists do not understand the properties of audio reproduction systems used for creating virtual systems. If available systems outside of the laboratory were closer to ideal, there would be no need for artists to understand their properties. However, existing systems are far from ideal." Similarly, but from the perspective of a composer, Natasha Barret finds that "spatialization equipment and technology have become readily available, but the users haven't caught up" (Otondo 2007, 17). Understanding this incoherence between engineers' advancements in technology and its creative use is needed to guide future research efforts toward fruitful directions.

The goal is therefore to understand how composers use spatialization, what spatial aspects are essential and what functionalities spatial audio systems should strive to include or improve. What is the motivation to use spatialization and how is it used? What quality and effects are sought by artists?

A web-based questionnaire was developed (see Appendix A) and announced on different appropriate web-sites and newsletters to reach a diversity of sound artists. This questionnaire is unique in its focus, and its analysis and discussion in **Chapter 2** is essential in making the research process more relevant to composers, musicians and listeners.



Part 2 - Developing Spatialization Tools fur Musical Applications

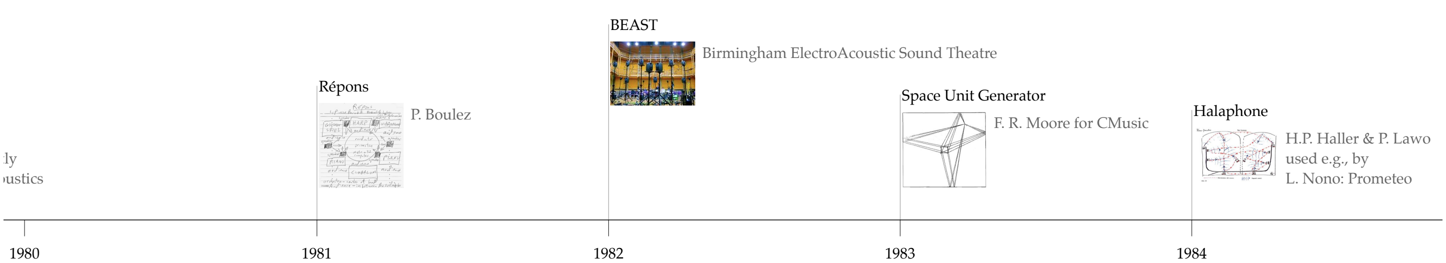


Currently, several spatialization concepts and implementations are available (Figure 1.1). For a detailed overview of spatialization tools, the SpatBASE [6], a freely accessible and collaborative WIKI database for spatialization software was initiated by the author. The goal of this database is to progressively review spatialization tools and to provide a collective documentation for users and developers, a missing resource for these communities. Because the SpatBASE also includes artwork that was facilitated with specific spatialization tools, it helps to promote art and technology and may even be a resource for future musicologists.

An initial review of existing sound spatialization tools revealed a lack of a coherent syntax and data formats across applications. This lack of a standardized format for controlling spatialization across different rendering platforms complicates the portability of compositions and requires manual synchronization and conversion of control data–a time-consuming affair. To facilitate collaboration between researcher, composer and institutions, the Spatial Sound Description Interchange Format (SpatDIF) was proposed primarily by the author (Peters et al. 2007; Peters 2008) and is now under development in the spatial audio community[7] (see also Kendall et al. 2008).

[6] http://redmine.spatdif.org/wiki/spatdif/SpatBASE, accessed June 2010

[7] At the time of writing, a SpatDIF meeting is being held at IRCAM, Paris; http://imtr.ircam.fr/imtr/GDIF_SpatDIF_Meeting, accessed May 2010



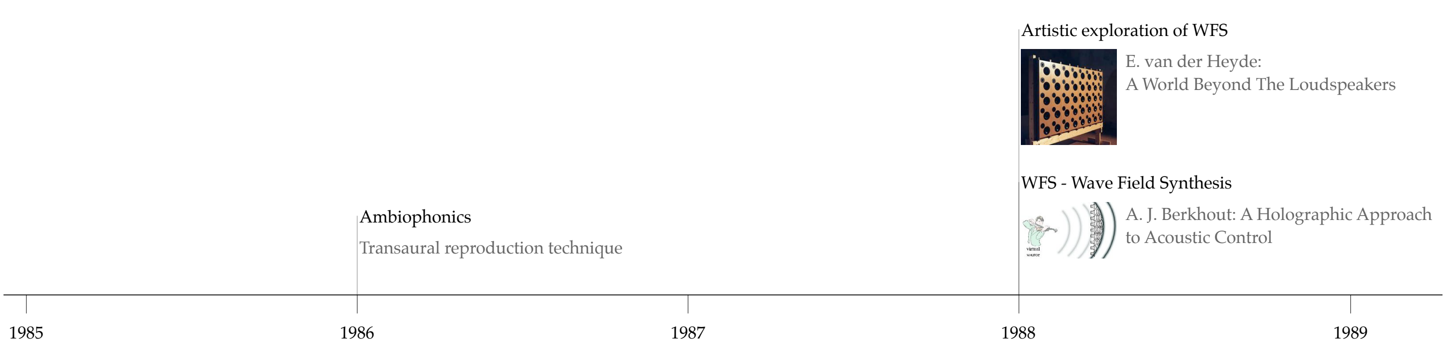
In Chapter 3, a novel approach to structuring technical aspects of spatialization systems is presented and aims to facilitate artistic work with such systems that are often found by users to be complicated and cumbersome (see for instance the previously cited quote by Blesser and Salter 2006, 194). This approach was developed through discussions with several artists and developers[8] and allows the combination of different spatialization tools in order to find new means of artistic expression.

Chapter 4 reports on the Virtual Microphone Control system (ViMiC), conceptualized by Braasch (2005), and greatly improved and refined in terms of stability, rendering quality, usability and integrability by the author and research assistant Tristan Matthews. From an initial proof-of-concept state in the form of a Pure Data patch prior to this dissertation, ViMiC is now freely available as a Max/MSP external, and as an AudioUnit plug-in for Digital Audio Workstations (DAW).

Chapter 5 shows how the the findings from Part I of this dissertation were applied in the development process of ViMiC. This chapter shows through use cases how ViMiC is now extensively employed in multiple fields. Further, new applications are outlined.

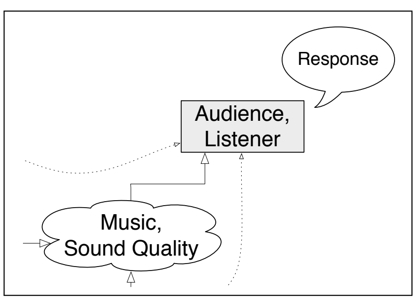
In addition to the ViMiC development, the Jamoma open-source project (Place and Lossius 2006, `http://jamoma.org`) was greatly extended and refined to facilitate the artistic and scientific work with spatial audio in Max/MSP, a prominent media programming environment (Zicarelli 1998). In co-existence with ViMiC, a number of different spatial rendering concepts are available in Jamoma. All are implemented with a common OSC

[8] See the list of SpatDIF meetings `http://redmine.spatdif.org/projects/spatdif/wiki/Meetings`, accessed July 2010



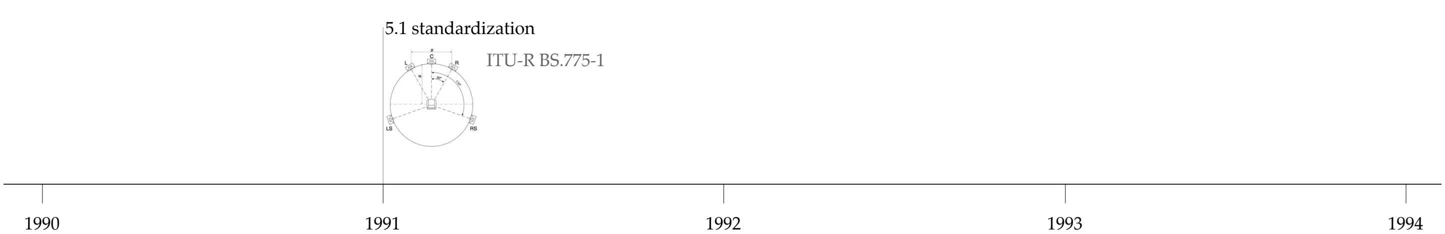
namespace, patching concept, and Graphical User Interface (GUI) so that one can easily exchange and explored the strengths and limitations of each rendering concept. Jamoma modules that were newly developed during the course of this dissertation are listed in Appendix C. Further, Jamoma was used in this dissertation as a platform to develop concepts to facilitate expressive work with control parameters (Place et al. 2008a;b) and to create more flexible audio processing environments (Place et al. 2010a;b), all of them being relevant for spatial audio applications. Resulting musical applications are TrakHue (Peters et al. 2007), DJ Spat (Marentakis et al. 2007b), and a system for gestural control of spatialization (Marshall et al. 2006).

Part 3 - Listener Perceptions of Sweet Spot and Off-center Listening Positions



As outlined in Section 1.1 and apparent in the timeline, for a number of decades, many performance venues for larger audiences have been equipped with a large number of loudspeakers, often differing in terms of room size, their technical specifications and the applied audio rendering concept. In spatial audio reproduction, a best listening position is usually implied (the sweet spot, or the *money seat*, Toole 2008, 301), thus making high-quality surround-sound reproduction difficult for audiences of more than one. This is different from conventional concert hall acoustics, which are designed to enhance natural sound sources on stage and produce a plurality of listening positions with perceptually good sound images of those sound sources.

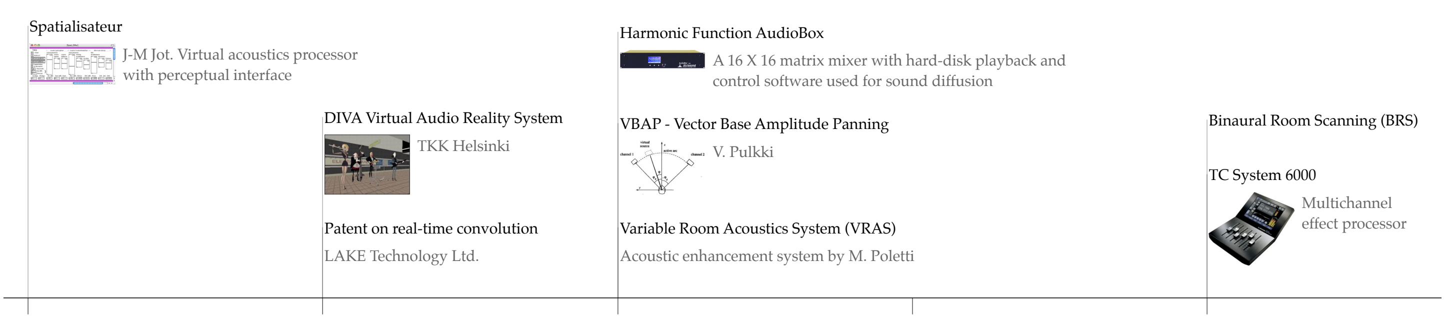
Spatial audio listening experiments usually focus on the central listening point (the



sweet spot) as mostly relevant for domestic listening scenarios. However, in the context of larger audiences, off-center listening perception becomes important because most of the audience is located at off-center listening positions. At these positions, the loudspeaker feeds coming from different directions and arriving with different time delays and level differences, create unique ear signals that differ for each off-center position. According to Toole (2008, 328), “Cinemas and home theatres are sometimes filled with as many seats as they can physically hold. We know that all of those seats are not equally good, but we try not to talk about it.” In concordance with Toole and to the author’s best knowledge, off-center listening perception has not been studied in enough detail. Perceptual studies on sound reproduction, including off-center listening positions for larger audiences, are increasingly becoming the subject of intense interest, both artistically and commercially. Recent relevant studies have been published by Frank et al. (2008), Bates et al. (2007) and Marentakis et al. (2008).

During the course of this dissertation, several listening experiments were carried out to specifically investigate off-center listening perception. We focused here on 5.0 loudspeaker setups (ITU 1992; 1997), because they are the most widely used and commercially accepted spatial audio reproduction format.

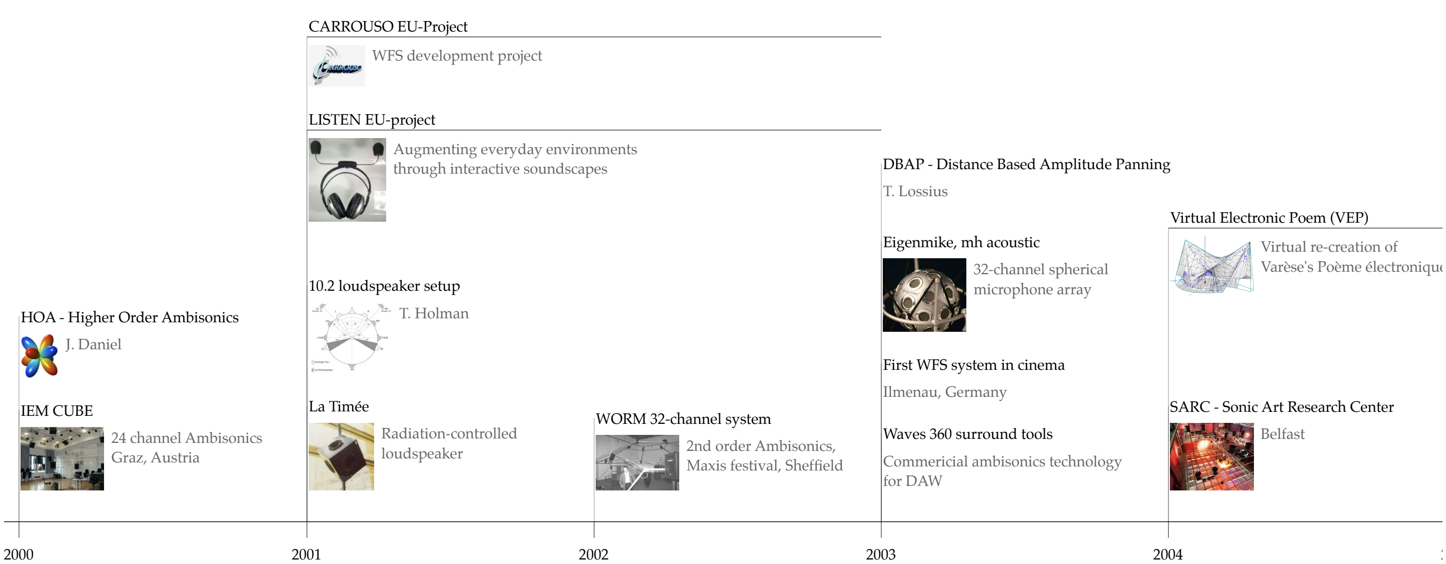
Chapter 6 investigates with two listening experiments the effect of multichannel microphone configurations on the perceived sound quality at off-center (non-sweet spot) listening positions in medium-size rooms and for larger audiences. Statistical models, which use stimulus sound features to predict the audience responses, were created with reason-



able accuracy to understand the sound features that lead to a degradation of perceived off-center sound. This investigation is also related to the ViMiC developments presented in Part II because it can be used to provide recommendations for an optimized arrangement of virtual microphones within the ViMiC system to accommodate larger audiences. Also in this chapter, a detailed examination of the geometric considerations and a review of potential perceptual artifacts is provided.

In **Chapter 7**, a novel experimental framework, developed to investigate off-center sound degradation, is used to study factors in the relation between listener and the loudspeaker setup geometry. Specifically, the three identified factors caused by geometrical relations are time-of-arrival differences, sound-pressure-level differences between the signal feeds, and the direction of the arriving wavefronts.

First, in a qualitative study, perceptual descriptors indicating this sound degradation are assessed. This step was necessary because, as mentioned in Section 1.2, perceptual descriptors that can be used to characterize off-center sound degradation have not been formally studied yet, but are important for further quantitative studies. Then, in a series of quantitative experiments, the individual contribution of the three geometrical factors to off-center sound degradation was studied in two room conditions using the most frequently

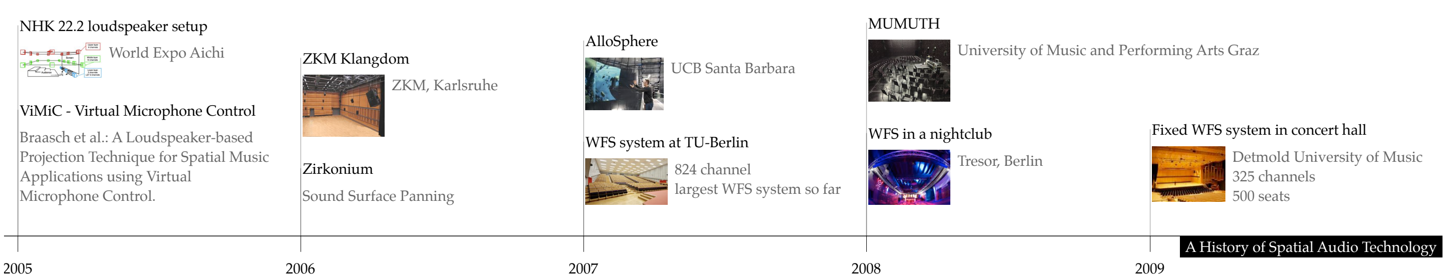


used perceptual descriptors as found and classified in the previous experiment. The relation between derived perceptual descriptors and overall listener preference is also examined in this chapter.

Although they are not included in any detail in this dissertation[9], other complementary perceptual experiments on the effect of the listening position were carried out in collaboration with Georgios Marentakis and Stephen McAdams, focusing primarily on sound localization of spatialized sounds. The effect of listening position and spatialization algorithm on the Minimum Audible Angle (MAA, Mills 1958) and Minimum Audible Moving Angle (MAMA, Perrott and Musicant 1977) of spatialized noise bursts were studied. MAA and MAMA have only been estimated for real sound sources and not for sources that are virtually spatialized over a loudspeaker array. For all experimental conditions, the MAAs for virtually spatialized sound sources were higher than those found for real sound sources. There were also significant effects between spatialization algorithms, the number of loudspeakers, and the dimensions of the 2D loudspeaker arrays (see for details Marentakis et al. 2007a; 2008a;b).

In another study (Marentakis et al. 2008), the influence of the performer's gesture on the perception of spatial sound trajectories was studied in a concert hall environment at different listening positions. For real-time sound spatialization, visible on-stage gestural control can improve the perception of sound trajectories, especially if the performer is equipped with a binaural reproduction system to control the spatialized sounds.

[9] The dissertation is solely based on first-authored publications.



1.5 Originality, Importance and Anticipated implications

Research in spatial audio quality has traditionally focused on a single reference listening point (the sweet spot) in order to improve the spatial audio technology used in the reproduction process. With the aim to improve spatial sound reproduction, especially for larger audiences, this dissertation investigates the use of spatialization in a more comprehensive way, starting from the artist's expectations and requirements for spatial audio technology, continuing to the development of spatialization tools for musical applications, and ending with the listeners' perception of spatially reproduced music as a function of different listening positions (see Figure 1.2). This approach is chosen because the author believes that an audience's appreciation of spatial sound reproduction relies on the production stage in the composers' studio, the employed spatialization technology, as well as on the reproduction context.

Exploratory research in the form of a study on the compositional use of spatialization was required due to a lack of information related to the use and understanding of spatialization concepts and technology by those who use it. A questionnaire was specifically developed for this purpose (see Appendix A). Novel in its depth and complexity, the results provide a rich insight into many compositional and technical aspects and are a valuable resource for other researchers. This study was integral to the development of the tools in Part II and helped focus the design of the listening experiments in Part III.

The listening experiments were designed to investigate sound perception in surround-sound loudspeaker setups outside of the ideal sweet spot based on preference and similarity ratings. To overcome logistical difficulties of listening experiments that investigate sound quality as a function of a listening position, a novel methodology was developed which allows simulation of a variety of listening positions in a laboratory. This methodology also allows for the decoupling of salient physical factors that characterize a listening position and can be used for other listening experiments in spatial audio.

This research may be one of the first in-depth and formal investigations of spatial audio quality perception measured at off-center listening positions. It is used to guide the development and refinement of a number of spatialization tools (see Part II and Appendix C). Because of the extensive work on the ViMiC spatialization concept for this dissertation, it is now usable in real-world conditions and is already employed in different artistic and scientific contexts (see Chapter 5).

This dissertation will provide new information to those who are connected to spatial audio research, production and reproduction, such as software developers, sound engineers, artists, (psycho)acousticians, loudspeaker manufacturers, and musicologists. This has already become apparent in the course of this research, through author-initiated projects already mentioned above, such as SpatBASE and SpatDIF.

Part I

Exploring the Compositional Use of Spatialization

Chapter 2

Current Technologies and Compositional Practices for Spatialization: A Qualitative and Quantitative Analysis

The following Chapter has been accepted for publication in:

Peters, N., Marentakis, G., and McAdams, S. (2011) Current technologies and compositional practices for spatialization: A qualitative and quantitative analysis. to appear in *Computer Music Journal*, 35(1)

The complete survey that was developed for this publication can be found in Appendix A.

2.1 Introduction

Spatialization, the synthesis of spaces and spatial properties of sounds for a listener, is a growing field of interest for researchers, sound engineers, composers and audiophiles. Due to broad and diverse viewpoints and requirements, the understanding and application of spatial sound is developing in many ways. To benefit from varying viewpoints, artistic practice and theoretical or applied research interests need to engage in a regular dialogue. Blesser and Salter (2006, 184) reported on the long-term relationship between artists and audio researchers regarding virtual spaces, "the story of an evolving relationship between sophisticated audio engineers, creating tools, and impatient artists, incorporating such tools long before they are fully defined". Otondo (2008) showed that over the last 10 years the technical equipment of composers has improved both in quality and quantity, with sound spatialization based on five or more loudspeaker channels being increasingly preferred over traditional two-channel stereo systems. However, novel spatialization tools available to composers have hardly found their way out of the research labs. Composers continue to use conventional and familiar spatialization techniques. As composer Natasha Barret said, "the spatialization equipment and technology have become readily available, but the users haven't caught up" (Otondo 2007, 17). We need to understand this lack of coherence between development and creative musical application in order to effectively conduct future research efforts.

2.2 Methodology

In our study, a web-based questionnaire was designed and presented to composers and sonic artists in order to elicit understanding of how they use spatialization, what spatial aspects are essential and what functionalities spatial audio systems should strive to include or improve. Additionally, we surveyed the degree to which recent development in spatial audio technologies is known and has already been applied by artists.

The survey consisted of two major parts with compositional (13) and technical (11) multiple-choice style and comment-form (open-ended) questions in English (see Appendix A for the full survey design). As opposed to the multiple-choice questions, answering open-ended questions was not obligatory. Each multiple-choice question included a comment text field to account for individual responses. The arrangement of the given choices was randomized across respondents to reduce order effects. To ease the response-entry process for the respondents, this survey was deployed over internet and could be stopped and continued at any time. Using the constant comparison method of analysis (Glaser and Strauss 1967), open-ended questions were independently analyzed by two researchers to control for biases in interpretation. Coding examples for the discussed open-ended questions are provided in Appendix B.

2.2.1 Respondents

This survey was announced in March 2008 on several appropriate web domains, such as SpACE-Net and mailing lists by the Canadian Electroacoustic Community (CEC), the British Sonic Arts Network (SAN), the Australasian Computer Music Association (ACMA) and Norwegian young composers. Also, several invitations were directed to specific contemporary composers, including the panelists of the 2008 CIRMMT/UCSD Music + Technology Incubator III workshop. Responses were collected for 14 days. Within this relatively short timeframe, 52 surveys were completed (approx. 55% of all the surveys that were started). This response rate can be considered high for a non-reward, web-based survey and suggests demand and interest.

2.3 Responses

Figure 2.1: Geographic distribution of the 52 respondents.

Respondents were primarily male (85%) and predominantly from Europe and North America (Figure 2.1). For musical education, more than 80 different universities/conservatories were named, of which the most frequent were Université de Montréal (17%), University of Birmingham (10%), and Stanford University (8%); several respondents were self-taught (11%). Respondents reported an overall composition experience of 20 years on average, 14 years of which was computer-aided and 10 years of which involved spatialization. Remarkably, several experienced composers reported a longer history of using spatialization than applying computer techniques to their work. Because we expected that work experience might affect responses, composers were separated into analytic groups according to their reported experience in using spatialization techniques: "beginners" (under 5 years), "intermediate" (5 - 10 years) and "advanced" (more than 10 years), resulting in equal-sized groups.

2.3.1 Fields and Forms of Application

In order to create valid use cases in research and development, we were interested in the artistic fields and presentation forms in which spatialization is applied. The upper part of Figure 2.2 shows that more than half of the respondents use spatialization for live electronics and/or for prepared electronics (fixed media). The "Acousmatique" classification was added here because several composers explicitly indicated this category in the comment text field. While prepared electronics seem to be equally distributed among experience groups, one can see that in live electronics, spatialization is used less by "beginners". One could speculate that a live electronics project might generally be a bigger challenge for an artist than a fixed media production, such that "beginners" are likely to reduce a project's complexity by avoiding spatialization. Similarly, we find, on average, fewer "beginners" in the mixed-music category, which combines electronics with instrumentalists. An Analysis of Variance (ANOVA) revealed that the number of orchestrations per respondent increases significantly among experience groups ($F(2, 46) = 5.6, p < 0.007$).

Almost every respondent (more than 90%) presents spatial music in a concert situation (lower section of Figure 2.2). The second most frequent presentation form is sound installations (more than 60%). New media forms, represented through the categories "Web application" and "Film, video", are the least common forms for respondents with more than 10 years of experience.

2.3.2 Compositional Motivation and Realization

In an open-ended form, we asked composers why they use spatial aspects in music. The comments were sorted and clustered into response categories (Table 2.1).

Most frequently (58%), composers intentionally use spatialization to enhance the listening experience. Multiple responses suggest that such an augmented experience is often

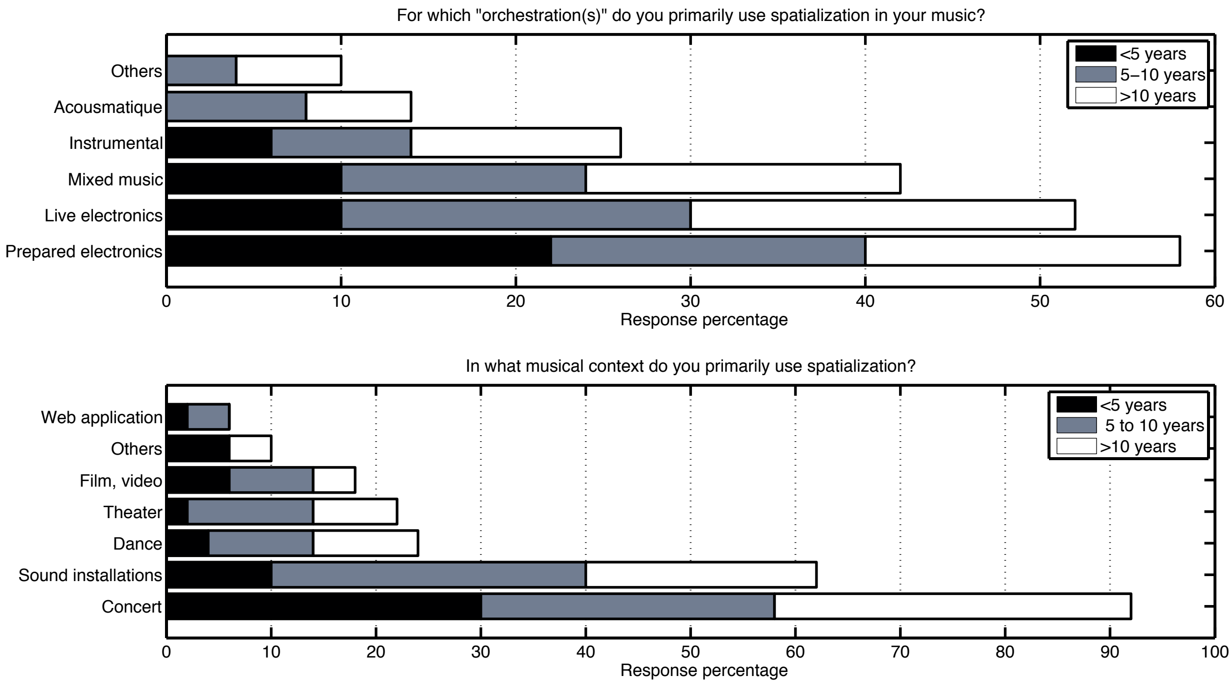


Figure 2.2: Distribution of "orchestrations" and contexts in which spatialization is used, as a function on compositional experience. Multiple categories could be selected. (In this and similar figures, the rightmost edge of an entire bar measures the total response percentage across all three experience groups. Within a bar, the response percentage for a given experience group is given by the width of the corresponding colored section of the bar, not by the absolute position of that section's rightmost edge.)

achieved through immersing the listener in sound. Spatial aspects heighten the experience of space and time and therefore "intensify the sensory experience for the listener", according to a British composition student. Several artists believe that listeners find it more interesting to hear sounds coming from a variety of directions than from only the traditional frontal stage direction. One response reminded us that the perception of music is always linked to space and time, but spatialization is a powerful method to heighten this experience for the listener. For almost half (44%) of the respondents, spatialization is a compositional paradigm. "There is no other way to express the ideas I am working with", said a 39-year-old artist who works on spatial sound installations, internet applications and concert music. Another composer, working on prepared electronics for sound installations and film/video said that spatialization is “a subtle but important part of the whole in [his] compositions”. More precisely, a composer who works with Wave Field Synthesis (WFS) said that he is not interested in the accuracy of movements and localization of WFS, but in the way an individual sound can create a space itself without changing acoustical properties of the room, e.g., by using additional reverb. Forty percent mentioned that spatial aspects help to organize and structure music. “The spatial structure of a work may be of equal importance as its organization in terms of pitch, timbre or rhythm”, stated a composer who primarily presents fixed media pieces within a concert situation. Another person mentioned that spatialization “adds one or more artistic layers to a piece.”

Table 2.1: Why composers use spatial aspects in their music. Fifty respondents in open-ended form. Respondents chose multiple categories, resulting in 116 responses.

Category	Total responses	%
To enhance the listening experience	29	58
Paradigm for artistic expression	22	44
To organize and structure sounds	20	40
To experiment with technology and spatial effects	12	24
Perceptual motivation	12	24
Segregation and blending of sounds	11	22
To add motion and dynamic	9	18
To make sounds more natural and vivid	7	14

A quarter of the responses (24%) indicate that many composers are attracted to the experimental and exploratory side of spatial effects and spatial sound technology. “Because I can” or “I like to play with sounds” were typical responses. Someone else reported that in spatialization “there is still lots of room for innovation, which I like.” A composer working in the field of live electronics said that spatial parameters are available *a priori* in the world of computer music and therefore have to be at least considered. Further, experienced composers are attracted by the novel possibilities spatialization offers. A composer from the "advanced" group mentioned that "spatialization gives the composer the means to expand their gestural palette into the spatial domain in a dynamic way not previously possible." Nearly a quarter (24%) of the responses mentioned that spatial aspects are perceptual attributes of hearing sounds and music. "Life is spatial, music is spatial," therefore those attributes should be addressed in music. Composers also explicitly mentioned the use of sound segregation and sound blending as a motivation for spatialization (22%), concepts which are known from auditory scene analysis research (Bregman 1990; Harley 1998). A composer working for more than 5 years with spatialization said he can "present more sound material at the same time without losing [...] clarity." Others feel that "complex music becomes more comprehensible." Other responses mentioned that spatialization is applied to add motion and dynamism (18%) or to make electroacoustic sounds more real and vivid (14%), "to give sounds an identity."

Very often, composers addressed several of these response categories (Table 2.1). For a more comprehensive understanding, we studied these intercategorical relations and found a strong connection between the three categories "To enhance the listening experience", "To organize and structure sounds" and "As a Paradigm for artistic expression". Many responses that addressed "To experiment with technology and spatial effects" also relate to these three response categories and indicate how strongly the experimental nature permeates current spatial music. Furthermore, "Segregation and blending of sounds" is strongly connected to "To organize and structure sounds" and to "To enhance the listening experience". Sound segregation is known to facilitate a listener's processing of compositional structure[1].

Further to the previous question, we wanted to know how composers configure sound ele-

[1]For instance, Meyer (2009, 264) discusses the importance of the spatial organization of musicians within an orchestra to support compositional ideas such as dialogue-like interchanges between similar instrument groups.

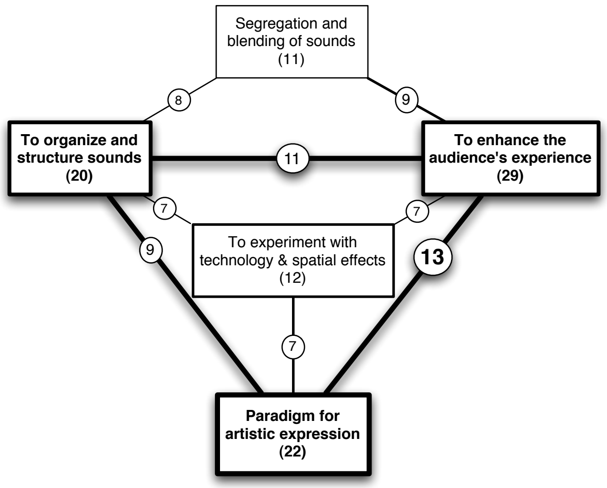


Figure 2.3: Intercategorical relations. Numbers in circles show how often two connected response categories were mentioned within one response. The plot presents response categories and connections that have more than seven shared responses.

ments in order to realize spatialized sound experience. If developers knew more about such methods, development could be better applied and could increase the usability of spatialization tools, for instance. This question triggered a variety of unique responses that are related to musical context and site-specific aspects, and are therefore hardly generalizable. The responses (see Table 2.2) mainly addressed methods of moving sound sources and their distribution in space and could involve experimentation with the sound material and the acoustics of the listening room. Often these methods are used to achieve a contrasting perception of sounds (i.e., clarity vs. blurriness, close vs. far, reality vs. surreality, thick sounds vs. thin sounds, dense vs. open).

Table 2.2: How do composers configure sound elements in order to achieve spatialized sound experience? Forty-seven responses in open-ended form.

Category	Example
Distribution and Position	Distance and depth
	Algorithmically generated distributions
	Spatial organization according to timbre, texture, musical function
	Spatial granularization
	Stereo tracks as source material
	Diffuse sounds (multi-speaker distribution)
	Mono sound reproduced from one or two adjacent speakers
	Image size
	Planes, subspaces and hierarchical sound layers
	Front (stage) is the focal point
Movements	Slight movements
	Contrast static vs. dynamic
	Many movements with a small number of sounds
	Strength of movements according to musical function: melody moves more than other sounds
	Prepared trajectories
	Moving instrumentalists
	Live performer moving speakers
Others	Changing loudness and dynamics
	Spectral filtering
	Reverberation
	Simulating room acoustics
	Surreal spatial impressions

2.3.3 Working Environments

In the Composition Studio

Figure 2.4 shows how often different reproduction systems are used in a composition studio. The responses were given with respect to a list of reproduction setups according to the frequency categories: Never, Seldom/Sometimes, Usually, and Always. As the main reproduction system in a composition studio, 35% of the composers Always use a two-channel stereo loudspeaker setup and 20% Always use headphones. Multi-loudspeaker arrangements such as quadraphonic or 5.1 are generally used only Sometimes or Seldom. More than 50% of the composers Never use or do not have access to seven- or eight-loudspeaker configurations. Because stereo and headphone reproduction is so common in the compo-

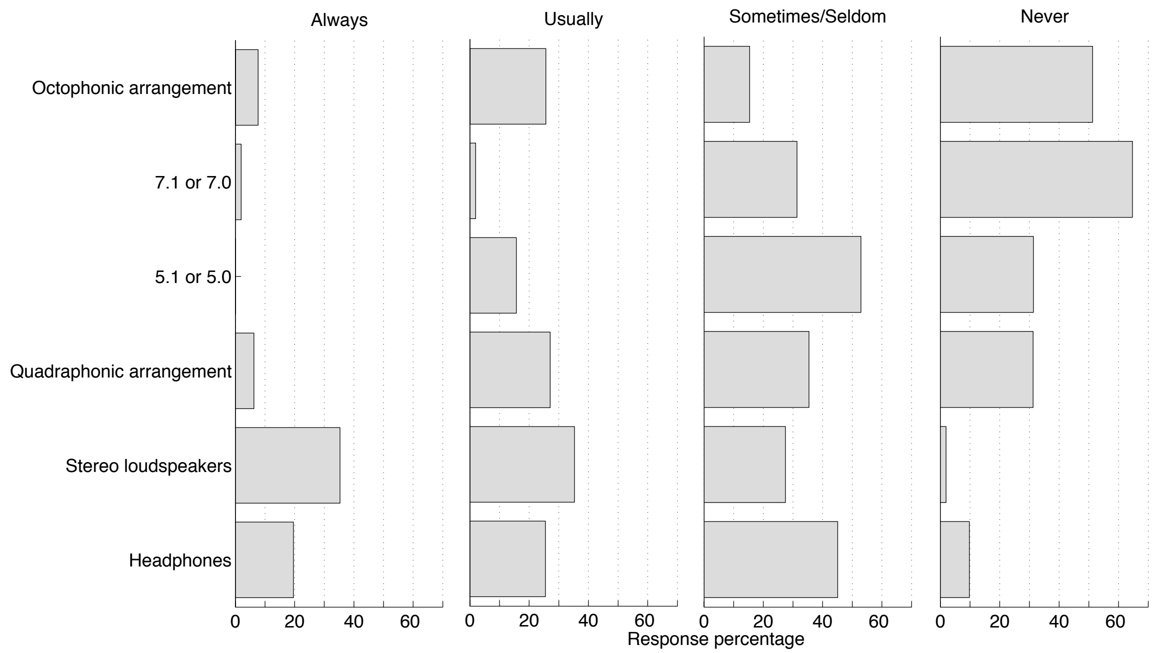


Figure 2.4: For composing spatialized music in your studio, how often do you use the listed reproduction systems?

sitional phase, one might speculate whether, instead of using expensive multi-loudspeaker setups, composers use binaural or transaural processes to spatialize audio around their head with only two audio channels while composing. Responses to a question concerning whether binaural or transaural versions of their spatial music have been released suggest that these techniques are widely uncommon (see Table 2.3). This response is not surprising. Fontana et al. (2007) reported that binaural versions of popular music were not preferred over standard stereo mixes by listeners, although the binaural versions were recognized to have better spaciousness and out-of-head localization.

Table 2.3: Do you release your music as binaural/transaural preprocessed versions?

Category	%
No	80
I don't know what that means	16
Yes, binaural	4
Yes, transaural	0
Yes, binaural and transaural	0

At the Performance Venue

Most composers (76%) consider the loudspeaker arrangement in the performance situation as different from their studio environment. Very often, when performing in a venue, there are more loudspeakers available than in the composition studios. An Acousmatique composer said "I have eight loudspeakers in my studio, but most of my work is intended for sixteen loudspeakers, and more recently, for twenty-four speakers. The eight loudspeakers in my studio give me an idea." Another person said that he works in a variety of composition studios with usually two to five loudspeakers, whereas the performance venue can contain more than fifty loudspeakers. In contrast to reproduction systems found in composition studios, today an eight-loudspeaker array can be considered as the most common loudspeaker configuration in venues and electroacoustic music festivals (Lyon 2008). However, loudspeaker setups also differ in terms of sound quality and in the distance and height between a listener and the speakers. Several composers mentioned that their studio is acoustically treated to minimize the effect of room reflections, in contrast to the performance venue, where reverberation can be expected. In terms of performance venues, traditional concert halls, specialized venues for electroacoustic music, as well as gallery spaces, are most common (Figure 2.5). It is surprising that cinemas are generally unused, although they provide standardized (e.g., the THX standard) acoustic treatment and multichannel loudspeaker systems. On average, composers work in approximately four different kinds of venues, this number increases with more experience (right side Figure 2.5).

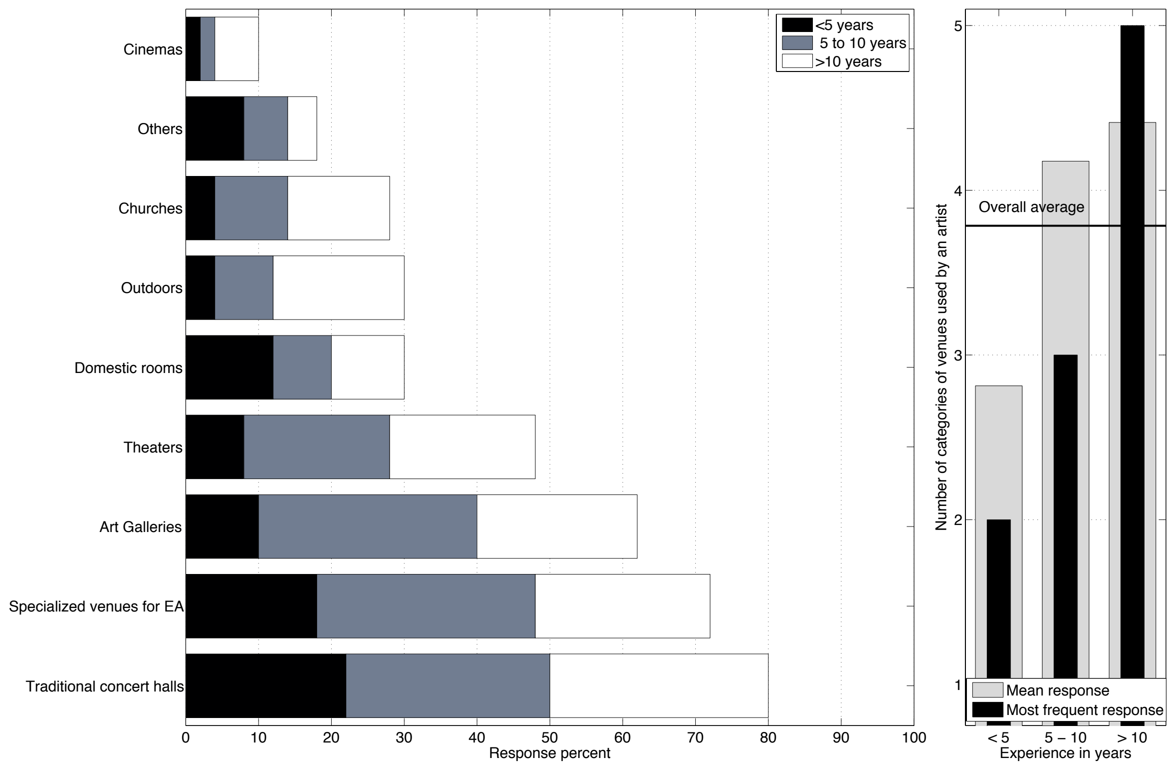


Figure 2.5: In what kind of venues are your spatial compositions performed? Others: bars and clubs, classrooms, spatial audio labs, virtual reality caves, spaces with remarkable acoustics.

Artists were also asked about the main challenges they have faced in venues regarding their compositional aspirations (Table 2.4). Main challenges were found to be related to the acoustic conditions, to technical limitations of the venue, and to time constraints[2] within the venue. In terms of the acoustical conditions, challenging attributes of venues are room dimensions, raked seats (i.e., an inclined floor), modes and resonances, and too much reverberation, which dominates the perception of spatial elements in a composition and causes microphone feedback for live electronics. The "sonic leakage" from one exhibition to another is an additional problem for sound installations. Regarding the technical limitations, the main complaint mentioned was the limited number and quality of loud-

[2]In a personal conversation, Scott Wilson from the BEAST in Birmingham said that before a concert, an artist gets normally three times the length of the piece as rehearsal time.

speakers. Almost every third response reported that the time allocated for arranging and optimizing the loudspeaker configuration at the venue is too short. Further, non-ideal locations of loudspeakers with respect to the audience and a small listening area (sweet spot) were reported. Many venues have room dimensions that complicate the setup of equidistant loudspeakers as required by most spatial rendering algorithms[3]. Composers who work with elevation and height face the difficulty that loudspeakers are often configured in a horizontal-only arrangement; hanging loudspeakers is almost always impossible in traditional concert spaces and opera theatres. It was reported that venue managers without the agreement of the composer repositioned the seats of the audience or placed extra furniture resulting in seats being too close to loudspeakers and walls where the sound quality for listeners is degraded. Sometimes an ideal placement of loudspeakers may not be possible due to aesthetic constraints of the stage or lighting designer.

Table 2.4: What are the main challenges of venues you have faced so far with respect to your compositional aspirations? Forty-two respondents.

Category	Total responses	%
Acoustical conditions	20	48
Technical limitations of the venue	16	38
Time constraints	13	31
Non-ideal loudspeaker and audience location	13	31
Staff and audience	8	19
Sweet spot	7	17
Cost of production	2	5
No problems	3	7

Because of the diverse reproduction conditions across venues, artists have developed (compositional) strategies to adapt their work. “I tend to accept the effect of venue as part of the concretization of my ideas,” said an artist who performs in traditional concert halls, specialized venues for electroacoustic music and art galleries. Other composers reduce their

[3]To a certain degree, non-equidistant loudspeaker set-ups can be calibrated by electronic means.

technical requirements from the beginning, thereby limiting the spatial possibilities. "As I have moved more towards visual arts, I have discovered that even getting adequate stereo playback in a venue is problematic. I certainly don't try for anything beyond 5.1," explains an Australian artist. Some responses suggest that composers tend to work with more extreme and obvious spatial properties, such as heavy panning rather than using more subtle spatialization techniques, to ensure that at least these gestures will be perceived. Composers of fixed-media pieces create different versions to account for different loudspeaker arrangements. If there are more loudspeakers than tracks, some tracks may be assigned to more than one loudspeaker. Therefore, "the more the number of tracks, the less the adaptability" according to one artist. "In the studio, I usually use a stereo system. At the performance place, I then adapt my work to the diffusion system," added a Canadian composer. An Austrian composer reported simulating "real-world condition[s]" by using outdoor loudspeakers in his studio when he works on outdoor installations.

2.3.4 Preservation of Spatial Music

The questions of how to preserve electroacoustic music is becoming an increasingly important topic which is especially challenging when spatialization is involved. This questionnaire probes this area in terms of media formats and notation practices.

Media Formats

One multiple-choice question asked what media formats are used for publishing spatial music. The two-channel audio CD is the most common medium (80%), while all other formats are used by less than 40% of the respondents (Figure 2.6). It should be mentioned that in a sound installation project, for instance, a two-channel recording is often made mainly to serve as documentation. One composer will use the "good old Audio CD" until there is a proper storage standard for multichannel files. Spatial music is also often stored on data DVDs that either contain PCM audio files for each individual loudspeaker feed (fixed media) or the composition within the audio software project files. One of the 10% who do not use any media, an Acousmatique composer, stated, “Publication is not important for me, I mostly work for live performance”. We could also find differences among the experience groups. Notably, the composers of the beginner group are more experienced with new media formats, such as mp3 and mp3-surround. In particular, mp3-surround, a technology to encode spatial audio in a mp3 file, is exclusively used by these composers. Remarkably, DVDs, the most common medium for multichannel audio work, are being used as often as conventional mp3. “I am interested in surround-sound technologies that are becoming available on the internet”, said a composer who uses audio CD and mp3 in performances and installations. Another respondent anticipates that fixed media will fade away in a few years in favor of sound files on generic media that can be adapted for different listening scenarios.

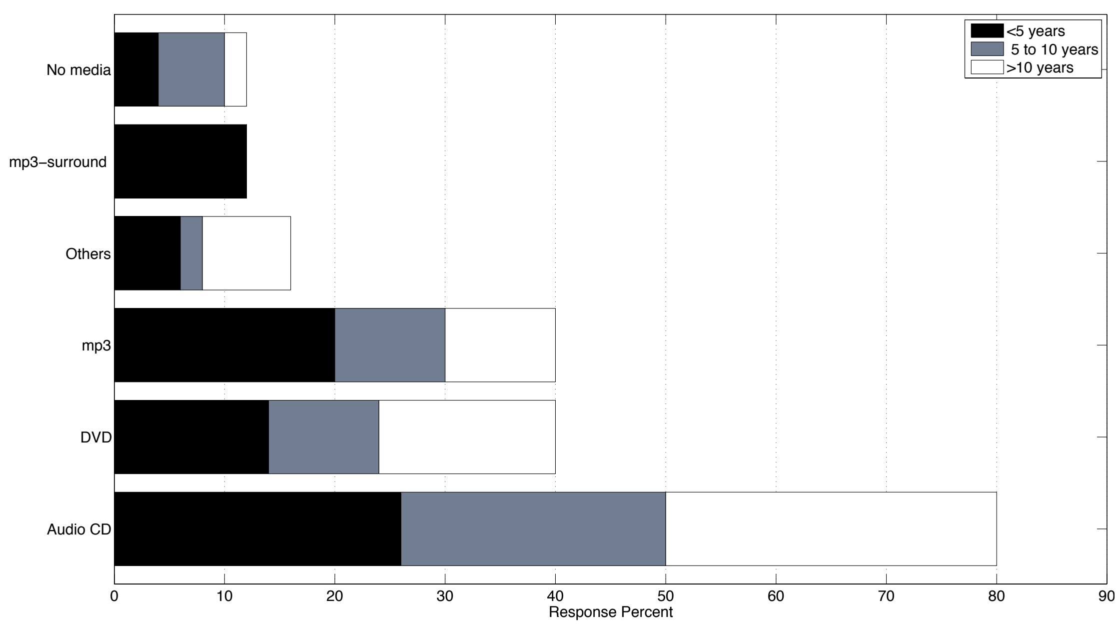


Figure 2.6: In what consumer media format(s) is your music published?

Notation

Through scores, compositions can be stored, exchanged, studied and performed, but also revised and adapted after their initial creation. To the authors' knowledge, there is no common language or notation format for spatial parameters as described by Rumsey (2002) or Kendall and Ardila (2008) for instance. Therefore, it seems reasonable that 62% of the respondents do not use any sort of notation for spatial aspects. The composers' arguments are that for (fixed) tape music, there is no need for a score, or they haven't found a satisfying way to notate spatialization. Others said that there is no score because spatialization is created through improvisation and experimentation, or generated by algorithms in real time. Composers seem to have developed individual spatial notations, ranging from photos and drawings over diffusion guides, poems and descriptive text to sonograms of the music with annotations. Some forms of spatial notation also depend on the production environment. When working in media programming environments such as Max/MSP or SuperCollider, spatial parameters are stored in data arrays within the composition patch, in contrast to DAWs (Digital Audio Workstations), where built-in track automation is used to store and recall parameter changes. "When working with the WFS system, the drawings become one of the main parts" was a response that suggests the rendering concept and technical environment have an impact on notation schemes.

2.3.5 Artists and Their Spatialization Tools

We are interested in what composers think about the software and hardware tools they use for spatialization and how their feedback can impact future development. We also are interested in the degree to which composers are aware of recent developments in spatial audio technologies. Figure 2.7 visualizes the responses to the question "What software and hardware tools have you used for spatial compositions?". The responses were given with respect to a list of spatialization tools according to the categories: Never heard, Heard about but never used, No longer in use, Currently in use, and Planning to try it. The list of spatialization tools was developed through a review of spatialization applications by the authors and is a mix of concepts and products. The experimental approach to spatialization (Table 2.1) is supported in the choice of tools: 20% of the people use self-made or custom-made tools and approximately a third (31%) use a media programming environment, such as Max/MSP, or SuperCollider. The primary spatialization tools are the built-in panning devices of DAWs and audio sequencers (75%) and panning performed with a hardware mixing console (58%). It seems that older technologies, such as the pan-pots in mixing consoles, are well known, but many are no longer used (in case of pan-pots, 37% of the respondents no longer use hardware mixers). Similarly, IRCAM's *Spatialisateur*,

Table 2.5: Concepts, hardware and software products included in the survey.

Name	Concepts/ Hardware/ Software	Reference
Vector Base Amplitude Panning (VBAP)	C	Pulkki (2001b)
Distance Based Amplitude Panning (DBAP)	C	Lossius et al. (2009)
1st-order Ambisonics	C	Gerzon (1973)
Higher Order Ambisonics (HOA)	C	Daniel (2000)
Wave Field Synthesis (WFS)	C	Rabenstein et al. (2004)
Space Unit Generator (SUG)	C	Moore (1983)
Virtual Microphone Control (ViMiC)	C	Braasch et al. (2008)
ZKM Zirkonium	S	Ramakrishnan (2009)
IRCAM Spatialisateur (Spat∼)	S	Jot (1992)
GMEM Holophon tools	S	Pottier (1998)
Waves 360° Surround tools	S	Waves (2009)
Vortex Surround tools	S	Vortex (2009)
Panning within Audio Sequencer	S	-
Pan-pots in mixing Console	H	-
Virtual Surround Panning in Studer-digital mixer (VSP)	H	Horbach et al. (2000)
TC-Electronics S6000	H	TC Electronics (2009)
TiMax Audio Imagine System	H	TiMax (2009)
Fraunhofer IOSONO	H	IOSONO (2009)
Sonic Emotion M3S	H	Sonic Emotion (2009)

under development since 1991 (Jot 1992; 1999), and VBAP (Pulkki 2001b) from 1998 are widely known, but are also often replaced by other tools. It is quite likely that several built-in panners in audio sequencers use VBAP or Ambisonics algorithms internally, but for some reason offer less explicit background information about the rendering concept.[4] For instance, Waves 360° is a known spatial processing toolbox for DAWs that employs Ambisonics concepts.

[4]See the amount of information on Apple's Core Audio Panners: http://developer.apple.com/releasenotes/MacOSX/WhatsNewInOSX/Articles/MacOSX10_5.html

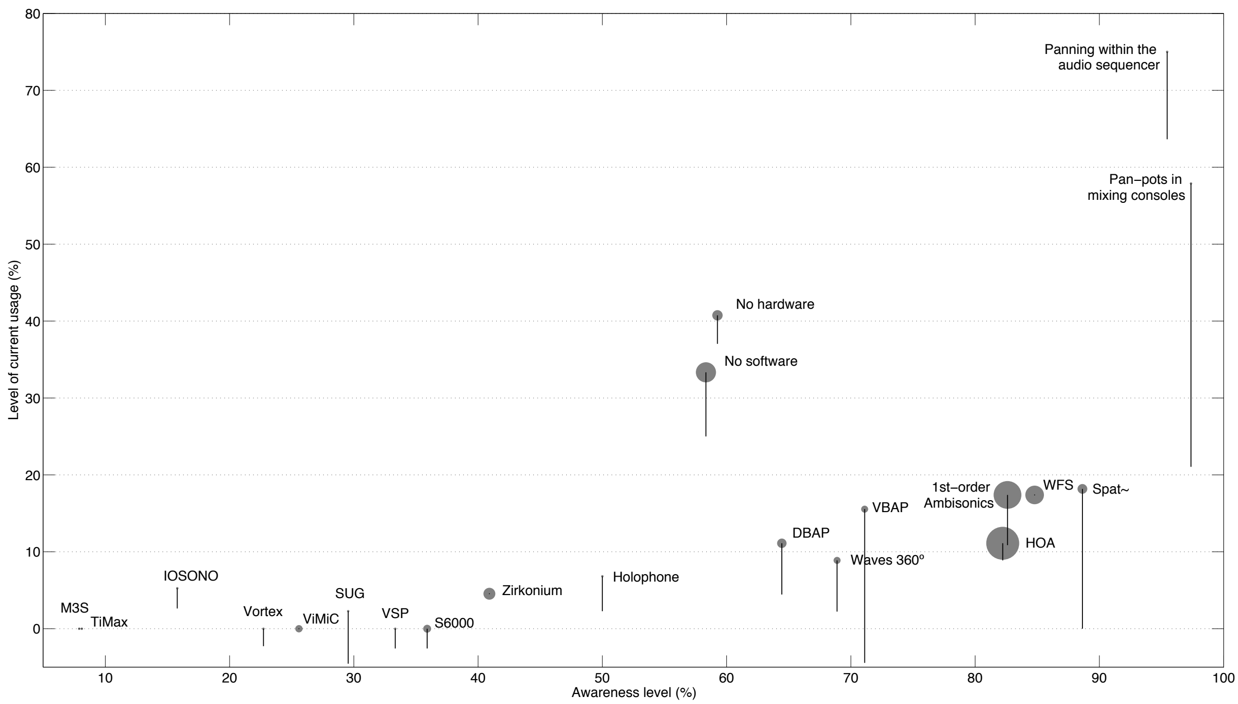


Figure 2.7: What software and hardware tools have you used for spatial compositions? The longer the vertical line under the bubble, the less the composer continues to use this tool. The bigger the bubble, the more the composer plans to try it.
M3S - Sonic Emotion M3S WFS system, TiMax - TiMax Audio Imagine System, IOSONO - IOSONO WFS system, Vortex - Vortex Surround tools, ViMiC - Virtual Microphone Control, SUG - Space Unit Generator, VSP - Virtual Surround Panning in Studer-digital mixer, S6000 - TC-Electronics S6000, Zirkonium - ZKM Zirkonium, Holophon - GMEM Holophon tools, DBAP - Distance Based Amplitude Panning, Waves 360° - Waves 360° Surround tools, VBAP - Vector Base Amplitude Panning, HOA - Higher Order Ambisonics, WFS - Wave Field Synthesis, Spat∼ - IRCAM Spatialisateur

The categories "No hardware" (41%) and "No software" (33%) also account for composers who work with instrumentalists and without any electronics. It is surprising to see that several respondents are planning to work without software. Does that mean that composers are frustrated with current spatialization tools? The categories "Pan-Pots in mixing console" (58%) and "No hardware used" (41%) sum up to 99% of the responses and show little awareness of other hardware-based spatialization tools.

According to Figure 2.7, the rendering concepts of 1st-order and Higher Order Ambisonics (HOA) seem to be the most interesting techniques for future compositions. However, people currently seem more aware of Wave Field Synthesis (WFS, Rabenstein et al. 2004) than Ambisonics[5]. Although HOA and WFS are both high-resolution spatial sound reproduction techniques that are conceptually based on the physical reconstitution of a wave field (Spors and Ahrens 2008), one wonders what makes HOA more attractive. Might it be Ambisonics' ability to store spatial audio material independently of the reproduction scenario? Or, is it that by defining the "Ambisonics order" and choosing the number of loudspeakers, the size of the listening area can be scaled to different listening scenarios? For WFS, the trade-off between the numbers of loudspeakers and spatial quality creates artifacts throughout the listening area. By using a periphonic loudspeaker dome, Ambisonics can reproduce elevated sounds, which is an important feature for many composers (Figure 2.9(a)). Also, WFS systems are still rare; approximately 20 research labs, and about 20 auditoriums are equipped with a permanent WFS system (De Vries 2008). Ambisonics systems may simply be more easily accessible to artists. There are several Ambisonics tools up to 11th order that are freely available, e.g., Schacher (2010). Although these Ambisonics tools do not work in such higher Ambisonics orders to compete with WFS's sound reproduction abilities, composers can get used to the Ambisonics concept by using it in

[5] Frank (2009) reported that Tonmeister students were much more aware of WFS than HOA.

production studios and for concert performances (Färber and Kocher 2010) and can more easily switch to higher-order systems in the future. However, the small number of loudspeakers that are currently used in composers' studios (Figure 2.4) suggests that they need to be equipped with more loudspeakers so that spatial rendering concepts requiring many loudspeakers (e.g., WFS and HOA) can be applied. Despite the higher interest in HOA, WFS offers many compositional possibilities (cf. Baalman 2007). In contrast to HOA, if an artist only wants to present sounds from one side (e.g., from the front), loudspeakers do not have to surround the audience–an economic, aesthetic[6], and pragmatic aspect. A composer working mainly on WFS systems reported that the most interesting aspect is not the accurate movement of sounds nor their localization, but the way that an individual sound can create a space itself: "The WFS system definitely helps better perception of sound movement [..]. However, that doesn't mean that the music will be more interesting."

Besides the above-discussed concepts and products, the survey revealed that many tools are unknown to the majority of artists. Why are there so many relatively unknown tools, although even at least one of them is an award-winning[7] computer software? Are most users satisfied enough with their current choice of tools and not looking for other (perhaps more suitable) tools? Do composers rely mainly on audio sequencer software with integrated common spatialization features? The motivation to look for third-party tools that fit into the technical architecture of this particular audio sequencer may be diminished. The next section may help provide some insight.

[6]The composer Harrison (1999) for instance, referred to the way loudspeakers, e.g., in Ambisonics rendering systems, equidistantly surround the audience as the "Stonehenge deployment".

[7]ZKM Zirkonium won the 2nd prize LOMUS award 2008.

Choice of Tools

We asked participants about the rationale for choosing their current spatialization tools. The responses in comment form are analyzed and grouped together in Table 2.6. Almost every other response is related to the usability of tools. Simplicity, intuitive use or ease of use are common replies. Also, the challenge of learning how to use a tool is included in this category. Sixty-one percent of the people think the time spent with spatialization tools could be reduced with optimization. The importance of time and usability are likely connected to the pressure composers feel to meet commission deadlines and to maximize work and creative outcomes within limited studio time and resources. People who invest time in creating their own spatialization tools reported that the advantage of self-written software is to create a personal approach to spatialization and to have control over all the essential parameters, which suggests that readymade tools do not provide this. Every second respondent uses less features than their spatialization tools offer. The use of features is probably person-, situation- or context-dependent – not every feature needs to be always applied. However, perhaps current tools focus on some features while excluding other important or desirable features. Further research in human-computer interaction in collaboration with composers could improve usability of future spatialization tools. Thirty-six percent of respondents base their choice of tools on the degree to which a specific tool can be applied in achieving compositional and aesthetic goals. A long-time spatially experienced American composer explains, “I have a system that does everything I am interested in doing.”

Another question revealed that 30% of the artists are constrained by the number of sound sources that can be spatialized using one’s current palette of tools. Besides trying less limiting and more appropriate tools, faster and multiple CPUs will definitely also help these composers. Flexibility and versatility were mentioned in 29% of the responses, whereas only 12% addressed the reliability of their tools. One would expect more responses for this fundamental aspect. Flexibility and versatility seems reasonable, because many artists work on different kinds of projects (see Figure 2.2) that require different configurations of their working environment (i.e., context, musical material, venues, reproduction systems). Therefore their tools have to be able to accommodate different scenarios and circumstances. Also, for experimenting and developing ideas, tools needs to be flexible and easily usable. Seventeen percent of the responses mentioned the ability to integrate tools into the existing technical framework in a composition studio. For instance, because composition studios are hardly equipped with more than eight loudspeakers (see Figure 2.4), the use of a rendering tool that requires more than the available number of loudspeakers would be inappropriate. In another example, a tool that creates binaural audio signals is not useful if music is presented to a large audience over loudspeakers.

Table 2.6: What is your motivation for working with your current spatialization equipment (versus using other tools)? Forty-two respondents.

Motivation	Total responses	%
Usability, learning curve	20	48
Quality of spatialization, fit to aesthetic goals	15	36
Availability, accessibility and cost	15	36
Flexibility, versatility	12	29
Integration into existing technical framework	7	17
Reliability	6	12
Other	2	5

Importance of Technical Features

According to a list of 10 technical features, we asked respondents to rate within five categories the relative importance of these features for their work. The response categories were: Not important, Slightly, Fairly, Very, and Extremely important. The ratings result in relatively little variance, all technical features having an average around Fairly important (left side of Figure 2.8). The feature “Spatial Rendering in Real-time” received the highest ratings (Very important), whereas “Visual 3D representation of a sound scene” was rated lowest (less than Fairly important). The most frequent responses demonstrate that there are three features rated as Extremely important: “Integration into digital audio workstations as plug-ins”, “Controllability via external controller” and “Spatial rendering in real time”. In their comments, respondents added to the listed features entries such as “Level visualization of each speaker feed”, “Up- or downmixing to eight output channels”, “Managing trajectories, patterns, and direct control protocols within a database”, and “Adaptation to different loudspeaker configurations”. Some of the technical features correlate moderately with each other (Spearman correlation $\rho(52) \approx 0.5$). A cluster analysis was performed on the responses in order to group technical features together that were similarly rated (right panel of Figure 2.8). This cluster analysis shows the technical features that developers might want to conjointly address for different use cases. For instance, a spatialization tool that can be integrated as a plug-in into DAWs should be equipped with a graphical user interface, whereas an application that renders a spatial scene into an audio-file in non-real-time should be equipped with a visual representation of the sound scene. Another grouping of features addresses real-time rendering in combination with external controllers and the rendering of non-standardized speaker configurations. The advantage of real-time rendering is that spatial parameters can be manipulated, e.g., via a gestural controller in real time (Marshall et al. 2006). Furthermore, the rendering process can be adapted to accommodate a given loudspeaker arrangement and certain acoustic conditions.

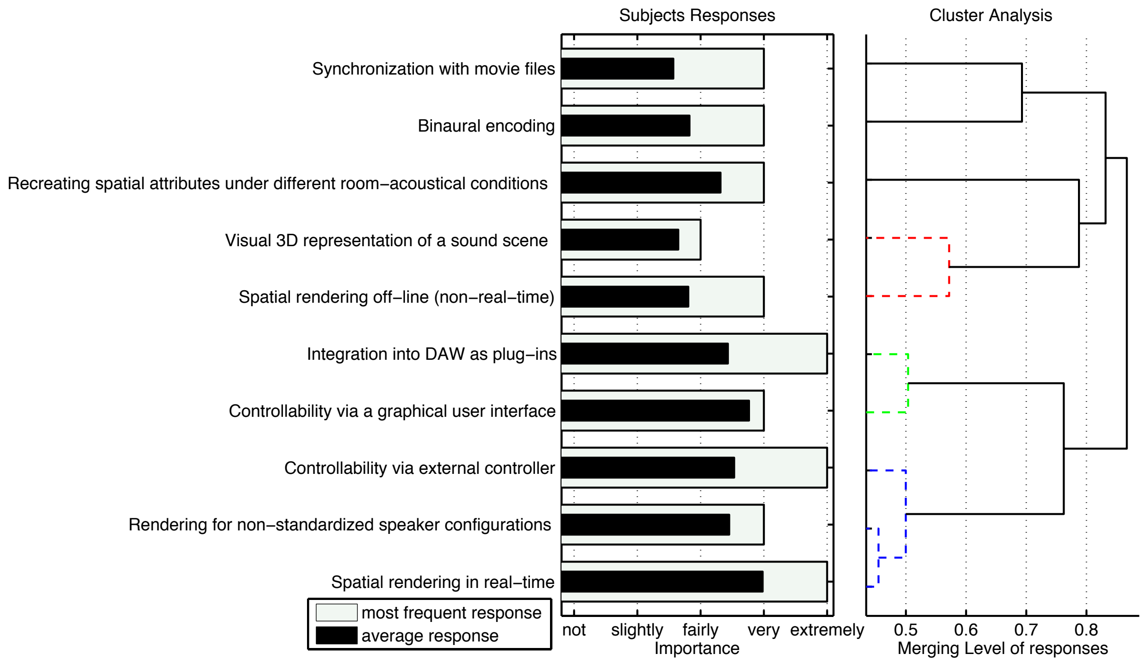


Figure 2.8: Rate the Importance you attach to the listed technical features.

Desired Features

When asked if features were missing, 41% responded "yes" while 20% said "no". Often an improvement in the usability of the spatialization tools was requested, such as "intuitive interfaces [...] to control spatialization processes from a high level" or "scalable interfaces that can go from 2 to 500 channels." Another respondent wishes to draw trajectories for multiple sound sources in one scene editor, rather than having multiple trajectory editors, one for each audio track, in a DAW. Others addressed the bus architecture in the DAW application, which often limits the number of loudspeaker feeds: "I prefer to use commercial DAW software, [...] but find most of the available multi-speaker tools inflexible and too cinematic." Another composer said that DAWs are too inflexible for live work and non-standardized playback scenarios. He misses features that facilitate flexible routing and control of stereo planes to various speaker sets. Generally a higher degree of flexibility was requested. One respondent said that each tool has strengths and weaknesses, and it is very difficult for him to imagine a tool that does everything perfectly. To benefit from the power of individual tools, he proposed a framework that interconnects them. This would appeal to another survey respondent who wants to map parameters of time, pitch, timbre and space in his music “through expressive software tools.” Other people would like to have tools that help adapt music to the varying acoustical and technical situations of the performance venues. Lastly, respondents expressed the desire to easily apply different interface devices for controlling spatialization, e.g., for drawing trajectories of sound sources. Besides using common human-computer interfaces, such as a joystick or keyboard, it was suggested to develop input devices that are tailored to the specific needs of controlling spatialization, e.g., multi-touch interfaces.

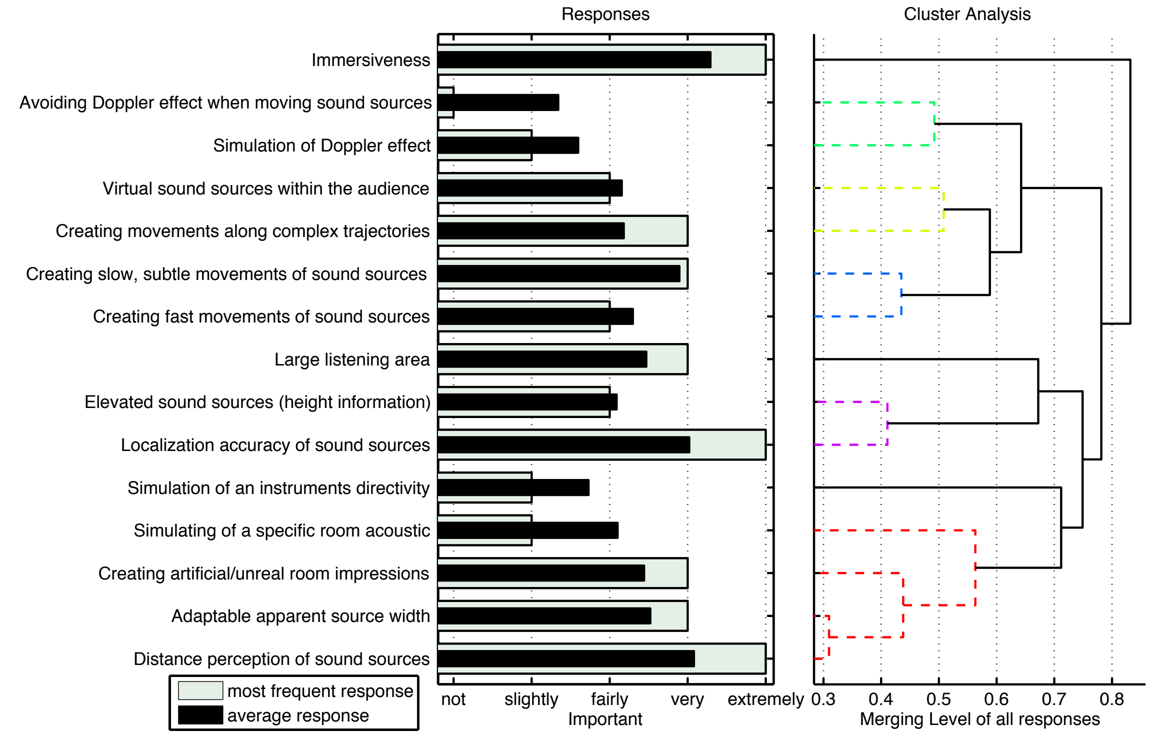
2.4 Analysis

2.4.1 Spatial Aspects: Compositional Importance and Their Fulfillment

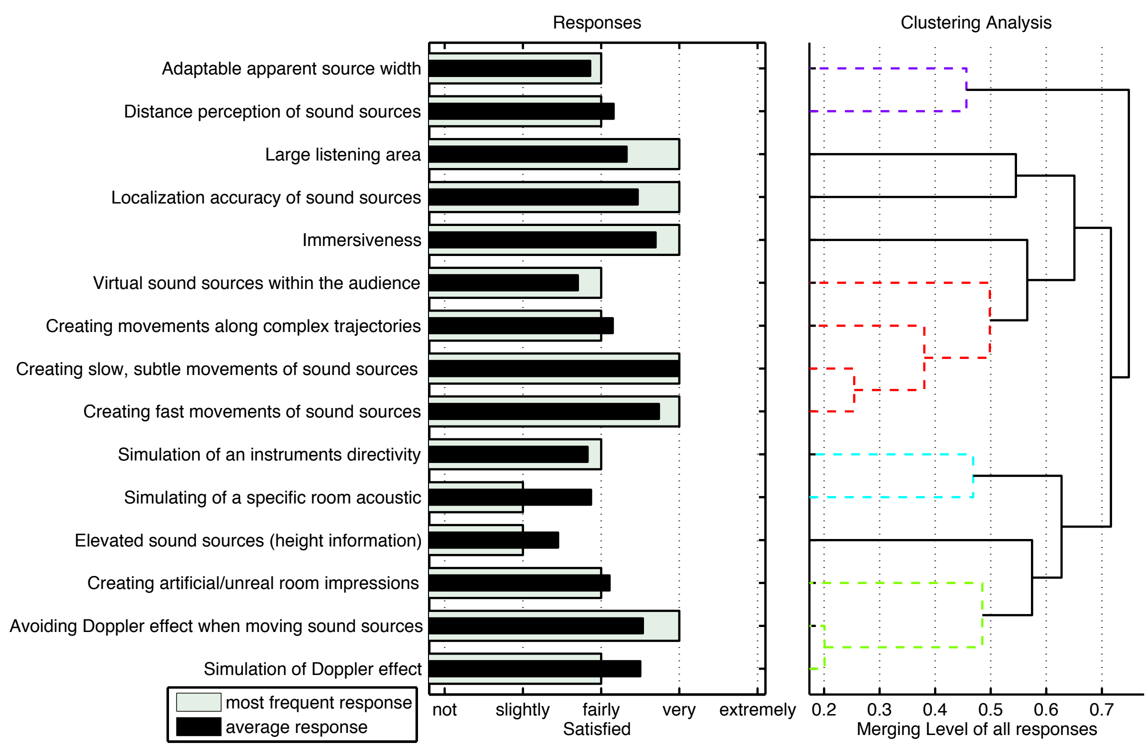
We provided a list of 15 spatial aspects to discover which ones artists consider to be important, and to what degree they can be effectively created through their tools. First, the importance of spatial aspects was addressed and the responses were categorized as: N/A, Not, Slightly, Fairly, Very, and Extremely important. The distribution of responses was N/A 6%, Not 10%, Slightly 17%, Fairly 20%, Very 24% and Extremely 23%. Hence, almost every second rating was Extremely or Very important (left panel of Figure 2.9(a)). Some composers mentioned that the degree of importance can change according to the compositional situation and the musical material. Other comments provided additional aspects not included in this question. These aspects include “Spatial clarity and density”, “Timbre spatialization” as described by Normandeau (2009) and “Perspective”, maybe related to the perspectives in spaces as elaborated by Stockhausen in Cott (1973, 45-46). The highest rated spatial aspects are “Immersiveness”, “Distance perception of sound sources” and “Localization accuracy of sound sources”. The aspect “Large listening area”, a feature of emerging high-resolution reproduction techniques (WFS, HOA), was rated as Very to Fairly important. The aspects of “Avoiding” or “Simulating a Doppler effect”, a natural pitch change of fast moving sounds, were surprisingly rated as least important. Considering that artists who were new to WFS reported disturbing sound coloration when moving sounds, one would have expected this feature to be rated with a higher importance. The higher ratings of “Creating slow, subtle movements of sound sources” compared to “Creating fast movements” suggests that the unwanted Doppler effect might occur less due to the preference for slow movements, which minimize this percept. The aspect “Virtual sources within the audience” (also associated with WFS and HOA) was rated similarly to “Elevated sound sources” as Fairly important.

The right panel of Figure 2.9(a) visualizes the cluster analysis performed on the responses according to a Spearman correlation. “Distance perception” of sound sources was very similarly rated to “Adaptable apparent source width” (ASW) and also to the aspects about creating room impressions, e.g., through synthetic reflection patterns. Another interesting cluster was created from “Elevated sound sources” and “Localization accuracy”. It seems that composers favoring accurate localization are also interested in perceiving elevated sounds. Remarkably, “Immersiveness”, the highest rated aspect, is unrelated to any other presented aspect.

After rating the importance of spatial aspects, respondents were asked to rate in a similar way their satisfaction with the ability of their spatialization tools to (re)create those spatial aspects. The intention behind this question was to find out if there is a gap between the creative desire and the ability to achieve compositional aims with available tools. The distribution of the responses among the satisfaction categories was N/A 18%, Not 8%, Slightly 14%, Fairly 23%, Very 27%, and Extremely satisfied 9% (left panel of Figure 2.9(b)). On average, six of the fifteen listed spatial aspects that can be produced by the currently used tools were rated Very satisfied. Five aspects have mean ratings below Fairly satisfied: “Elevated sound sources”, “Virtual sound sources within the audience”, “Simulation of a specific room acoustic”, “Simulation of an instrument’s directivity”, and “Adaptable apparent source width”. As previously described, a cluster analysis was performed on the ratings. The resulting dendrogram (right panel of Figure 2.9(b)) shows similarities to the cluster analysis of the importance ratings (Figure 2.9(a)), e.g., aspects regarding the Doppler effect and source movements were also found to be correlated.



(a) Rate the importance you attach to the listed spatial feature.



(b) How satisfied are you with the ability of your preferred spatialization tools to produce the listed spatial aspects?

Figure 2.9: Importance vs. Satisfaction ratings of spatial aspects.

Comparison

In comparing the average responses in Figure 2.9(a) with Figure 2.9(b), one sees that generally the satisfaction ratings are lower than the associated ratings about importance. Figure 2.10 yields insight into the inter-individual differences by showing the ratings of three composers with different levels of spatialization experience. The slight tendency is that with more experience, the differences between importance and satisfaction ratings increase. “Nothing is perfect - dissatisfaction is a state of mind” commented composer C from Figure 2.10. In contrast, a generally satisfied artist with less years of experience than composer C said that he can usually get adequate results. His two most important spatial aspects are “Immersiveness” and “Creating slow subtle movements of sound sources”.

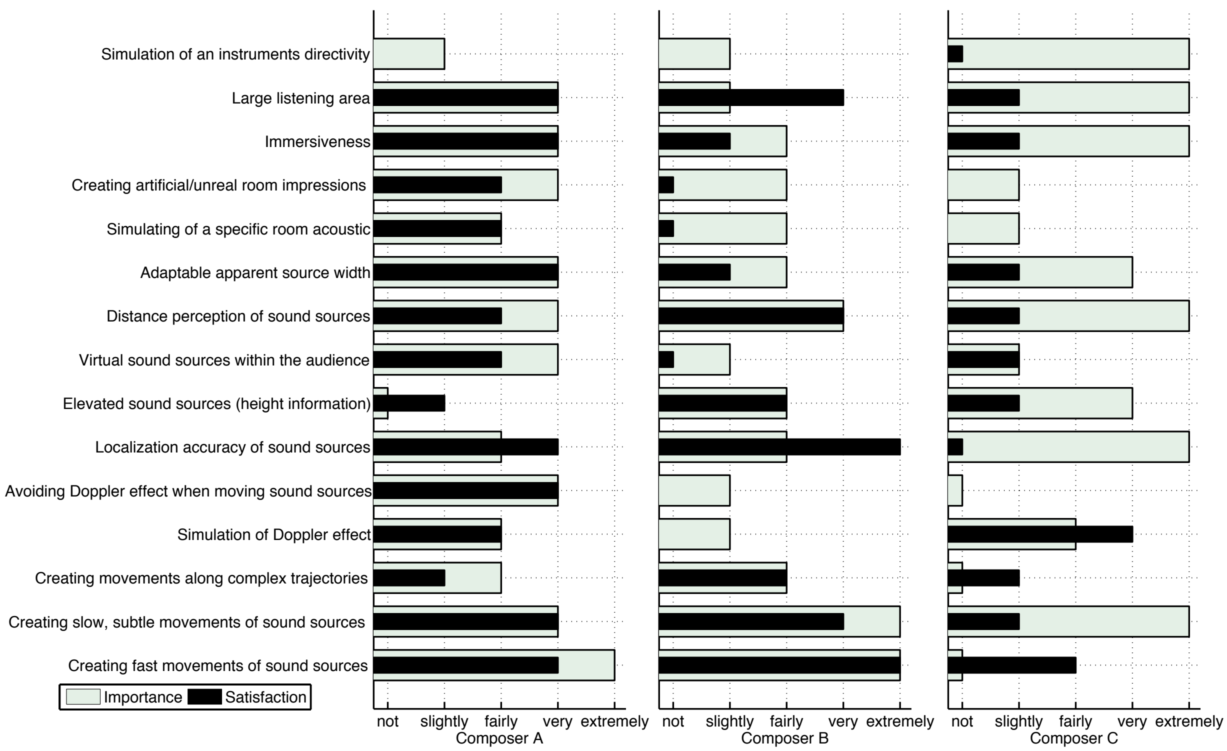


Figure 2.10: Spatial Aspects, comparison of Importance and Satisfaction ratings from three different composers. Composer A: 1 year of experience; Composer B: 8 years of experience; Composer C: 30 years of experience.

For research and development in spatial audio, the aspects that were rated with the highest Importance but with low Satisfaction indicate a demand for better tools. There would be less interest in investing in further improvements of spatial aspects that received either a high satisfaction rating or a low importance rating. Table 2.7 shows the region of potential research and development interest, which are marked with a “ϟ”, whereas the regions not requiring improvement are marked with a “+” (highest importance and satisfaction values). The responses by the respondents were sorted according to this matching matrix.

Table 2.7: Matching Matrix showing the regions of interest. “+” symbolizes cells with the best values in Importance and Satisfaction; “ϟ” symbolises cells related to a low Satisfaction but high Importance.

Importance \ Satisfaction	not	slightly	fairly	very	extremely
not					
slightly					
fairly	ϟ				+
very	ϟ	ϟ		+	+
extremely	ϟ	ϟ	ϟ	+	+

For instance, composer A’s rating of the “Large Listening Area” shown in Figure 2.10 fits into a +-region (Very important and Very satisfied), whereas composer C’s rating would be sorted into a ϟ-region (Very important and Slightly satisfied). The distribution of all responses can be seen in Figure 2.11. The higher the blocks, the higher the number of responses in this Importance/Satisfaction category. One can see that many ratings were given with the combination Extremely important/Very satisfied and Very important/Very satisfied. The middle-ground responses Fairly important/Fairly satisfied and Very important/Fairly satisfied were less frequent. Figure 2.12 shows the center of gravity, a geometric measure that roughly indicates about the central tendency of the ratings and tends on average, toward "Very important" and "Fairly satisfied". This figure also shows the center of gravity of the responses for each composer and its relation to the average response. One can see according to the different experience groups that with greater experience, composers tend to deviate more strong from the mean center of gravity, suggesting that with more experience in spatialization, the individuality of the ratings increases. The ϟ-region in Table 2.7 represents more than 17% and +-region represents about 32% of all responses. An examination of these response regions according to the spatial aspects is shown in Figure 2.13. In the ϟ-region, the aspects with the most responses are "Elevated sound sources", "Distance perception", and "Virtual sound sources within the audience". In the +-region, the most frequent aspects are "Creating slow subtle movements of sound sources", "Immersiveness" and "Localization accuracy of sound sources". The two latter aspects were also judged with Very high importance ratings (Figure 2.9(a)). Sound-source localization is traditionally a strong field in psychoacoustics (e.g, Blauert 1997) and a lot of research in spatial audio research has evaluated this aspect (e.g., Pulkki and Hirvonen 2005; Marentakis et al. 2008a). The analysis of the composers' responses suggests that research and development efforts have benefitted from the field of musical applications. Remarkably, the aspects "Distance perception of sound sources" and "Creating unreal room responses" have relatively similar contributions in both regions, indicating a polarization of the responses.

2.5 Conclusion and Recommendations

The high response rate of this survey suggests that this kind of questionnaire is well received by artists. Researchers from related fields might be encouraged to similarly gather feedback from artists. The responses of 52 composers regarding technical and compositional aspects

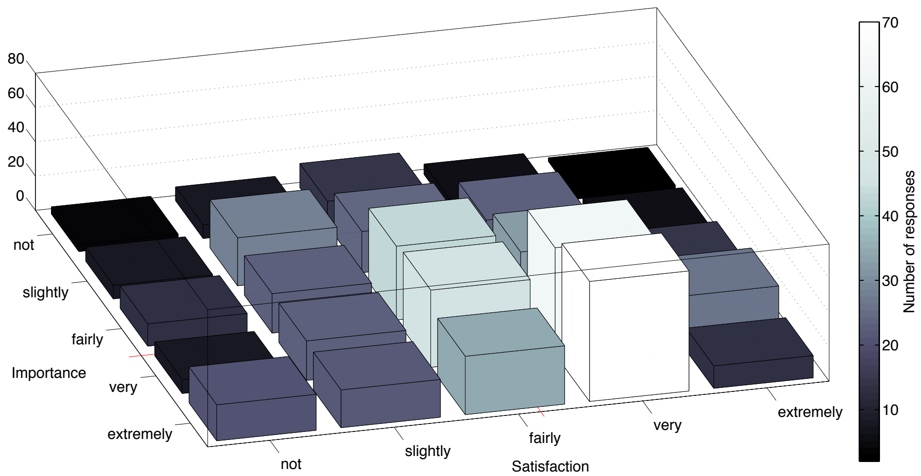


Figure 2.11: Spatial aspects: All responses are sorted according to importance and satisfaction ratings. The bar height indicates the number of responses in a given category.

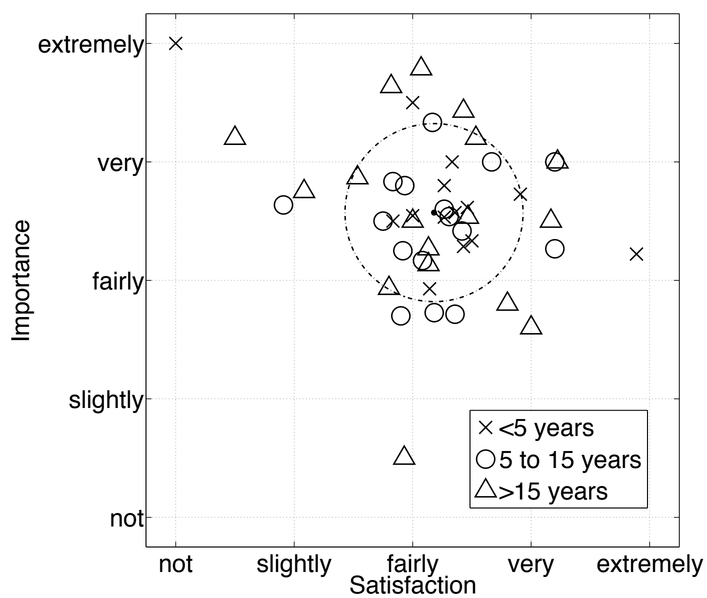


Figure 2.12: Spatial aspects: center of gravity of the responses according to importance and satisfaction per respondent. The dash-dotted circle shows the average distance of the responses to the mean center of gravity (the center of the dash-dotted circle).

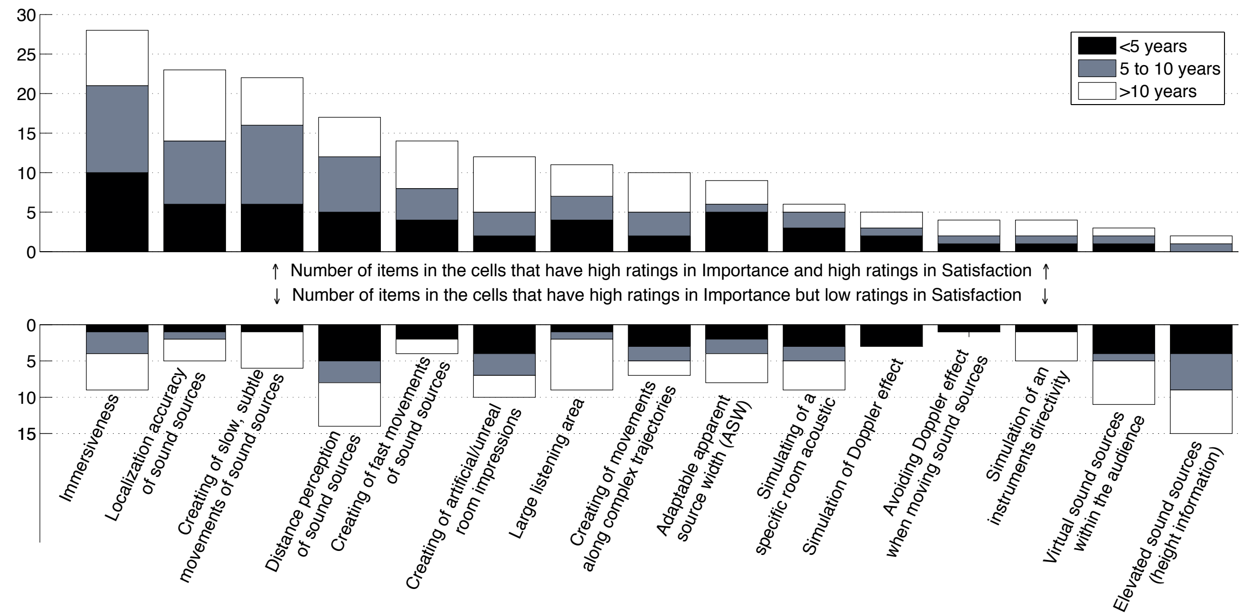


Figure 2.13: Spatial aspects: items within the best and worst cells according to Table 2.7. Y-axis: number of responses. The barplot also shows the responses according to the experience groups.

were analyzed to find general tendencies in current usage, while acknowledging the artistic individuality of each composer. According to our interpretation of the survey findings, we derived the following recommendations for collaborative work between composers and researchers.

The technical and practical challenges that the application of multichannel sound reproduction systems can create for some artists is related to their under-utilizing the available spatial aspects. To address these challenges, one has to acknowledge the higher technical complexity of multi loudspeaker setups, especially for the emerging high-quality spatialization techniques WFS and HOA, which require careful calibration of the equipment to take full compositional advantage of them. In order to familiarize composers with new technology, the learning curve must be kept reasonably shallow. Good usability (e.g., avoiding cumbersome command line control), as well as the possibility to integrate new tools into common compositional environments, are paramount and can lower the entry barriers for artists. Many DAWs (the most commonly used compositional environment, Figure 2.7), are limited to eight-channel spatialization. To accommodate the needs of high-quality spatial rendering concepts, the bus architecture must be extended to allow for massive output channels. We also saw the demand for technology to give composers a feeling of the venue acoustics while working in the studio. Regarding new technologies, some responses suggest that a group of artists is motivated and technically experienced enough to become "early adopters" of new or unreleased tools in order to explore their artistic potential. Real-world loudspeaker setups often differ from the standardized systems usually employed in listening experiments. More ecologically valid speaker configurations in the labs yield more meaningful data that better generalize to real-world conditions. The first part of the survey tells us where, why and how spatial concepts are applied. This information can be used by researchers and developers to create meaningful test environments which approximate real-world scenarios in the labs. Composers could support this effort by making parts of a composition accessible to researchers and by supporting the development of a spatial-notation or description systems as proposed by Kendall et al. (2008). Apart from increasing the potential of preserving a composition, through a common description format, developers could more authentically re-render spatial music and evaluate novel rendering methods or loudspeaker configurations, for instance. It is necessary to investigate how diffusion practice, as a prominent form of sound spatialization, can be incorporated into notation/description approaches.

2.6 Summary

This paper presented an analysis of a survey on the compositional use of spatialization by composers. The purpose of the survey was to give an overview of the current state of practice in order to guide future research and development of spatial audio technology. The respondents showed a great diversity in terms of experience in composing, age, place of education and residence. Therefore, we believe that the responses represent a meaningful cross-section on 24 questions about compositional and technical aspects of spatialization. Besides the expected individual differences in composers' responses, we also extracted common themes in motivation, compositional practice, preferences and critiques of available audio technologies. We hope that our findings help enable communication between artists and researchers, in order to refine spatial audio technologies that will enhance future artistic practice.

2.7 Acknowledgement

This work was funded by a grant from the Canadian Natural Sciences and Engineering Research Council and the Canada Council for the Arts (NSERC, CCA) to Stephen McAdams and by the Centre for Interdisciplinary Research in Music Media and Technology (CIRMMT). Thanks to Catherine Guastavino, Matthias Geier, Sandra Duric and Finn Upham for discussions and suggestions during the different phases of this research.

Part II

Development of Spatialization Tools

Chapter 3

A Stratified Approach for Sound Spatialization

The following Chapter is based on the peer-reviewed publication:

Peters, N., Lossius, T., Schacher, J., Baltazar, P., Bascou, C., and Place, T. (2009). A stratified approach for sound spatialization. In *Proc. of the 6th Sound and Music Computing Conference (SMC)*, pages 219–224, Porto, PT.

This paper introduces a multi-layer structure to mediate essential components in sound spatialization. This approach will facilitate artistic work with spatialization systems, a process which currently lacks structure, flexibility, and interoperability.

3.1 Introduction

The improvements in computer and audio equipment in recent years make it possible to experiment more freely with resource-demanding sound synthesis techniques such as spatial sound synthesis, also known as spatialization. To seek new means of expression, different spatialization applications should be able to be readily combined and accessible for both programmatic and user interfaces. Furthermore, quantitative studies on spatial music (see Otondo 2008; Peters et al. 2011) remind us that there are great individual and context-related differences in the compositional use of spatialization and that there is no one spatialization system that could satisfy every artist. In an interactive art installation, the real-time quality of a spatial rendering system in combination with the possibility to control spatial processes through a multitouch screen can be of great importance. In contrast, the paramount features in a performance of a fixed-media composition may be multichannel playback and the compensation of non-equidistant loudspeakers (in terms of sound pressure and time delays). Additional scenarios may require binaural rendering for headphone listening, multichannel recording, up and down mixing, or a visual representation of a sound scene. Moreover, even during the creation of one spatial art work, the importance of these requirements may change throughout different stages of the creative processes.

Guaranteeing efficient workflow for sound spatialization requires structure, flexibility, and interoperability across all involved components. As reviewed in the following section, common spatialization systems too often give no consideration to these requirements.

3.2 Review of Current Paradigms

3.2.1 Digital Audio Workstations–DAW

Many composers and sound designers use DAWs for designing their sound spatialization primarily in the context of fixed media, tape-music, and consumer media production. A number of DAWs are mature and offer a systematic user interface, good project and sound file management, and extendability through plug-ins to fulfill different needs.

DAWs mainly work with common consumer channel configurations: mono, stereo and 5.1. However, through focusing on consumer media products, multichannel capabilities are limited. ITU 5.1 (1992), a surround sound format with equidistant loudspeakers around an ideally located listener, is the most common multichannel format. Its artistic use may be limited because 5.1 favors the frontal direction and has reduced capabilities for localizing virtual sources from the sides and back. Recent extensions up to 10.2 are available[1], but are insufficient for emerging reproduction techniques such as Wave Field Synthesis (WFS) or Higher-Order Ambisonics (HOA). Also, in art installations or concert hall environments, non-standard loudspeaker setups are common due to artistic or practical reasons, varying in number and arrangements of loudspeakers. These configurations are not typically taken into account in DAWs and are therefore often difficult to use.

DAW surround panners often comprise a parameter named *blur*, *divergence*, or *spread* that controls the apparent source width through modifying the distributed sound energy among loudspeakers. Although this parameter enriches the creative possibilities, it is often either missing or only indirectly accessible, e.g., by changing the distance of the sound source.

3.2.2 Media Programming Environments

Various media programming environments exist that are capable of spatial sound synthesis, e.g., SuperCollider, Pure Data, OpenMusic, and Max/MSP. In order to support individual approaches and to meet the specific needs of computer music and mixed media art, these environments enable the user to combine music making with computer programming. While aspiring to complete flexibility, they end up lacking structured solutions for the

[1]A comparison of DAWs concerning their multichannel audio capabilities can be seen on http://acousmodules.free.fr/hosts.htm, accessed Jun 2010.

specific requirements of spatial music as outlined in Section 3.1. Consequently, numerous self-contained spatialization libraries and toolboxes have been created by artists and researchers to generate virtual sound sources and artificial spaces, such as Space Unit Generator (Yadegari et al. 2002), Spatialisateur (Jot 1992), or ViMiC (Braasch et al. 2008). Also toolboxes dedicated to sound diffusion practice have been developed, e.g., BEASTmulch System[2], ICAST (Beck et al. 2006). Each tool, however, may only provide solutions for a subset of compositional viewpoints. The development of new aesthetics through combining these tools is difficult or limited due to their specific designs.

3.2.3 Stand-alone Applications

A variety of powerful stand-alone spatialization systems are in development, ranging from directional-based spatialization frameworks, e.g., SSR (Geier et al. 2008), Zirkonium (Ramakrishnan et al. 2006), and Auditory Virtual Environments (AVE), e.g., tinyAVE (Borß and Martin 2009) to sound diffusion and particle oriented approaches, e.g., Scatter (McLeran et al. 2008). Although these applications usually promote their graphical user interfaces as the primary method to access their embedded DSP-algorithms, one can find a few strategies that allow external communication through self-contained XML, MIDI or OSC (Wright and Freed 1997) protocols.

3.3 A Stratified Approach to the Spatialization Workflow

When dealing with spatialization in electroacoustic composition or linear sound editing, the workflow comprises a number of steps in order to construct, shape and realize the spatial qualities of the work. The creative workflow might appear to be different when working on audio installations or interactive/multimedia work. Still, we have identified underlying

[2] http://www.beast.bham.ac.uk/research, accessed Jun 2010

common elements when spatialization is used. For this reason we propose a stratified approach, where the required processes are organized according to levels of abstraction.

This model is inspired by the Open Systems Interconnection network model (OSI)[3], which is an abstract description for layered communications and computer network protocol design. OSI divides network architecture into seven layers that range from top (Application) to bottom (Physical Layers). Each OSI-layer contains a collection of conceptually similar functionalities that provide services to the layer above it and receive services from the layer below it. As depicted in Figure 3.1, six layers have been defined in our model.

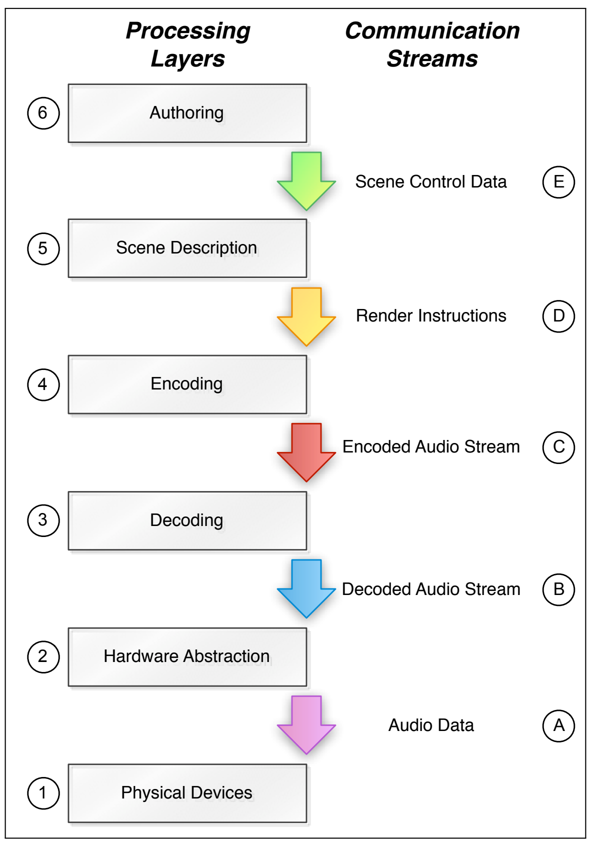


Figure 3.1: Layers and streams in sound spatialization.

[3] http://en.wikipedia.org/wiki/OSI_model, accessed Jun 2010.

3.3.1 Physical Device Layer

The major functionality of this layer is to establish the acoustical connection between computer and listener. It defines the electrical and physical specifications of devices that create the acoustical signals, such as soundcards, amplifiers, loudspeakers, and headphones.

3.3.2 Hardware Abstraction Layer

This layer contains the low level audio services that run in the background of a computer OS and manages multichannel audio data between the physical devices and higher layers. Examples are Core Audio or ALSA. Prominent software protocols that communicate with those low level audio services are PortAudio, JACK, Soundflower or Rewire which can be used for more complex distributions of audio signals among different audio clients.

3.3.3 Encoding and Decoding Layers

In the proposed model, the spatial rendering is considered to consist of two layers. The Encoding Layer produces encoded signals containing spatial information while remaining independent of and unaware of the speaker layout. The Decoding Layer interprets the encoded signal and decodes it for the speaker layout at hand. According to Wiggins (2004, 99), this makes the creative process and the created piece more portable and future-proof, because different speaker layouts can be used as long as a decoder is available. Examples of such hierarchical rendering methods are Ambisonics B-Format, Higher Order Ambisonics, DirAC (Pulkki 2007), MPEG Surround, AC-3, or DTS.

Not every rendering technique generates intermediate encoded signals, but instead can be considered to encapsulate the Encoding and Decoding Layers in one process. Some examples of such renderers are VBAP (Pulkki 1997), DBAP (Lossius et al. 2009), ViMiC (Braasch et al. 2008) and Ambisonics equivalent panning (Neukom and Schacher 2008).

Processing of sources to create an impression of distance, such as Doppler effect, gain attenuation and air absorption filters, are considered to belong to the Encoding Layer, as does the synthesis of early reflections and reverberation, i.e. as demonstrated by surround effects that employ B-format impulse-response convolution.

3.3.4 Scene Description Layer

This layer mediates between the Authoring Layer above and the Decoding Layer below through an abstract and independent description about the spatial scene. This description can range from a simple static scene with one virtual sound source up to complex dynamic audio scenes including multiple virtual spaces. This data could also be stored to recreate spatial scenes in a different context. Specific (lower-level) rendering instructions are communicated to the Encoding Layer beneath. Examples are ASDF (Geier et al. 2008), OpenAL (Hiebert 2005) or SpatDIF (Peters 2008).

3.3.5 Authoring Layer

This layer contains all software tools for the end-user to create spatial audio content without the need to directly control underlying processes. Although these tools may remarkably differ from each other in terms of functionality and interface design to serve the requirements for varicolored approaches to spatialization, the communication to the Scene Description Layer must be standardized. Examples are symbolic authoring tools, generative algorithms, and simulations of emergent behaviors (swarms or flock-of-birds); or, more specifically as discussed below, Holo-Edit, and ambimonitor/ambicontrol.

3.3.6 Concluding Remarks

OSI provided the idea that each layer has a particular role to play. The stratified model does not enforce one particular method for each layer; rather, a layer offers a collection of conceptually similar functions. This is analogous to how TCP and UDP are alternative protocols working at the Transport Layer of the OSI model.

Spatialization processes should be modularized according to the layered model when feasible. With standardized communication between the layers, one method for a layer can easily be substituted for another, enhancing a flexible workflow that can rapidly adapt to varying practical situations and needs.

3.4 Stratified Tools

Below, the authors discuss several of their developments that strive to establish and evaluate the proposed stratified concept.

3.4.1 SpatDIF

The goal of the Spatial Sound Description Interchange Format (SpatDIF) is to develop a system-independent language for describing spatial audio (Peters 2008) that can be applied around the Scene Description Layer to communicate between authoring tools down to the Encoding/Decoding Layers.

Formats that integrate spatial audio descriptors, such as MPEG-4 (Vaananen and Huopaniemi 2004) or OpenAL (Hiebert 2005), did not fully succeed in the music or fine arts community because they are primarily tailored to multimedia or gaming applications and don't necessarily consider the special requirements of spatial music, performances in concert venues, and site-specific media installations. To account for these specific requirements, the SpatDIF development is consequently a collaborative effort that jointly involves researchers and artists.

A database[4] has been created to gather information about syntax and functionalities of common spatialization tools and to identify the lowest common denominator, the "Auditory Spatial Gist", for describing spatialized sound. Besides these essential Core Descriptors, a number of extensions have been proposed to systematically account for enhanced features, for example: the Directivity Extension, which deals with directivity information of a virtual sound source; or the Acoustic Spaces Extension that contains acoustical properties of virtual rooms; or the Ambisonics Extension that handles ambisonics-only parameters. The latter is an example where SpatDIF mediates between the processing layers, starting from Layer 3 to Layer 6.

Although SpatDIF does not imply a specific communication protocol or storage format at present, OSC for streaming and SDIF (Schwarz and Wright 2000) as a storage solution are used for piloting.

3.4.2 ICST Ambisonics

The ICST Ambisonics Tools is a set of externals for Max/MSP (Schacher and Kocher 2006). The DSP externals ambiencode∼ and ambidecode∼ generate and decode Higher Order Ambisonics and are part of the Encoding and Decoding Layer.

Ambimonitor and ambicontrol complete the set as control tools for the Authoring Layer. Ambimonitor generates coordinate information for the DSP objects, presents the user with a GUI displaying point sources in an abstract 2D or 3D space, and is equipped with various key commands, snapshot and file I/O capabilities. Ambicontrol provides a number of methods that control the motion of points in the ambimonitor's dataset. Automated

[4] http://redmine.spatdif.org/wiki/spatdif/SpatBASE, accessed Jun 2010

motions, such as rotation or random motion, optionally constrained in bounding volumes and user defined trajectories can be applied to single or grouped points. Trajectories, and state snapshots can be imported/exported as an XML file, which will be replaced with SpatDIF compliant formatting in a later release.

A novel panning algorithm (Neukom and Schacher 2008) was derived from in-phase ambisonics decoding and implemented as a Max/MSP external entitled ambipanning∼. It encapsulates the Encoding and Decoding Layer by transcoding a set of mono sources in one process onto an ideally circular speaker setup with an arbitrary number of speakers. The algorithm works with a continuous order factor, allowing the use of individually varying directivity responses.

3.4.3 Jamoma

Jamoma[5] is a framework (Place and Lossius 2006) for structuring and controlling modules in Max/MSP. Work on spatialization has been of strong interest to several of the developers, and solutions for spatialization in Jamoma have a stratified approach in accordance with the proposed model.

The Max/MSP signal-processing chain only passes mono signals, and for multichannel spatial processing the patch has to be tailored to the number of sources and speakers. If Max/MSP is considered a programming environment and the patch is the program, a change in the number of sources or speakers requires a rewrite of the program, not just a change to one or more configuration parameters. Jamoma addresses this by introducing multichannel audio signals between modules with all channels wrapped onto a single patch cord. Jamoma Multicore[6] is being developed as a more robust solution than the current approach for handling multichannel signals, which are also used between the Encoding,

[5] http://www.jamoma.org, accessed Jun 2010

[6] http://github.com/tap/JamomaAudioGraph, accessed Jun 2010

Decoding and Hardware Abstraction Layers.

Jamoma modules have been developed to convert multichannel signals, play and record multichannel sound files, perform level metering and pass multichannel signals on to the sound card or virtual auxiliary bus. These are supplemented by modules compensating for sound-pressure and time-delay differences in non-equidistant loudspeaker arrangements.

Ambisonics is the only spatialization method implemented in Jamoma that separates spatial encoding and decoding. First- to Third-order B-format encoding of mono sources is implemented using the ICST externals (Schacher and Kocher 2006). Other modules are available to encode recordings made with the Zoom H2 and to encode UHJ signals. Encoded signals can be manipulated, i.e. the balance between the encoded channels can be adjusted, or the encoded signal can be rotated, tilted and tumbled. The decoding module for up to third-order B-format signals uses the ICST externals, whereas a module for binaural decoding uses Spatialisateur (Jot 1992). B-format signals can also be decoded to UHJ.

Several other popular spatialization algorithms are available as Jamoma modules: VBAP (Pulkki 2000), ViMiC (Braasch et al. 2008) and DBAP (Lossius et al. 2009). Consequently, one rendering technique can easily be substituted for another, or several rendering techniques might be used in tandem for a variety of spatial expressions, analogous to how an artist will use many different brushes in one artwork.

Prior to rendering, additional modules offer Doppler, air absorption and distance attenuation source pre-processing. All modules operating at the Encoding Layer are SpatDIF-compliant and hence provide the same interface for controlling modules operating at higher layers.

At the Scene Description Layer, a module provides a simple interface for defining the position of sources. The same module can be used to set loudspeaker positions for the

Decoding Layer.

At present, two modules operate at the Authoring Layer; Boids simulates co-ordinated animal motion and a scene manipulator allows geometric transformations (e.g., scaling, skewing, rotation) and stochastically driven manipulations of the whole scene in three dimensions. In addition, Jamoma can be bridged to Holo-Edit as discussed in the next section.

3.4.4 GMEM Holo-Edit

Initiated by Pottier (1998), Holo-Edit is part of the GMEM Holophon project and conceptualized as an authoring tool for spatialization.

This standalone application uses the timeline paradigm found in traditional DAWs to record, edit, and play back control data. The data is manipulated in the form of trajectories or sequences of time-tagged points in a 3D space, and the trajectories can be generated or modified by a set of tools allowing specific spatial and temporal behaviors including symmetry, proportion, translation, acceleration, and local exaggeration. Different scene representation windows allow the user to modify data from different (compositional) viewpoints: *Room* shows a top view of the virtual space, the *Time Editor* shows the traditional DAW automation curve view and, finally, the *Score Window* represents the whole composition in a multi-track block-based view. Holo-Edit's space and time representations are generic and can be adapted to any renderer at the Encoding Layer. To allow precise alignment of sound cues to desired spatial movements, waveform representations of sounds and associated trajectories are displayed and can be edited together.

Holo-Edit uses OSC for communicating with the desired spatial sound renderer. Here, the main challenge is to adapt and format the data stream that fits the specific rendering algorithm syntax (e.g., coordinate system, dimensions, units). To overcome this challenge,

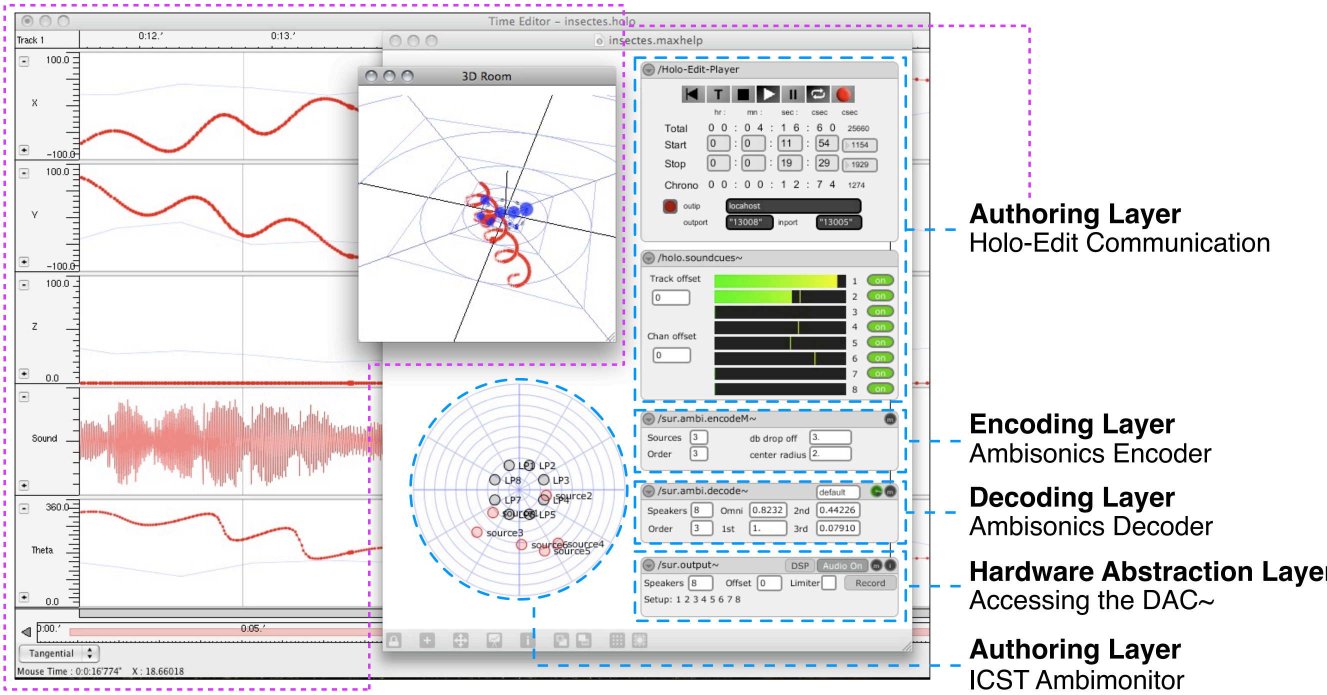


Figure 3.2: Holo-Edit, Jamoma and ICST Ambisonics Tools unified.

a Holo-Edit communication interface that handles sound file playback and position data of loudspeakers and sound sources through its standardized OSC-namespace was developed for the Jamoma environment. Therefore, Holo-Edit can be used as the main authoring tool for spatialization, whereas all DSP audio processes are executed in Jamoma (Figure 3.2). The communication between Holo-Edit and Jamoma is full-duplex, thus also enabling the recording of trajectories in Holo-Edit from any real-time control interface addressable through Jamoma.

3.5 Discussion And Conclusion

The examples from the previous section illustrate that a stratified model can be fruitful for development within media programming environments. The modular framework TANGA (Reiter 2007) for interactive audio applications reveals a related separation of tasks.

A few stand-alone applications are designed with a similar layered approach that allows control of different spatial rendering algorithms from one common interface, (e.g., Geier et al. 2008). Artists and researchers would benefit greatly if all these "local solutions" could be accessed by any desired authoring tool and integrated into existing environments.

An ICMC 2008 panel discussion on interchange formats for spatial audio scenes[7] (Kendall et al. 2008) and further informal discussion showed that adequate spatialization tools for working in DAWs are missing, but strongly desired. The proposed stratified approach would be more flexible than the current DAW architecture, in which tools for spatialization are tied to a number of consumer channel configurations. The object-oriented mixer approach proposed in Meltzer et al. (2008) suggests that stratification can be employed in DAWs. A potential limitation might be imposed by the fact that automation in DAWs is generally represented as time-tagged streams of one-dimensional values, whereas spatial information is generally multidimensional.

One keystone may be to define and agree on a meaningful communication format for spatialization. Therefore SpatDIF needs to be further developed to culminate in an API that can be easily integrated into any spatialization software.

3.6 Acknowledgment

The concepts proposed in this paper were to a large degree developed during a workshop at GMEA Centre National de Création Musicale as part of the Virage research platform funded by the French National Agency for Research. This work is also partly funded by the Canadian Natural Sciences and Engineering Research Council (NSERC), the Canada Council for the Arts, the COST IC0601 Action on Sonic Interaction Design (SID) and the Municipality of Bergen.

[7] http://redmine.spatdif.org/wiki/spatdif/Belfast_2008, accessed Jun 2010

Chapter 4

Spatial Sound Rendering with ViMiC in Max/MSP

The following Chapter is based on the peer-reviewed publication:

Peters N., Matthews, T., Braasch, J., and McAdams, S. (2008). Spatial Sound Rendering in Max/MSP with ViMiC. In *Proc. of the International Computer Music Conference (ICMC)*, pages 755–758, Belfast, UK.

The paper presents the Virtual Microphone Control (ViMiC) sound rendering technique implemented by Nils Peters and research assistant Tristan Matthews for the real-time media programming environment Max/MSP (Zicarelli 1998).

Unfortunately, due to a technical error on the part of the editors, this paper appeared in the electronic proceedings of the ICMC under the title "ViMiC - A Novel Toolbox for Spatial Sound Processing in Max/MSP". The user manual for this implementation can be found at: `http://github.com/Nilson/ViMiC-and-friends/raw/master/ViMiC_manual.pdf`.

Abstract

ViMiC (Virtual Microphone Control) is a tool for real-time spatialization synthesis, particularly for concert situations and site-specific immersive installations, and for larger or non-centralized audiences. Based on the concept of virtual microphones positioned within a virtual 3D room, ViMiC supports loudspeaker reproduction up to 24 discrete channels for which the loudspeakers do not necessarily have to be placed uniformly and equidistant around the audience. Originally conceptualized by Braasch (2005) for the software Pure Data, ViMiC was recently refined and extended with regards to usability, efficiency, and sound rendering quality for release to the Max/MSP community. Through the integrated Open Sound Control protocol (OSC, Wright and Freed 1997), ViMiC can be easily accessed and controlled.

4.1 Introduction

Besides the traditional concepts of pitch, timbre, and temporal structures, composers have long felt the desire to integrate a spatial dimension into their music. First, through static placement and separation of musicians in the concert space, and later, through dynamic modifications of the sound source position, effects of spatial sound segregation and fusion were discovered. In the 20th century, especially due to the invention and integration of microphones and loudspeakers into musical performance, spatialization has become popular. One of the earliest composers using the newly available electronic tools was Karlheinz Stockhausen. For his composition *Kontakte* (1958-60), he developed a rotational table, mounting a directed loudspeaker surrounded by four stationary microphones that receive the loudspeaker signal. The recorded microphone signals were routed to different loudspeakers arranged around the audience. Due to the directivity and separation of the microphones, the recorded audio signals contained Inter-Channel Time Differences (ICTDs) and Inter-Channel Level Differences (ICLDs). Depending on the velocity of the speaker rotation, the change in ICTDs can create an audible Doppler effect. To some degree, ViMiC follows Stockhausen's tradition and pioneering works by Steinberg and Snow (1934), Chowning (1971), and Moore (1983) by using the concept of spatially displaced microphones for the purpose of sound spatialization.

4.2 Spatialization with Max/MSP

This section briefly describes available loudspeaker spatialization techniques for Max/MSP. For further details, refer to the indicated references.

4.2.1 Vector Base Amplitude Panning (VBAP)

VBAP is an efficient extension of stereophonic amplitude panning techniques, applied to multi-loudspeaker setups. In a horizontal plane around the listener, a virtual sound source at a certain position is created by applying the tangent panning law between the closest pair of loudspeakers. This principle was also extended to project sound sources onto a three-dimensional sphere and assumes that the listener is located in the center of the equidistant speaker setup (Pulkki 2000). An extension to VBAP, entitled RVBAP (reverberated VBAP), was created by Olaf Matthes which aims to create a sense of distance and acoustic space by the usage of artificial reverb.

4.2.2 Distance Based Amplitude Panning (DBAP)

DBAP, conceived in 2003, also uses intensity panning applied to arbitrary loudspeaker configurations without assumptions as to the position of the listener. All loudspeakers radiate coherent signals, whereby the underlaying amplitude weighting is based on a distance-attenuation model between the position of the virtual sound source and each loudspeaker (Lossius et al. 2009).

4.2.3 Higher Order Ambisonics (HOA)

HOA extends Blumlein's pioneering idea of coincident recording techniques. HOA aims to physically synthesize a soundfield based on its expansion into spherical harmonics up to a specified order. To date, Max/MSP externals up to the 3rd order for horizontal-only or periphonic speaker arrays have been presented by Schacher and Kocher (2006) and Wakefield (2006).[1]

4.2.4 Space Unit Generator

Moore's Space Unit Generator, also called the room-within-the-room model, dates back to Moore (1983). Four loudspeakers represented as *open windows* are positioned around the listener and creates an *inner room*, which is embedded in an *outer room* with virtual sound sources. Sound propagation of the virtual source rendered at the *open windows* creates ICTDs and ICLDs. Some early reflections are calculated according to the size of the outer room. Pure Data and Max/MSP implementations were presented by Yadegari et al. (2002).

4.2.5 Spatialisateur

Spatialisateur, or its short name Spat∼, is in development at IRCAM and Espaces Nouveaux since 1991. It is a library of spatialization algorithms including VBAP, first-order Ambisonics and stereo techniques (XY, MS, ORTF) for up to 8 loudspeakers (Jot 1999). It

[1] Addition: A preliminary version of the Ambisonics externals by Schacher (2010) allows for up to 11th order Ambisonics encoding/decoding.

can also reproduce 3D sound for headphones (binaural) or 2/4 loudspeakers (transaural). A room model is included to create artificial reverberation controlled by a perception-based user interface.[2]

4.3 Virtual Microphone Control

ViMiC is a computer-generated virtual environment, where gains and delays between a virtual sound source and virtual microphones are calculated according to their distances, and the axis orientations of their microphone directivity patterns. Besides the direct sound component, a virtual microphone signal can also include early reflections and a late reverb tail, both dependent upon the sound absorbing and reflecting properties of the virtual surfaces.

4.3.1 ViMiC Principles

ViMiC is based on an array of virtual microphones with simulated directivity patterns placed in a virtual room.

Source - Microphone Relation

Sound sources and microphones can be placed and moved in 3D as desired. Figure 4.4(a) shows an example of one sound source recorded with three virtual microphones. A virtual microphone has five degrees of freedom: (X, Y, Z, yaw, pitch) and a sound source has currently four: (X, Y, Z, yaw). The propagation path between a sound source and each microphone is accordingly simulated. Time-of-arrival (ToA) and attenuation due to distance are estimated, depending on the speed of sound c and the distance d_i between a virtual

[2]Addition: After this paper was published, Spat∼ has been greatly overhauled, HOA and DBAP rendering algorithms were added and the binaural model was improved. Further, the software now supports more loudspeakers.

sound source and the i-th microphone. This attenuation function, (Equation 4.1) can be greatly modified by changing the exponent q. Thus, the effect of distance attenuation can be boosted or softened. The minimum distance to a microphone is limited to 1 meter in order to avoid high amplification gains.

$$g_i = \frac{1}{d_i^q} \qquad d \geq 1 \tag{4.1}$$

As an alternative to this inverse proportional decrease, a second distance function is implemented. Here the decrease in dB per unit can be controlled by the parameter k:

$$g_i = 10^{\frac{-k}{20} \cdot d_i} \tag{4.2}$$

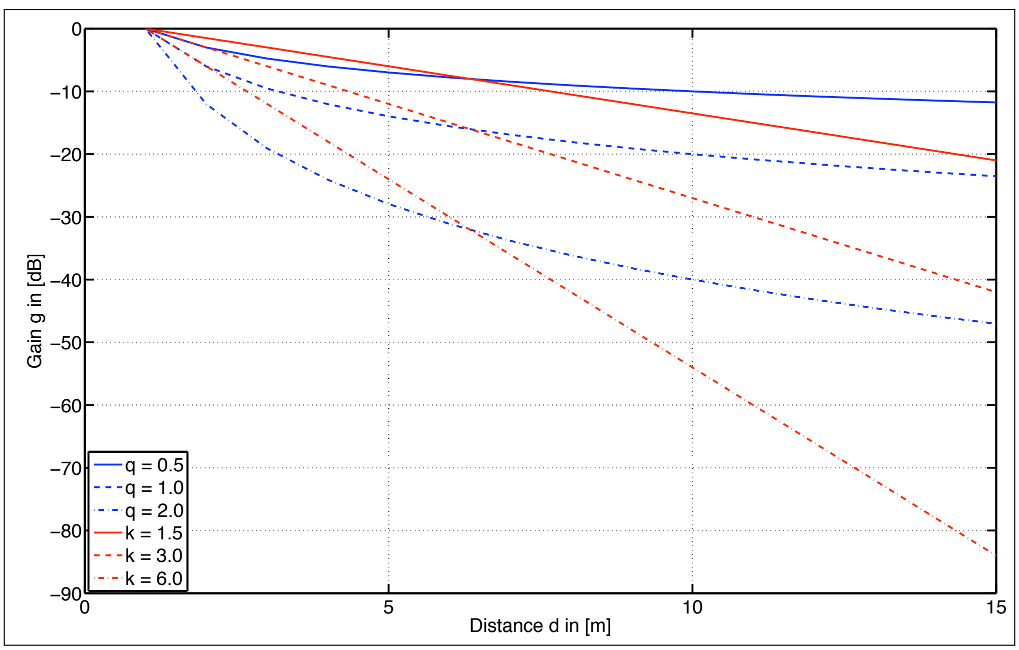


Figure 4.1: Comparison of the distance-attenuation functions with different settings: "inverse proportional decrease" model (blue lines, Equation 4.1) and "exponential decrease" model (red lines, Equation 4.2).

Further attenuation happens through the chosen microphone characteristic and source directivity (see Figure 4.3). For all common microphone characteristics, the directivity for a certain angle of incidence δ can be imitated by calculating the following equation:

$$\Gamma = (a + (1 - a) \cdot \cos\delta)^w \qquad 0 \leq a \leq 1 \tag{4.3}$$

The variable a is drawn from Table 4.1 to simulate different directivity characteristics. Increasing the exponent w to a value greater than 1 will produce an artificially sharper directivity pattern. Unlike actual microphone characteristics, which vary with frequency, microphones in ViMiC are designed to apply the concept of microphone directivity without simulating undesirable frequency dependencies.

Table 4.1: Coefficients for calculating microphone directivity in Eq. 4.3.

Characteristic	a	1-a	w
Omnidirectional	1	0	1
Subcardioid	0.7	0.3	1
Cardioid	0.5	0.5	1
Supercardioid	0.33	0.67	1
Hypercardioid	0.3	0.7	1
Figure-of-8	0	1	1

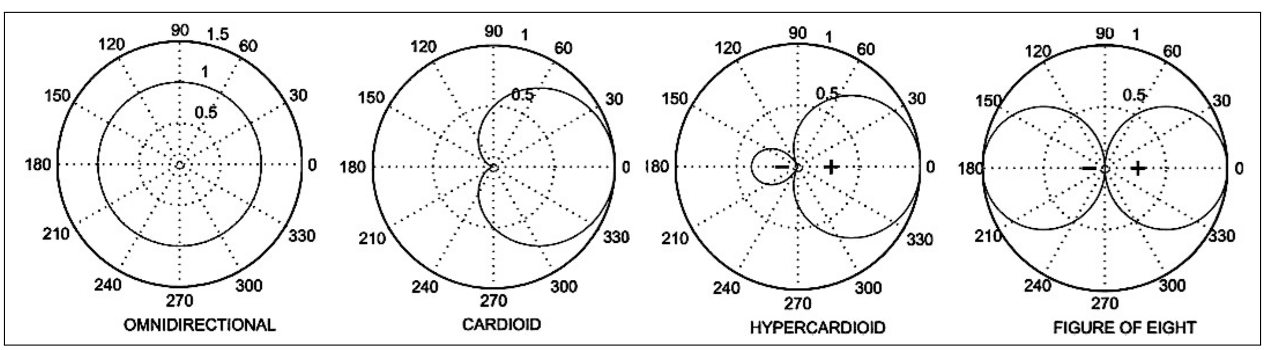


Figure 4.2: Common microphone characteristics, adopted from Pulkki and Hirvonen (2005).

Source-directivity is known to contribute to immersion and presence. Therefore ViMiC is also equipped with a source directivity model. For the sake of simplicity, in a graphical control window, the source directivity can be modeled through a frequency-independent gain factor for each radiation angle in steps of 1°.

Room Model

ViMiC contains a (rectangular) shoe-box room model to generate time-accurate early reflections that increase the illusion of this virtual space and envelopment as described in the literature (Allen and Berkley 1979; Pellegrini 2002). Early reflections are strong auditory cues in encoding the sound source distance. According to virtual room size and position of the microphones, adequate early reflections are rendered in 3D through the well-known image method by Allen and Berkley (1979). For a general discussion of the image method, see Cremer and Müller (1982) and Berman (1975). Each image source is rendered according to the time of arrival, the distance attenuation, microphone characteristic and source directivity, as described in Section 4.3.1. Virtual room dimensions (height, length, width) modified in real time alter the reflection pattern accordingly. The spectral influence of the wall properties are simulated through high-mid-low shelf-filters. Because larger propagation paths increase the audible effect of air absorption, early reflections in ViMiC are additionally filtered through a 2nd-order Butterworth lowpass filter with adjustable cut-off frequency.

Also, early reflections must be discretely rendered for each microphone, as propagation paths differ. For eight virtual microphones, 56 paths are rendered if the 1st-order reflections are considered (8 microphones · [6 early reflections + 1 direct sound path]). Although time delays are efficiently implemented through a shared multi-tap delay line, this processing can be computationally intensive.

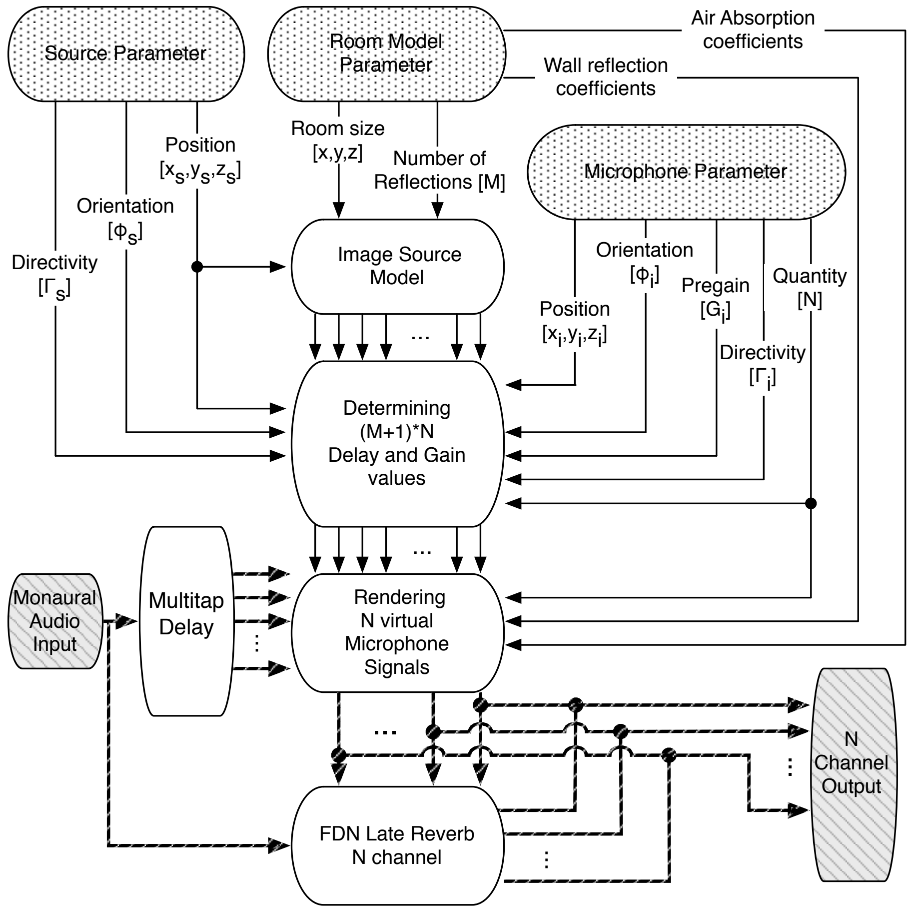


Figure 4.3: Flowchart of the Max/MSP processing.

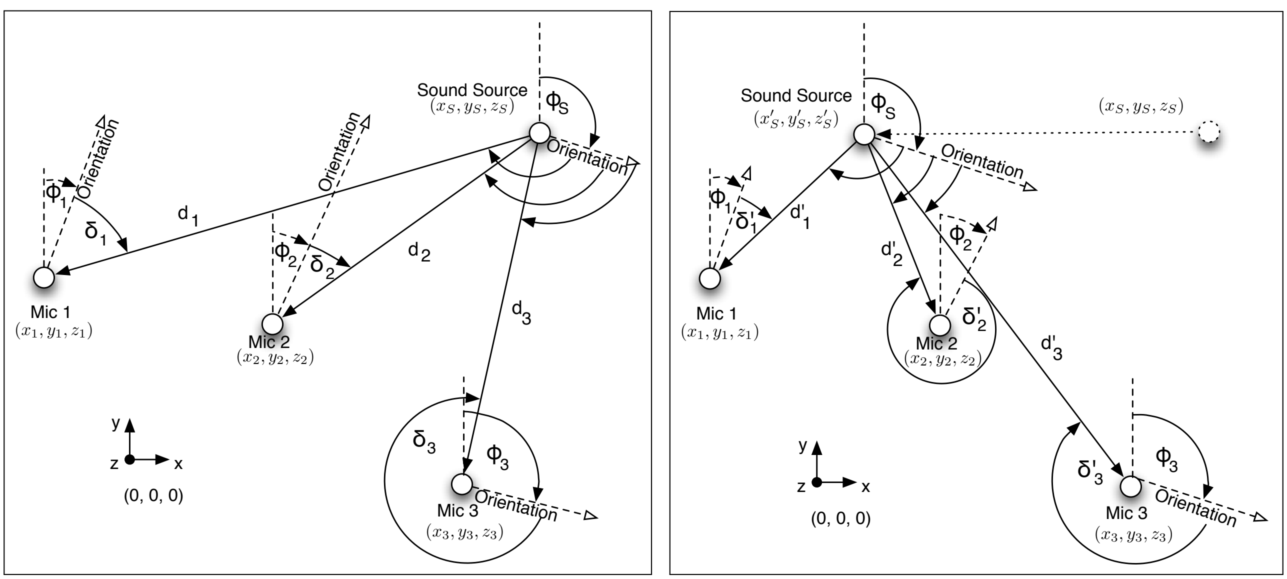
It has been shown that when increasing the level of individual early reflections in a simulated small room environment, the differences are first observed in sound coloration and then (after an increase of 2-4 dB) for spatial aspects, such as spatial impression (Bech 1998). In ViMiC, the levels of direct sound and early reflections can be modified independently to optimize the perceived sound field.

4.3.2 Late Reverb

The late reverberant field of a room is often considered to be nearly diffuse without directional information. Thus, an efficient late reverb model, based on a feedback delay network by Jot and Chaigne (1991) with 16 modulated delay lines diffused by a Hadamard mixing matrix, is used. By feeding the outputs of the room model into the late reverb a diffused reverb tail is synthesized, for which timbral and temporal character can be modified (see Figure 4.3). This late reverb can be efficiently shared across several rendered sound sources.

4.4 Moving Sources

In Figure 4.4(b), the sound source moved from (x, y, z) to (x', y', z'), changing the propagation paths to all microphones, and also, the time delay and attenuation. A continuous change in time delay engenders a pitch change (Doppler effect) that creates a very realistic impression of a moving sound source. The Doppler effect might not always be artistically desired, and consequently ViMiC offers solutions for both cases, spatialization with and without Doppler shift as further discussed below.



(a) Static source (b) Moving source

Figure 4.4: Geometric principles.

4.4.1 Rendering with Doppler effect

For each sound path, the changes in time delay are addressed through a cubic interpolated delay-line which leads to a significantly better perceived quality than an economical linear delay line interpolation. To save resources, interpolation is only applied when moving the virtual sound source, otherwise the time delay is being rounded to the next non-fractional delay value. At $fs = 44.1$ kHz and a speed of sound of $c = 344$ m/s, the roundoff error is approximately 4 mm. Some discrete reflections might not be perceptually important due to the applied distance law, microphone characteristics, and source directivity. To minimize the processor load, an amplitude threshold can be set to prevent the algorithm from rendering these reflections.

4.4.2 Rendering without Doppler effect

This rendering method works without interpolation: the time delays of the rendered sound paths remains static until one of the paths has been changed by more than a specified time delay. In this case, the sound paths of the old and new sound positions are cross-faded within a window of 50 ms, in order to avoid strongly audible phase modulations. Different cross-fade functions are offered to avoid audible cross-fade artifacts that might occur for certain sound material.

4.5 ViMiC in Relation to Other Rendering Techniques

As shown in Figure 1.1, several other mature spatialization techniques exist. Because of their relation to conventional microphone recording techniques, ViMiC can emulate those principles.

Ambisonics: By using one virtual omnidirectional and three virtual figure-of-eight microphones (perpendicular to each other), a First-Order Ambisonics B-Format can be synthesized with the W, X, Y and Z components and decoded with any appropriate Ambisonics decoder, e.g., those as part of the previously mentioned tools in Section 4.2.3. The synthesis of Higher Order Ambisonics through higher-order microphone directivity patterns is technically possible, but is not implemented yet.

Distance Based Amplitude Panning: ViMiC offers the possibility to render only the Inter-Channel Level Differences of the direct sound paths between sound source and omnidirectional microphone, thus enabling the imitation Distance Based Amplitude Panning (DBAP, Lossius et al. 2009), which is popular in the sound installation context.

Space Unit Generator: By placing virtual omnidirectional microphones at the position of the loudspeaker, and rendering first-order reflections in a virtual room, the Space Unit Generator, developed for CMusic by Moore (1983), can be imitated.

Wavefield Synthesis, WFS: Valente and Braasch (2006) demonstrated that with an appropriate number of closely spaced virtual microphones arranged according to the positions of a WFS-loudspeaker array, ViMiC can be used to apply the concepts of Wave Field Synthesis (Berkhout et al. 1993).

4.6 Acknowledgment

This work has been funded by the Canadian Natural Sciences and Engineering Research Council (NSERC) and the Centre for Interdisciplinary Research in Music Media and Technology (CIRMMT).

Chapter 5

Case Studies for Real-time Spatialization with ViMiC

The following Chapter was submitted as:

Peters N., Braasch, J., and McAdams, S. Learning from Users - Case Studies for Real-time Spatialization with ViMiC. Submitted.

Because the ViMiC principles were already explained in the Section 4.3, a similar explanatory part of this publication is omitted here.

Abstract

ViMiC (Virtual Microphone Control) is a real-time multi-channel spatial sound projection and rendering technique based on sound recording principles. The system is very intuitive for audio engineers, who typically think in terms of microphone arrangements, and was designed in order to be applied in a number of different music-related scenarios. This paper explains design criteria that have guided the development process and presents an overview of real-time applications where ViMiC is successfully used.

5.1 Structure

This paper is organized as follows. Section 5.2 explains details of the design approach that made the ViMiC System applicable for multiple real-world applications across disciplines. Section 5.3 shows how ViMiC is a versatile real-time application for composers, sound designers and Tonmeisters. Section 5.4 illustrates how ViMiC can also be used for education and in research (Section 5.5).

5.2 Design Approach

A previous study surveyed spatial sound synthesis techniques across media artists and composers (Peters et al. 2011). An effort has been made to apply these findings in the software design process to make ViMiC more applicable in real-world contexts.

For instance, one of the survey's questions asked about the importance of 10 technical features on a 5-point scale, ranging from "not important" to "extremely important". The three features with the highest importance ratings on average were "Spatial rendering in real time", "Controllability via graphical user interface" and "Controllability via external controllers" (see Figure 2.8 on Page 49). Further, for a subset of the participants the possibility to integrate the spatial sound renderer through plug-ins into Digital Audio Workstations (DAW) was rated as "extremely important". As will be shown, these four features in particular guided development.

5.2.1 Environments & Integratability

ViMiC was developed for two popular computer music software paradigms. First for real-time media programming environments Max/MSP (Peters et al. 2008); second as a multi-channel Audio Unit plug-in for DAWs.

For the media programming environment Max/MSP, the Jamoma platform (Place and Lossius 2006, http://www.jamoma.org) provides configurable, easy-to-use higher-level modules with a standardized graphical user interface (Figure 5.1). The Jamoma real-time environment currently comprises more than 120 high-level modules, ranging from controller and mapping modules, to specialized modules for video and audio effects, spatial sound rendering, or gesture and motion analysis (Jensenius 2007). The Jamoma distribution is freely available and includes ViMiC. Hence, ViMiC can be easily combined and applied in many different user scenarios. Customized stand-alone applications can also be created, thus liberating the application from any software dependency.

The survey indicated that DAWs are very popular environments for spatial audio productions. However, DAWs are often tailored to consumer media productions and therefore constrained in their multichannel capabilities (Peters et al. 2009). To make ViMiC accessible in DAWs, a ViMiC multichannel plug-in was developed that can be used with any DAW that supports Apple's Audio Unit plug-in format (Figure 5.7). For this implementation, the C++ DSP framework by Place et al. (2010a) was employed which also facilitates cross-compiling into other environments such as Ruby On Rails for web-based applications

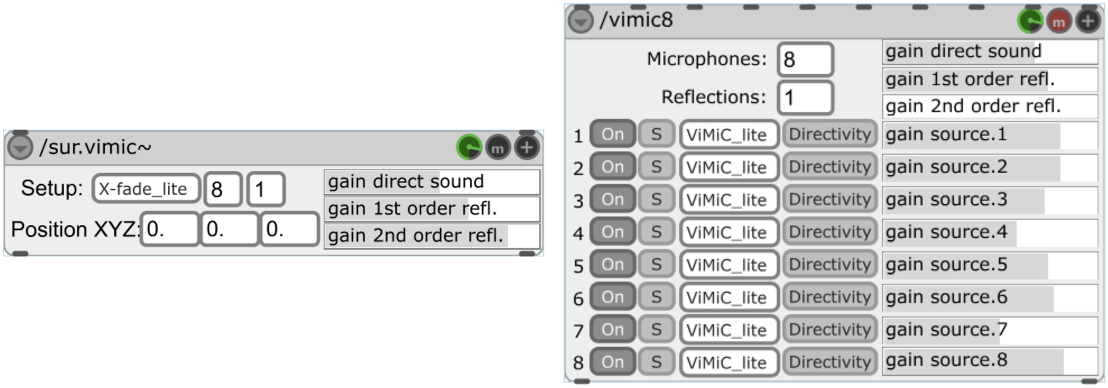


Figure 5.1: Jamoma Modules designed for ViMiC in Max/MSP. The left module renders one sound source and the right module eight sound sources in real time.

(Ruby et al. 2009) or into other DAW plug-in formats (e.g., VST, LADSPA).

5.2.2 Flexible Loudspeaker Settings

Today, many electroacoustic music festivals provide, as a quasi-standard, a loudspeaker array based on eight circular, horizontally-arranged full-range loudspeakers. For consumer media productions, the standard for surround sound is determined by DVD and Blu-ray Disc, with 5.1[1] and 7.1 discrete loudspeaker channels, respectively, according to the recommendation by the ITU (1992).

Despite those well-known loudspeaker settings, many performance and installation artists reported using non-standardized loudspeaker layouts differing in number, position and elevation of loudspeakers. ViMiC for Max/MSP is currently able to render 24 virtual microphone channels. If every microphone is routed to a discrete loudspeaker channel, 24 loudspeakers can be accommodated. Most DAWs usually only support standardized surround loudspeaker configurations to a maximum of eight loudspeaker channels sufficient enough for DVD or the emerging Blu-ray Disc. The ViMiC plug-in is capable of supporting all of these standardized configurations with a maximum of eight microphone channels. In

[1] The number behind the dot symbolizes the number of discrete subwoofer channels.

both environments, the virtual microphones can be placed with five degrees of freedom (position (x, y, z) and orientation (pitch, yaw)) at will. Therefore, there is a liberal relation between loudspeaker and virtual microphones and non-standardized loudspeaker settings can be accommodated.

5.2.3 Accessibility

Today's music software applications need to fit into existing environments. For instance, the developers of the spatial authoring software *MusicSpace* (Pachet and Delerue 1999), winner of the Bourges Music Software prize 2000, mention on their website[2] that although this stand-alone application received very positive feedback from artists such as Jean-Michel Jarre and Peter Gabriel, it failed to be used more widely because "MusicSpace was a closed system, not able to communicate easily with other music software".

To prevent such failure of communication in Jamoma, the Open Sound Control protocol (OSC, Wright and Freed 1997) is embedded for easy access and manipulates processes via external devices and interfaces in real time. Moreover, Jamoma contributes to the development efforts of OSC (Place et al. 2008) as acknowledged by Schmeder et al. (2010). ViMiC's OSC namespace reflects the three main categories (sound source, microphone and room) and is self-explanatory, avoiding cumbersome and potentially misleading abbreviations. It follows herewith the SpatDIF initiative in describing spatial sound information in a structured way to facilitate the exchange of spatial audio scenes (Peters et al. 2007; Peters 2008).

Many parameters are defined with unit information to allow flexible manipulation of parameters in different measures. For instance, the gain controllers by default are defined within the MIDI range (0.0 - 127.0), but other units can be declared via OSC messages;

[2] http://www.csl.sony.fr/~pachet/MusicSpace/, accessed July 2010

e.g., the OSC message `/room/reflection/gain.1 -3.0 dB` defines the gain value in decibels, whereas `/room/reflection/gain.1 0.708 linear` would create the same gain value, but defined as a linear amplitude value.

OSC namespace example

```
/source.2/position −1.5 4.0 0.0
/room/reflection/gain.1 95.7
/room/reflection/airfilter 6000
/microphone.5/directivity/preset supercardioid
```

5.2.4 Real-time rendering

ViMiC was deliberately implemented in the time-domain instead of in the frequency domain to circumvent latency due to FFT/IFFT block processes. This allows for high quality dynamic control of all parameters in real time. For instance, the virtual room dimensions can be modified, creating perceptual effects hardly possible in the real world, virtually transporting the listener from a reverberant cathedral into a small garage-band rehearsal room. Further, according to artistic aspiration or the limitation of the DSP resources, the ViMiC system also offers several levels of rendering quality that can be changed seamlessly at run time (see the User Manual for details).

5.2.5 Perceptual Parameters

The survey (Peters et al. 2011) also inquired into the perceptual aspects that composers and artists are striving to create. From 15 given spatial percepts, the three most highly rated aspects were “Immersiveness”, “Distance perception of sound sources” and “Localization accuracy of sound sources” (see Figure 2.9(a) on Page 53).

ViMiC's approach to create these three perceptual aspects is to simulate the sound propagation in a room as defined through the direct sound, early reflections and late reverb segments as previously described (Peters et al. 2008; Braasch et al. 2008).

Table 5.1 adapted from Theile (2001) shows the contribution of these segments to the perception of envelopment (as related to immersiveness), direction, distance and depth impression. Because early-reflections were found to be specifically important for distance, depth and spatial impression, ViMiC renders early reflections according to Allen and Berkley (1979) for each individual virtual microphone time-accurately.

Table 5.1: Contribution of Direct sound, Early reflections and Late reverb to various perceptual aspects. The number of stars represents the importance of the acoustical component for creating the percept. Table is adapted from Theile (2001).

Percept	Direct Sound	Early Reflections	Late Reverb Tail
Direction	⋆⋆	⋆	
Distance, Spatial Depth		⋆⋆	⋆
Spatial Impression		⋆⋆	⋆⋆
Envelopment		⋆⋆	
Sound Color	⋆⋆	⋆	⋆⋆

5.3 Musical Applications

Due to its flexibility in the number and placement of virtual microphones and sound sources, ViMiC can be used for non-standardized loudspeaker setups and non-centralized audiences as well as standardized contexts, thus making it applicable for a variety of sound and media applications. The following are examples from several applications, including sound and media installations, concert scenarios, telepresence, studio production, motion picture, and digital preservation of music technology.

5.3.1 Sound and Media Installations

Ricardo del Pozo - adaptation/volume

Ricardo del Pozo created the sound installation *adaptation/volume* in partial fulfillment of the requirements for a Master in Fine Arts at the National Academy of the Arts, Bergen, Norway. The 16-channel sound installation is made with the real-time media programming software Max/MSP (Zicarelli 1998) and Jamoma (Place and Lossius 2006). For the installation, pre-recorded sounds were manipulated through different real-time audio effects and spatialized for the 16 loudspeakers via ViMiC, with a Mac Mini computer.

The artist says that "the work deals with the aspect of acousmatics, space and the idea of the organized sound, structure and composition as a spatial and sculptural form, how organized sound achieves a body, a form, through spatialization techniques. It is a study into how virtual space is overlapping the physical, auditory perception of space and visual perception of space, one superimposed on the other." He further noticed that through the use of pre-recorded sound material processed in real time, the form of the sound material changes over time, while the source material remains accessible; thus time expands beyond the linear approach.

Pozo used an unorthodox loudspeaker arrangement for his work. While loudspeakers usually surround the audience, he arranged the 16 loudspeakers on a small circle, facing outwards, interacting with the gallery space (see Figure 5.2). The ViMiC algorithm was accordingly set up for 16 virtual cardioid microphones on a circle, pointing inwards to the circle's center point. The early reflection pattern, reverberation time and other properties of the virtual room, as well as the virtual sound source positions changed very slowly, provoking the audience' acoustic awareness. Pozo also said that a key factor in choosing ViMiC for this installation was that the positioning of the speakers was not predefined, that ViMiC can be quickly adapted to any loudspeaker setting. “This means a lot to me since it gives me great flexibility to use this system differently if the space I am working in is architecturally and acoustically challenging. I have several audio-visual works which I’m developing [...] with ViMiC. For me, there is no reason to use something else since it is in my opinion the best solution for working site specific with multichannel audio installations,” (Pozo 2010, personal email communication with the first author).

Prior to using ViMiC, the artist also experimented with ambisonics and other rendering approaches to create an immersive environment. “I found that I never really felt that the perception of depth was audible. Direction of the sound was convincing, but distance was not so noticeable, or perceptually convincing. When I first tried ViMiC I was struck by how believable it was. It really felt like the sound came from *behind* the speakers and not coming *from* the speakers. Also this perception of depth and distance was really remarkable in terms of what I was seeking in my work.”

Figure 5.2: Impressions of Ricardo del Pozo’s sound installation *adaptation/volume*, pictures provided by the artist.

The Wooster Group - There is still time . . . brother

In this interactive multimedia piece, the ViMiC system was used in conjunction with a 360° cylindrical digital video projection screen (McGinity et al. 2007). The work was commissioned by Rensselaer Polytechnic Institute's Experimental Media and Performing Arts Center (EMPAC) and has been presented at the Zentrum für Kunst und Medientechnologie (ZKM) in Karlsruhe, Germany, and other venues.

The audience is surrounded by the screen with one audience member taking place on a special rotating chair in the central position. Although the projected movie was created as a full 360° movie, it is not projected in its full frame width. Instead, only the video segment to which the person in the rotation chair is oriented is fully visible, whereas the other segments are darkened and blurred out, simulating the effect of peripheral field of vision. This "windowed" video segment follows the movement of the rotating chair. Because the timeline of the video is preserved, the audience can explore the entire movie content by viewing it several times from different angles.[3]

Similarly to the video projection, the accompanying sound field also dynamically adapts in real time to the movement of the rotating chair. ViMiC was used in this scenario to create spatial effects such as Depth illusion and Doppler shift and to simulate various classical microphone techniques. Surreal scenes could be created by assigning artificial directivity patterns to microphones or changing the laws of physics in the model. While the unprocessed audio content and pre-arranged control data were organized on a DAW and streamed to a dedicated ViMiC audio rendering computer, various ViMiC parameters were adjusted in real time, including the directivity patterns and orientations of both the microphones and sound sources, as well as their precise locations. For communication between the rotation

[3]A video can be seen at http://icinema.unsw.edu.au/projects/prj_wooster.html, accessed Jun 2010.

chair, video and audio components, the OSC protocol was used (Figure 5.3). In total, 54 sound sources were spatialized with ViMiC through 24 loudspeakers. To fully immerse the audience in the installation, these loudspeakers were configured as three loudspeaker rings at different elevations behind the acoustically transparent 360° video projection screen, spatializing the sounds, early reflections and late reverb in 3D.

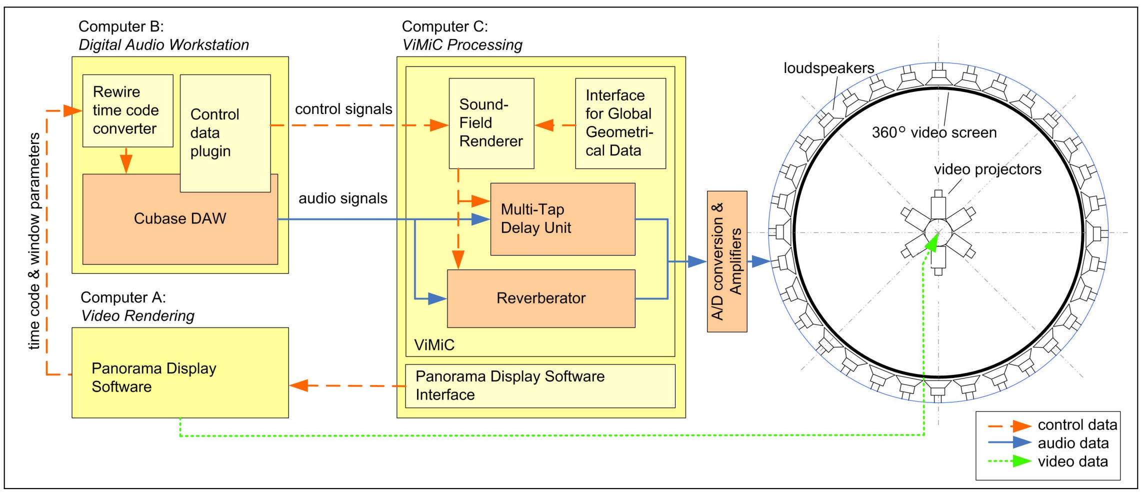


Figure 5.3: System architecture and data routing for Wooster Group's installation piece.

5.3.2 Concert Scenarios

Sean Ferguson - Ex Asperis

Composed for solo cello, gesture-controlled spatialization, live electronics and chamber orchestra, *Ex Asperis* had its world premier in Pollack Hall at the 2008 MusiMars Festival in Montreal and was performed by Chloé Dominguez (Solo Cello), Fernando Rocha (data gloves for manipulating sound spatialization) and the McGill Contemporary Music Ensemble (CME), directed by Denys Bouliane.

Pollack Hall is a traditional shoebox-style concert venue with raked seatings for up to 600 listeners. Altogether, 24 loudspeakers were arranged in two rings at different heights surrounding the audience. The system was further enhanced by four subwoofers.

The technical setup for this performance was rather complex, including bidirectional audio and a TCP/UDP network between stage and front of house (FoH), on-stage motion-sensors, data-gloves, as well as several decentralized computers used for real-time spatial-sound rendering, sample playback, and processing of the motion data. Further, audio feed support for CBC Radio (Canadian Broadcasting Corporation) was provided.

The composer's idea was to use the ViMiC spatialization system to expand the physical limits of the stage width and virtually stretch it during the performance completely around the audience, as controlled by the performers' gestures. To this end, a sensor system was attached to the right arm of the solo-cellist to measure the activity of bowing motions; another performer was equipped with data gloves and binaural headphones to act as a "spatial orchestrator" who arranged and manipulated the spatialized virtual sound sources around the audience. The sound of the solo cello was captured live (Figure 5.4(a)) to render virtual early reflections through ViMiC. These reflections, played back over the loudspeakers, enhanced the natural direct sound of the instrument in a subtle, yet perceivable way.

A few spatial sound layers were pre-rendered by the composer within a DAW using the built-in panning features for a 7.0 loudspeaker setup. In the concert hall, ViMiC was used to up-mix these 7-channel pre-rendered audio materials to the specific 24 loudspeaker configuration: according to the placement of the 7 loudspeakers in the studio, 7 virtual sound sources were arranged in ViMiC, behaving as virtual loudspeakers. Further, 24 virtual microphones were positioned, according to the placement of the loudspeakers in Pollack Hall. By feeding the pre-rendered audio material into ViMiC, the audio was reproduced at the positions of the virtual loudspeakers and "re-recorded" via the virtual microphones.

The creation and performance of this piece was funded by the Canadian Natural Sciences and Engineering Research Council (NSERC) and the Canada Council for the Arts (CCA).

(a) C. Dominguez, equipped with motion sensors in a rehearsal. In front, a feedback-protected microphone for sound capturing for real-time sound spatialization. Courtesy of M. Marshall.

(b) View from the Front of House (FoH) to the stage, ViMiC sound processing Max/MSP patch on the middle computer screen. Courtesy of R. McKenzie.

Figure 5.4: Rehearsal scenes from Sean Ferguson's *Ex Asperis*.

Marlon Schumacher - De Vive Voix II

For Montreal's MusiMars festival 2010, Marlon Schumacher performed his composition *De Vive Voix II* for voice, data glove and live electronics, including spatialization of real-time processed sounds, in the fairly reverberant Redpath Hall at McGill University (see Figure 5.5(a)). The arrangement of stage and audience seating area and the limited room size complicated the standard placement of equidistant surrounding loudspeakers. ViMiC was used in this scenario to develop a creative spatialization concept with eight loudspeakers, arranged as illustrated in Figure 5.5(b). The loudspeakers were positioned to provide three separate spatialization and amplification zones, and an overall layer of spatialization, each of which were separately connected to a different ViMiC system. The three zones consists of the left wing, the right wing, and the central area (Figure 5.5(b)). Different spatial sound layers were projected with ViMiC to each of the loudspeaker zones, enabling the composer to create “sound clouds”, rich in spatial and spectral detail. An overall sound layer used all eight loudspeakers and produced a more “global” sound spatialization, perceivable across the audience. Therefore, rather than using the concept of a single ideal listening point (sweet spot) and many low-fidelity listening positions, *De Vive Voix II* created different listening areas, presenting distinct perspectives on the musical material, while providing an overall shared music experience.

(a) Singer Juliane Klein (left) and Marlon Schumacher (with computer and data glove).

(b) Stage, audience, and loudspeaker configuration.

Figure 5.5: De Vive Voix II.

5.3.3 Telepresence Concerts

Telepresence, or live networked music performances have become popular over the last few years. Increased availability of fast and reliable broadband internet connections and other advances in computer technologies suggests this trend will continue. In these concerts, musicians perform together over the internet, while being physically located at two or more remote sites. In this context, the ViMiC system can be used to create a common auditory virtual space in which all musicians perform and interact with each other.

Since 2006, the ViMiC system is used for such tasks in the telepresence music improvisations between Pauline Oliveros' Tintinnabulate Ensemble (Rensselaer Polytechnic Institute, Troy) and Chris Chafe's Soundwire Ensemble (Stanford University) as a component of the Telematic Music System (Braasch 2009). The Telematic Music System also includes the Expanded Instrument System (Gamper and Oliveros 1998), JackTrip audio streaming software by Caceres and Chafe (2009) and the Ultravideo Conferencing system (Cooperstock et al. 2004).

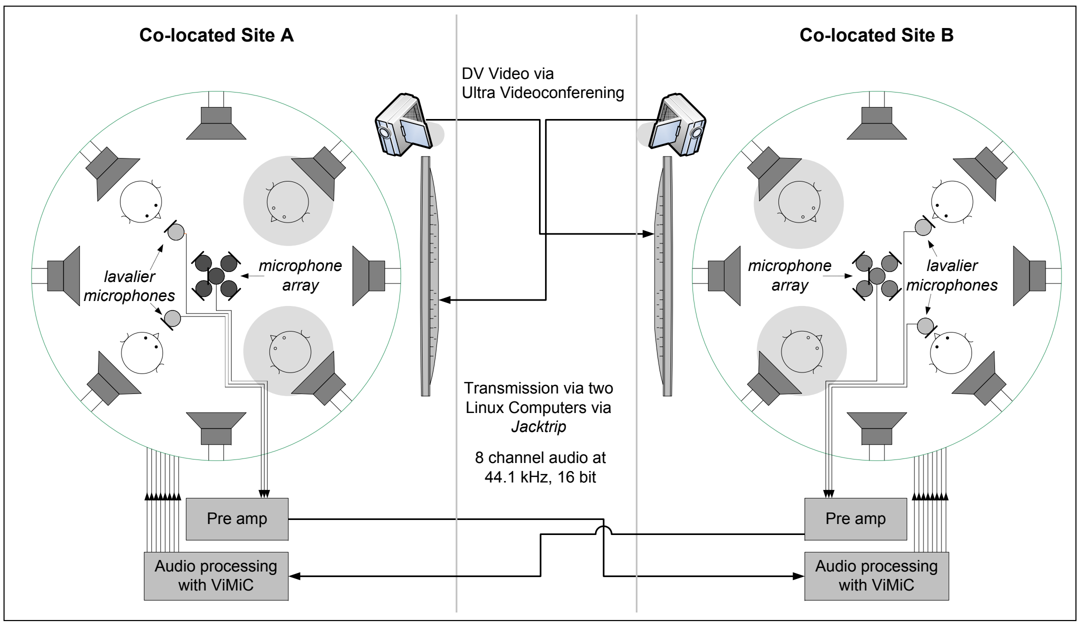


Figure 5.6: Sketch of the internet-based Telematic Music System used by Braasch (2009).

In this system (Figure 5.6), the sound of the musicians is captured using near-field microphones and a microphone array to localize them. The near-field microphone signals are transmitted via JackTrip and spatially recreated at the remote ends using ViMiC and a loudspeaker array. To simulate the same virtual room at all co-located sites, the ViMiC systems communicate using the OSC protocol to exchange room parameters and the room coordinates of the musicians. Using OSC, they also receive localization data from the microphone arrays. A bidirectional video stream (Ultravideo Conferencing) allows visual interaction of the musicians.

Many telepresence concerts by the RPI/CCRMA ensembles demonstrated the validity and reliability of this ViMiC use case. The first commercial album using ViMiC in a telepresence scenario is a 5-channel Quicktime video of a live-recording by Tintinnabulate & Soundwire et al. (2009). For the ICAD 2007 conference, they performed *Tele-Colonization* together at the co-located sites McGill University (Montreal, Canada), Rensselaer Polytechnic Institute (Troy, NY, US), Stanford University (Stanford, CA, US), and KAIST (Seoul, South Korea) (Stallmann 2007).

For a performance of *Dynamic Spaces* at SIGGRAPH 2007 in San Diego, ViMiC was used to create a dynamically changing acoustical space. In this performance, the room acoustics were remotely altered using a real-time controller. Reverberation time, room size, sound pressure level of early reflections, and frequency response were among the parameters they manipulated.

5.3.4 Studio Production

ViMiC is also used for studio production in commercial audio media formats; for example a two-channel auralization can be heard on the CD *Global Reflections* by Jonas Braasch (2006) and on a 5.1 DVD of Marlon Schumacher's composition *De Vive Voix II*.

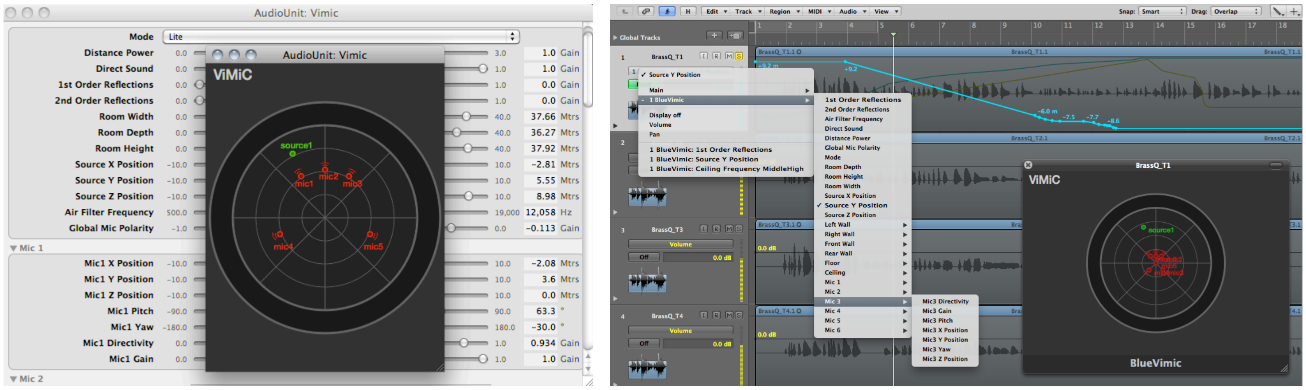
To arrange the virtual microphones for commercial media formats (e.g., stereo, ITU 5.1), Tonmeisters developed different microphone setups (e.g., Williams and Le Dû 2004; Rumsey 2001) easily applicable in ViMiC.

Because audio productions are most likely carried out on a DAW, the ViMiC Audio Unit plug-in can be used. In a DAW, the ViMiC plug-in is added to the desired (to be spatialized) audio tracks, and by manipulating the plug-in's GUI, the positions of microphones and sound sources are defined. Microphone settings can be synced across different ViMiC instances and all parameters can be dynamically controlled via the DAW's automatization features in real time (Figure 5.7(b)). The ViMiC plug-in does not render the late reverb, so that different late reverb plug-ins can be applied according to quality and CPU resources and can optionally be shared across ViMiC instances.

This flexibility uniquely positions ViMiC in the context of radio play productions, where dynamic changes of listening perspective and room environment are prominent dramaturgical principles.

One specific and hypothetical scenario within the audio production context should be highlighted here. Imagine a sound engineer who has completed a multichannel (multi-microphone) recording of an orchestra. Then the audio producer decides to add an extra sound layer on top of this recording. Usually, a simple amplitude panning would be used to position and distribute the sounds of the extra layer on top of the recording. Because of the missing Inter-Channel Time Differences in the added sound layer, the mix with the previously recorded orchestra may sound flawed. By arranging ViMiC's virtual microphones in a way similar to that of the real recording and by simulating the recording room parameter in ViMiC's virtual room, the extra sound layer can be spatially matched to the recorded material, thus creating a more homogenous spatial sound impression than could have been achieved by simple amplitude panning. Horbach et al. (2000) presented a related auralization approach for static sound sources which is now implemented as *Virtual Surround Panning* in Studer mixing consoles.

Also in the context of mixed music productions (music for acoustic instruments and electronics), ViMiC can help to blend electronic sounds with the sounds of the (live recorded) acoustical instruments, an effect that is often desired.



(a) ViMiC Audio Unit plug-in with custom (front) and generic user interface (background).

(b) ViMiC AU used in Apple Logic, automatization of parameters.

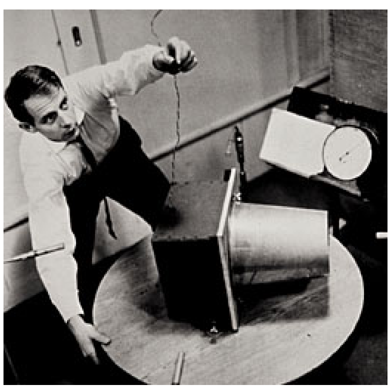
Figure 5.7: ViMiC as an Audio Unit plug-in.

5.3.5 Digital Preservation of Stockhausen's Rotation Table

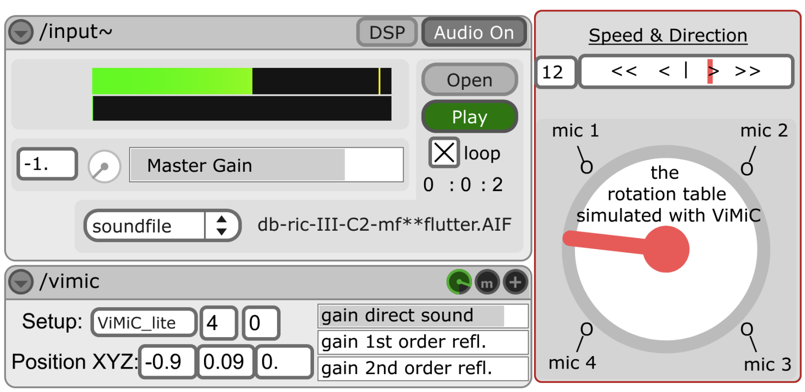
The preservation of electroacoustic music is becoming an important topic among composers, musicians, musicologists and researchers in Information Studies and Music Technology (Chadabe 2001). The fast development and generation changes in media formats and music technology complicate the issue, making technology too often obsolete and inaccessible even before its musical potential can be fully explored and appreciated. Several efforts have been made to digitally recreate old technology for analysis and experiential purposes (e.g., Pestova et al. 2008; Clarke and Manning 2009).

One of these technologies might be the *Rotation Table*, developed in 1958 by Karlheinz Stockhausen for his piece *Kontakte* and refined for *Sirius* (1975–77), which mounts a directed loudspeaker surrounded by four stationary microphones that receive the loudspeaker signal (Figure 5.8(a)). The recorded microphone signals were played back and routed to different loudspeakers arranged around the audience. Due to the directivity and separation of the microphones, the recorded audio signals contained Inter-Channel Time Differences (ICTDs) and Inter-Channel Level Differences (ICLDs). The speaker could be manually rotated up to about 7 Hz and depending on its velocity, the change in ICTDs can create audible phasing and Doppler effects. Stockhausen mentioned that for high rotation frequencies, the sound "starts dancing completely irregularly in the room – at the left, in front, it's everywhere"–even changing pitch depending on where the listener is standing. He also pointed out that these effects are caused by the natural effects of phase-shifting, which alternately stretches and compresses the sound - an effect that cannot be reproduced by simple digital amplitude panning (Maconie 2005).

ViMiC was used to emulate the loudspeaker-microphone configuration of the rotation table, producing the ICLDs and ICTDs that create the sound quality as reported by Stockhausen. Figure 5.8(b) shows the application where pre-recorded or live sounds can be spatialized in real time. The table's rotation speed and direction can be manipulated either via the GUI or remotely with an external controller. Consequently, the rotation table technology is digitally preserved and is reusable with a variable number of virtual microphones for today's musical applications.



(a) Stockhausen on his Rotation Table.



(b) The digital emulation using ViMiC.

Figure 5.8: Preservation of Stockhausen's Rotation Table using ViMiC.

5.3.6 Motion Picture

Spatial sound is a popular creative element in Cinema and Motion Picture and the interaction between picture and audio can create a unique experience for the audience. In combination with (motion) pictures, psychological studies have even shown that sound is better in generating entertainment pleasure and emotions than the visual inputs (Christensen and Lund 1999).

Molecules to the MAX!

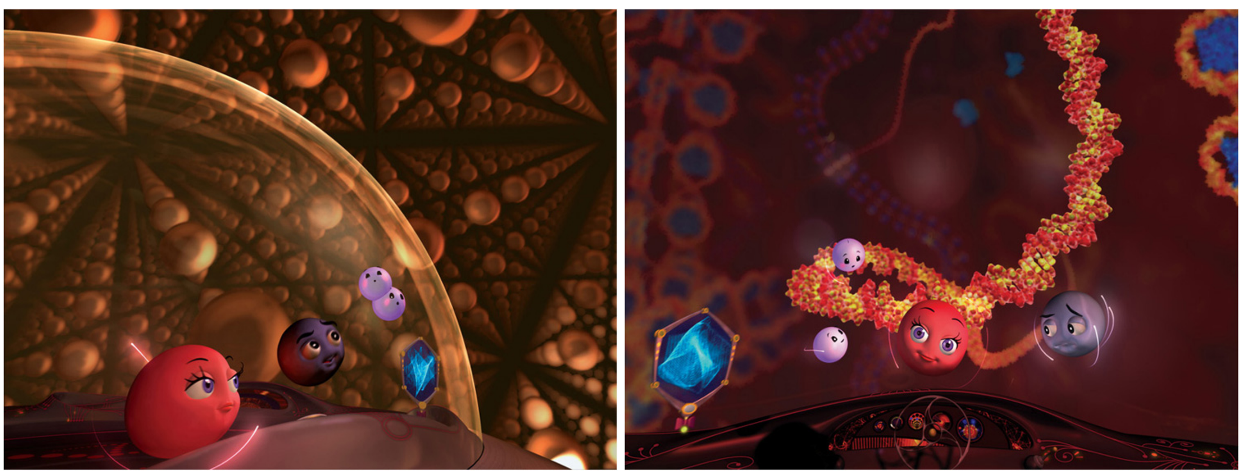
Premiered in 2009, the 3D-animation movie for IMAX® "Molecules to the MAX!", is an educational family adventure where molecules are the main characters, travelling with a spaceship through the universe. On their journey, Oxy and her crew explore on a microscopic level the molecular world of snowflakes, raindrops, a living cell and other objects.

The sound designer Jesse Stiles used an early version of the ViMiC plug-in made with Cycling'74's Pluggo environment to spatialize sounds using a DAW. From the visual rendering software Maya, used to create the animations, he received a real-time OSC data stream containing the spatial locations of the animated characters and sounding objects in Maya's virtual world. These position coordinates could then be used to spatialize sound effects and dialog with the ViMiC plug-in for the six-channel IMAX surround format according to the picture. Further, by properly placing the virtual microphones in ViMiC, the typical mismatch between the screen width and the width spanned by the frontal loudspeakers was compensated.

The virtual microphones were also oriented according to the camera perspective. Whenever the camera perspective changed, the virtual microphones automatically displaced and oriented themselves according to this real-time synchronization with Maya. For instance a 360° panning shot makes the virtual microphones rotate simultaneously with the camera.

With animation and sound design always in synchrony, the time-consuming need to create sound trajectories manually according to the picture's perspective was eliminated.

The multichannel audio tracks were sent to the Technicolor studio in Toronto for the final sound mix. Technicolor's sound engineers reported that with all their experience working on lots of major motion pictures, ViMiC's spatialization approach is a novel concept in the context of large-format films. After playing the ViMiC-spatialized scene in an IMAX theatre, in an email communication with J. Stiles they stated that "the spatialization is interesting and seems to work well with the image".



(a) Exploring a copper structure (b) Within a living cell

Figure 5.9: Scenes from the IMAX movie *Molecules to the MAX!*, courtesy of http://www.moleculestothemax.com, all rights reserved.

5.4 Education

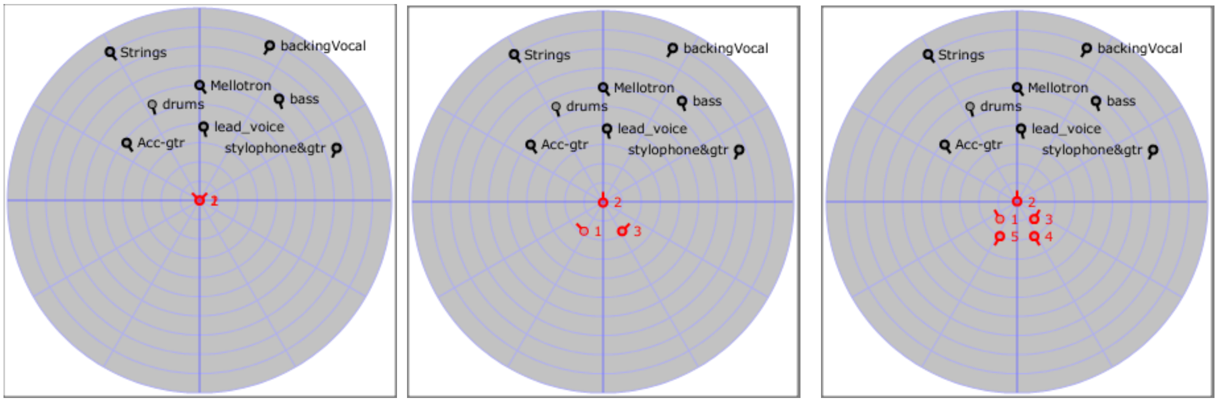
5.4.1 Tonmeister Training

Tonmeister students usually undergo technical ear training courses to sharpen perception and understanding of sound quality to improve their recording and production skills.

Timbral aspects and reproduction artifacts (e.g., bit-errors, amplitude/phase response differences between channels) are often prioritized in training, with less emphasis on spatial sound attributes (Neher 2004). ViMiC has the potential to be the missing educational tool for training recording engineers and Tonmeisters with regards to microphone settings. With ViMiC, students can create virtual recording scenarios and quickly experience the subtle differences between microphone directivities and various microphone techniques in terms of Inter-Channel Time Differences (ICTD) and Inter-Channel Level Difference (ICLD) and the effect of source directivity patterns and early room reflections. Because ViMiC is designed as a real-time application with a limited amount of processing power, the software is restricted to ideal microphones with frequency-independent directivity patterns, and does not simulate frequency-dependent directivity characteristics of specific microphone brands and models. However, with faster computer systems, this feature could be implemented to make ViMiC even more suitable for this application. To facilitate this application, ViMiC is equipped with a preset database to simulate popular stereophonic and multi-channel microphone settings, such as ORTF, Blumlein, AB, XY, OCT, or Decca Tree. An overview of multichannel microphone techniques can be found in Rumsey (2001).

5.4.2 Educational Events

For several educational events organized by the Centre for Interdisciplinary Research in Music Media and Technology (CIRMMT), ViMiC was used to educate and entertain children at science fairs and school events with an interactive spatial sound installation of up to 22 loudspeakers. Using Jazzmutant's Lemur multitouch interface (Figure 5.11(b)) the spatial location of eight sound sources in different sound scenes could be manipulated by the visitors. For instance, sound scenes included an orchestra, divided into eight instrumental sections, playing a Beethoven symphony. Children then could listen to these sections and



(a) Using a stereo XY-microphone setup (b) Using a Decca Tree setting (c) Using a Fukada Tree arrangement

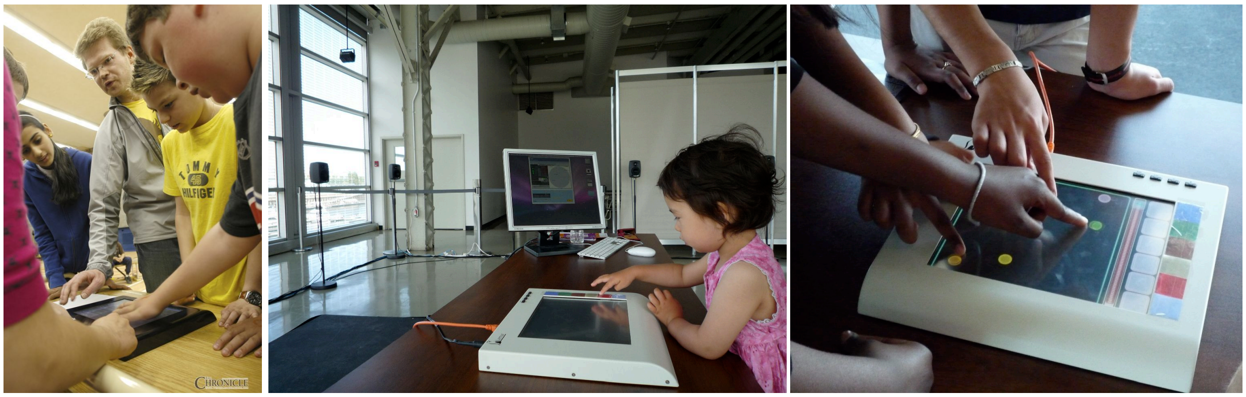
Figure 5.10: Top view of a virtual recording scene, using different microphone presets. Black dots with text label: sound sources and their frontal direction. Red dots with numbers: microphones. The "nose" on the circles illustrates the orientation of the sound sources and microphones.

explore the character of the individual instruments, before arranging them in the different formations in the virtual concert hall using the touch screen, perceiving the differences of orchestra settings (Meyer 2009). Different sound scenes were created by composer and sound designer Eliot Britton to account for different age groups, including a farm, rain forest, or a rock concert. Another user example is shown from the Eureka! science festival 2008 in Figure 5.11(b), where a child discovers and virtually arranges the soundscape of a farm. The entire installation was fully controllable through the Lemur multitouch interface connected via OSC to a Max/MSP patch running ViMiC.

5.5 Research Projects Using ViMiC

5.5.1 Medical Sector

ViMiC is currently used by Valente et al. (2010) at the Boys Town National Research Hospital in Omaha, NB, USA to assess children's speech intelligibility in the presence of noise and reverberation by using a virtual classroom paradigm. For the experiments,



(a) In an elementary school.

(b) At the Eureka! Science Festival, loudspeakers and the computer system running a ViMiC patch (left). Lemur multitouch interface (right) .

Figure 5.11: ViMiC in the museum installation context. (a) courtesy of Jacques Pharand.

audio/visual stimuli of children reading classroom lessons were created and processed with the ViMiC system to generate controlled room models. These models vary in terms of reverberation times (T_{30}) and definition ratings (C_{50}) by changing the energy ratio of early reflections to late reverberation in ViMiC. HVAC-simulated background noise was added at different sound pressure levels.

5.5.2 Sound Recording Research

With the advent of new surround reproduction standards (e.g., 7.1 ITU 1992, 10.1 Holman 2002, 22.2 Hamasaki et al. 2004), traditional five-channel recording techniques used for the 5.1 standard are insufficient. Consequently, adequate multichannel microphone techniques have to be developed and evaluated. ViMiC can help in this design process. With ViMiC, new microphone configurations can be virtually created and tested before time- and money-consuming recordings in real-world scenarios are created.

Related to the use case described in Section 5.3.3, Braasch et al. (2009) proposed a mixing console specifically design for telematic music. The software-based prototype mixer, which includes (beside traditional console elements) a number of use-case specific features, also deploys ViMiC for spatial sound rendering.

5.6 Conclusion

This paper has shown how the findings of a survey on the compositional use of spatialization helped to refine the development process of the ViMiC system which led to a variety of unexpected applications across fields. As documented through several of these applications, ranging from music reproduction and motion picture systems to medical research, ViMiC is evidently a flexible tool for spatial sound rendering in real time. The case studies exemplify the interest in and need for a spatial rendering approach with the exact features ViMiC was designed for.

We hope that this report encourage other people to use ViMiC and to compare it with other spatial rendering approaches. ViMiC is freely available through the authors.

5.7 Acknowledgment

This report was funded by a grant from the Canadian Natural Sciences and Engineering Research Council and the Canada Council for the Arts (NSERC, CCA) to Stephen McAdams, Jonas Braasch, Marcelo Wanderley and Sean Ferguson. The different presented projects may have received other sources of funding. We would like to thank all ViMiC users who provided feedback for this review.

Part III

Listener Perceptions of the Sweet Spot

Chapter 6

Auditory Perception at Off-center Listening Positions in Surround Sound Environments

The following Chapter was submitted as:

Peters N., Braasch, J., and McAdams, S.: Auditory Perception at Off-center Listening Positions in Surround Sound Environments, *J. Acoust. Soc. Am.*, (under revision).

Abstract

Assessments of listener preferences for different multichannel recording techniques typically focus on the sweet spot, the spatial area where the listener maintains optimal perception of the reproduced soundfield. The purpose of the present study is to determine whether multichannel microphone configurations affect the sound quality at off-center (non-sweet spot) listening positions in medium-sized rooms for larger audiences. In two different rooms, listening impressions of two musical excerpts created by three different multichannel recording techniques for several off-center positions are compared with the impression at the sweet spot. It was found that spaced microphone techniques create less sound degradation for off-center listening than the coincident Ambisonics technique. In a non-ideal listening environment (non-ideal loudspeaker configuration and listening room acoustics), the degree of sound degradation was different between the musical excerpt and none of the tested recording techniques was superior for both excerpts. Relevant spectral, spatial and energetic sound features, extracted from binaurally captured sound fields, are used to predict the mean behavioral data. These prediction models suggest that the perception of sound degradation is first due to spectral and spatial alterations in the higher frequency range at and above 2.5 kHz and that the listening environment and properties of the audio material both affect the strength of the off-center sound degradation.

6.1 Introduction

A concert hall is designed to enhance natural sound sources and produce a plurality of listening positions with perceptually good sound images of those sound sources (Ando 1998). In spatial audio reproduction, however, a best listening point is usually implied, thus making surround-sound reproduction difficult for larger audiences. Although several types of microphone techniques exist to record music performances in surround-sound (Section 6.2.2), and all techniques aim to give listeners the impression of *being there*, they tend to favor the centralized listener and yield a degraded sound image for the others (Section 6.1.1).

Perceptual artifacts that degrade the sound image may be due to the way binaural processes form a plausible auditory scene from the environment (Blauert 1997). These artifacts are under-investigated and little understood (Section 6.1.4). Bech and Zacharov (2006, 259) suggest that this understanding is critical because "suboptimal listening locations can provide significant information regarding the general performance of the [reproduction] system" and that "off-centre locations may well be more representative of typical [listening] situations".

This paper investigates off-center listening experiences, specifically: the degree of degradation in sound quality; the extent to which a recording (microphone technique, recording room, instrumentation) and the reproduction (loudspeaker, listening room) have an effect on off-center sound degradation; and the identification of the primary sound features leading to a perceived sound degradation.

Griesinger (2001) and Martin (2006) have suggested that inter-channel decorrelation of the loudspeaker feeds increases the sweet spot, which can be achieved for instance by spacing the microphones. To the author's knowledge, no formal listening tests are known to have confirmed this hypothesis.

6.1.1 Definition of Center and Off-center Listening Position

Audio recording and reproduction techniques usually refer to a reference listening point, called the "sweet spot", which draws from perceptual or geometric concepts. The perceptual concepts suggest a vague consensus that the sweet spot is the point in space where a listener is fully capable of hearing the intended audio recording, a "spatial bubble of head positions where the listener maintains the desired perception ...” as described by Rose et al. (2001). Such definitions are ambiguous, because the intentions of a sound design are usually unknown for a listener, and the audio material will likely be reproduced with different loudspeakers in different listening rooms.

Landone and Sandler (2001), define the sweet spot as “the point where the combined wavefront generated by the loudspeaker in the reproduction layout is coherent, or, alternatively, where the listener is roughly equidistant from the radiators”. For domestic sound reproduction, the ITU (1992; 1997) placed the reference listening point in the center of a loudspeaker setup in which the left and right frontal loudspeakers create the correct listening angle of 60° (Figure 6.1).

Because the term sweet spot is ambiguous, we will use Central Listening Position (CLP) to describe the reference listening point where all loudspeakers are equidistant and equally calibrated in Sound Pressure Level (SPL). An Off-Center Listening Position (OCP) refers to all other positions within the loudspeaker array. Multiple loudspeaker feeds coming from different directions will create signals at the ears that differ for each OCP. We present a detailed examination of the geometric considerations that lead to perceptual artifacts.

6.1.2 Geometric Considerations

Time-of-Arrival Differences - ToA

Wavefronts from the loudspeakers will arrive at an OCP with different temporal delays due to path-length differences. The maximal temporal delay t_{max} is calculated in Equation 6.1 from the distance of the closest and farthest loudspeakers s_{min}, s_{max} and the speed of sound v. The further away the OCP is from the CLP, the greater is t_{max}.

$$t_{max} = (s_{max} - s_{min}) \cdot v^{-1} \tag{6.1}$$

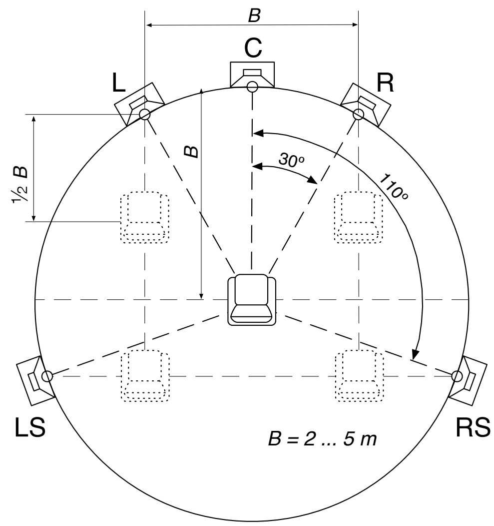


Figure 6.1: ITU BS.1116-1. CLP (solid seat) and worst case OCPs (dashed seats). The recommended listening area is 0.7 m around the CLP.

Unbalanced Sound Pressure Level

If all loudspeakers are calibrated for the CLP, at an OCP a closer loudspeaker will produce a higher SPL than a loudspeaker that is farther away. For a conventional loudspeaker, the attenuation of the direct sound is ca. 6 dB SPL per doubled distance for the direct sound component. This inverse square law determines that the SPL changes very quickly near a loudspeaker but less rapidly at greater distances. Therefore this effect is most prominent in small loudspeaker arrays. Loudspeaker level differences at OCPs depend also on room characteristics and on loudspeaker directivity (Bauer 1960), because of the contribution of reflected sound energy. For steady-state sounds, the attenuation is closer to 3 dB SPL per doubled distance (Toole 2008, 46). Therefore, the degree of off-center sound degradation is probably different for transient and for sustained sounds, and must consequently depend on the musical material.

Direction of Arriving Wavefronts

A wavefront emitted by loudspeaker R in Figure 6.1 arrives from a direction of 30° at the CLP, whereas for a listener at the upper right dashed seat, the same wavefront arrives from the front (0°).

6.1.3 Perceptual Artifacts

Effects on Localization - the Precedence Effect

Depending on ToA differences, angular direction of the loudspeaker feeds, level and spectral content, the sound image might be shifted or even collapse toward the direction of the most prominent feed. The Precedence Effect explains this perception (see Litovsky et al. (1999) and Blauert (1997) for reviews).

Because the Precedence Effect is related to localization processes in the presence of early reflections, it is primarily investigated for localization in rooms, but is also important in multichannel audio reproduction. The main difference between these two scenarios may be that a real sound source has one direct wavefront, from which directional information is decoded, and multiple (to-be-suppressed) early reflections. On the contrary, in multichannel audio the location of an auditory event is formed by the superposition of coherent wavefronts emitted from several loudspeakers. Consequently at OCPs, the auditory system may misapply the Precedence Effect and incorrect signal components could be fused. Each loudspeaker can also cause individual reflections in the listening room that will be superimposed upon the early reflections of the room in which the recording was made.

Effects on Image Stability

One of three factors in the definition of Overall Spatial Quality by the IEC (1997) is Image Stability which relates to the loss in image precision. A few other spatial descriptors were created to study this effect, such as Spatial Clarity, Readability, Locatedness, or Image Focus. Lund (2000) derived a localization-consistency score from subject responses to Robustness, Diffusion and Certainty of Angle of phantom sources.

Distortion of the Spatial Impression

Barron (1971) defined the perceptual effects of a single lateral reflection as Spatial Impression. This term was later replaced by Apparent Source Width (ASW) and Listening Envelopment (LEV) to account for the different effects of early reflections as a function of the arrival time at the listeners ears relative to the direct sound.

ASW describes the spatial extent of a sound source influenced by early lateral room reflections (up to 80 ms). ASW is correlated with Lateral Fraction (LF_E) and with the Inter-Aural Cross-correlation Coefficient calculated from the early energy ($IACC_E$, Okano et al. 1998). ASW was also found to be generated by frequencies above 1 kHz. Lee and Rumsey (2005) found that ASW is closely related to Image Stability, but for certain sounds these two aspects are perceptually distinct. It seems unclear whether the same preference for large ASW as found in concert hall studies exists in sound reproduction (Rumsey 2001).

Late lateral reflections, as described by LEV, contribute to the fullness of sound images around the listener. LEV depends on the front/back energy ratio (Morimoto 1997), on the loudspeaker and the direction of their wavefronts (Hiyama et al. 2002; Muraoka and Nakazato 2007), on the listening level (Soulodre et al. 2003) and on the frequency spectrum primarily below 1 kHz (Bradley and Soulodre 1995). At OCPs, the LEV can become unstable so that “the envelopment illusion is seriously diminished” (Toole 2008, 336).

Timbral Effects

People can distinguish between timbre and spatial aspects in audio reproduction (Bech 1998). Rumsey et al. (2005a) examined their relative importance. Basic Audio Quality (BAQ), a global judgment for sound quality, was evaluated by manipulating spatial and timbral reproduction quality of 5.0 surround audio material perceived at the CLP. A regression model determined that timbre fidelity has a weight of ca. 70% on the BAQ, whereas spatial factors accounted for ca. 30% of the variance. This ratio was found to be somewhat different between naive and experienced listeners (Rumsey et al. 2005b).

Similar but slightly delayed signals create comb filtering. Depending on bandwidth and on the ratio of transient to steady-state sound components, this effect can be perceived in rooms with strong early reflections (Halmrast 2000). The absolute threshold for an audible timbre change rises when these delays increase and thus, when the spacing between the comb-filter peaks becomes narrower (Kuttruff 2009). More complex reflection patterns that create ASW and LEV can mask disturbing timbre changes. In sound reproduction at OCPs, the misalignment of the loudspeaker wavefronts can also lead to perceptible comb-filtering effects. Results from stereophonic listening experiments suggest that it is unclear how this effect is processed by the auditory system (Blauert 1997; Theile 1980, 11). Koenig (1950) and Zurek (1979) found that the perception of sound coloration due to reflections is weaker for binaural compared to monaural listening, which suggests the existence of a binaural decoloration mechanism (Brüggen 2001) or a *central spectrum* (Bilsen 1977). Rakerd et al. (2000) concluded that there may be an echo suppression mechanism mediated by higher auditory centers in which binaural and spectral cues for location are combined so that a timbre change creates a localization change of an auditory event. Because loudspeakers usually face the CLP, their off-axis frequency performance may become apparent due to a contribution to the direct sound at OCPs, causing timbre changes especially for higher frequencies. An accurate prediction of this effect is limited, because off-axis responses vary greatly across manufacturers and models (Toole 2008, 342).

6.1.4 Previous Studies

Several listening tests have been performed to assess differences among surround microphone techniques primarily at the CLP (e.g., Irimajiri et al. 2007; Kassier et al. 2005; Berg and Rumsey 2002). The work by Camerer and Sodl (2001) and Kim et al. (2006) is highlighted here because the present research uses a selection of their surround recordings. In both studies, experienced Tonmeisters simultaneously recorded musical performances with several multichannel microphone techniques and mixed each technique appropriately into 5.0. To ensure a fair comparison, attempts were made to optimize each microphone technique. For the listening tests, expert and trained listeners were asked to judge these recordings according to several spatial and timbral aspects. Spaced microphone techniques were preferred over coincident microphone techniques, the latter represented by the Ambisonics approach. However, studies focused primarily on the CLP and excluded OCPs. An exception is Kamekawa (2006), who attempted to find a microphone configuration that maintains locational, timbral, and spatial attributes at OCPs as well. He tested different 3-channel microphone configurations and found that for timbral and spatial attributes, the microphone arrays accounted for most of the variance. In terms of localization, however, there were no significant differences due to microphone array. The influence of listening position in sound reproduction was primarily studied for localization errors via listening tests (e.g., Marentakis et al. 2008a) or psychoacoustic models (e.g., Pulkki 2002). As mentioned by Pulkki, the success of these models depends on the extent to which they can account for the precedence effect (Section 6.1.3) in the presence of ToA differences in the loudspeaker feeds.

6.2 General Methods

In a listening experiment, the reproduced sound field at different OCPs is compared with the sound field a listener perceives at the CLP of the system.

For this experiment, two sets of 5.0 multichannel recordings were used (EXC, Section 6.2.1). Each set of 5.0 multichannel recordings were previously created simultaneously with three different multichannel microphone techniques (MRT, Section 6.2.2). By reproducing these recordings in two different medium-sized rooms through 5.0 multichannel loudspeaker systems, binaural stimuli were captured at different listening positions (POS, Section 6.2.3). Summarizing, for each tested listening position, six binaural stimuli were captured in total (2 EXC · 3 MRT). In a sound-proof booth and using equalized headphones, trained listeners with normal hearing were asked to compare these binaural stimuli on a computer (Figure 6.2).

The two medium-sized rooms are further described in Section 6.3.1 and Section 6.4.1. These are not domestic listening rooms, but venues that are regularly used for multichannel sound reproduction for larger audiences, which is the focus of this study.

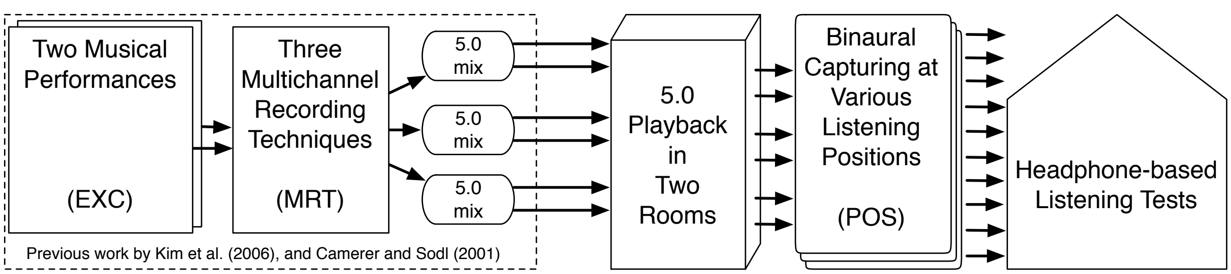


Figure 6.2: General experimental method.

6.2.1 Musical Excerpts–EXC

Each musical excerpt is a 5.0 multichannel recording created from the perspective of a concert audience facing a stage with the instrument sounds arriving from the front and ambient sounds and room response from the sides and behind.

EXC 1.: J.S. Bach "Variation 13", Goldberg Variationen for solo piano (BWV 988).

EXC 2.: W.A. Mozart "Maurische Trauermusik" in c-minor for symphony orchestra (KV 477).

Detailed information concerning the recording and mixing procedures can be found for EXC 1 in Kim et al. (2006) and for EXC 2 in Camerer and Sodl (2001). Microphone settings used for recording the musical excerpts are illustrated in Figure 6.3.

6.2.2 Multichannel Recording Techniques–MRT

Both excerpts were recorded with three prominent multichannel recording techniques (MRT). These techniques differ in their strategy for avoiding inter-channel crosstalk. In one, the microphone directivity pattern increases amplitude separation, and in the other, the displacement of the microphones creates inter-channel time delays. These may be combined. We provide a short overview of these techniques. A more detailed review of them can be found in Rumsey (2001).

Coincident Microphone Technique–Ambisonics

Ambisonics extends the recording technique by Blumlein (1931) by adding an omnidirectional microphone to the pair of figure-eight units. With a third figure-eight unit perpendicular to the other two directional microphones, the vertical component of the sound field is also captured. All microphone capsules are meant to be at exactly the same spot in the sound field, which is encoded in the 4-channel B-Format. To reproduce the sound field, the B-format recording has to be decoded according to the loudspeaker setup. Some computational models have been proposed to evaluate Ambisonics decoders across the listening area (e.g., Daniel et al. 1998). Both excerpts were recorded using a Soundfield MKV microphone and SP451 surround encoder.

Spaced Cardioid Microphone Technique

The Optimized Cardioid Triangle (OCT) proposed by Theile (2001) reduces channel crosstalk by creating both inter-channel amplitude and time differences. Two outer hyper-cardioid microphones face $\pm 90°$ sideways from the center cardioid microphone, which is usually placed 8 cm forward. Optional lowpass-filtered omnidirectional microphones may be applied to enhance the low-frequency response. The OCT array can be extended to OCT surround, or combined with a rear array such as the Hamasaki Square as used in both excerpts.

Spaced Omni Microphone Technique

The applied omnidirectional microphones are widely spaced, primarily creating inter-channel time differences. To account for the differences in source width dimension between a piano and an orchestra, different versions of this technique were used for the two excerpts.

Decca Tree + Hamasaki-Square: The Decca Tree is designed for three omnidirectional microphones arranged in a triangle. The center microphone is placed 0.7–1.0 m forward, whereas the right and left capsulars are spaced at a distance ranging from 1.4 to 2.0 m. For recording large sound sources (e.g., orchestra), the array can be extended with additional microphones to the side. The Decca Tree has been widely used for large-scale recordings and is a favorite among film scoring mixers because of its ability to maintain imaging and separation through the various matrix systems employed in the distribution of film soundtracks. To feed the rear channels in a surround speaker setup, the Decca Tree is usually expanded with a rear array, such as the Hamasaki Square (2003) as used in the recording of the Mozart symphony (EXC 2).

Polyhymnia Pentagon: This technique invented by Polyhymnia International (formerly Philips Classics) uses five widely spaced omnidirectional microphones and is often described as a multichannel version of the Decca Tree. The microphones are arranged in a large circle and their positions correspond to the azimuthal angles of the loudspeaker as in Figure 6.1. The Polyhymnia Pentagon was used for the piano recording (EXC 1).

6.2.3 Listening Positions–POS

The musical excerpts were reproduced through surround loudspeaker setups. At specific listening positions (POS), six sets of binaural stimuli (2 EXC · 3 MRT) of a duration of about 7 sec were made to capture the sound field a potential listener would perceive. Additionally, binaural room impulse responses (BRIR) for each loudspeaker at each POS were measured. All signals were recorded at 48 kHz and 16 bit with an RME Fireface 800. An Apple Macbook running Pure Data was used for playback and recording purposes.

6.2.4 Procedure and Apparatus

The subjects were explicitly asked to "Rate the degradation in sound quality of sound B relative to sound A." Sound A represented one of the six CLP recordings, whereas sound B could be: a) one of the OCP recordings of the same EXC and MRT; b) the same CLP recording as sound A (a hidden reference); or c) a monaural recording, taken at a far corner

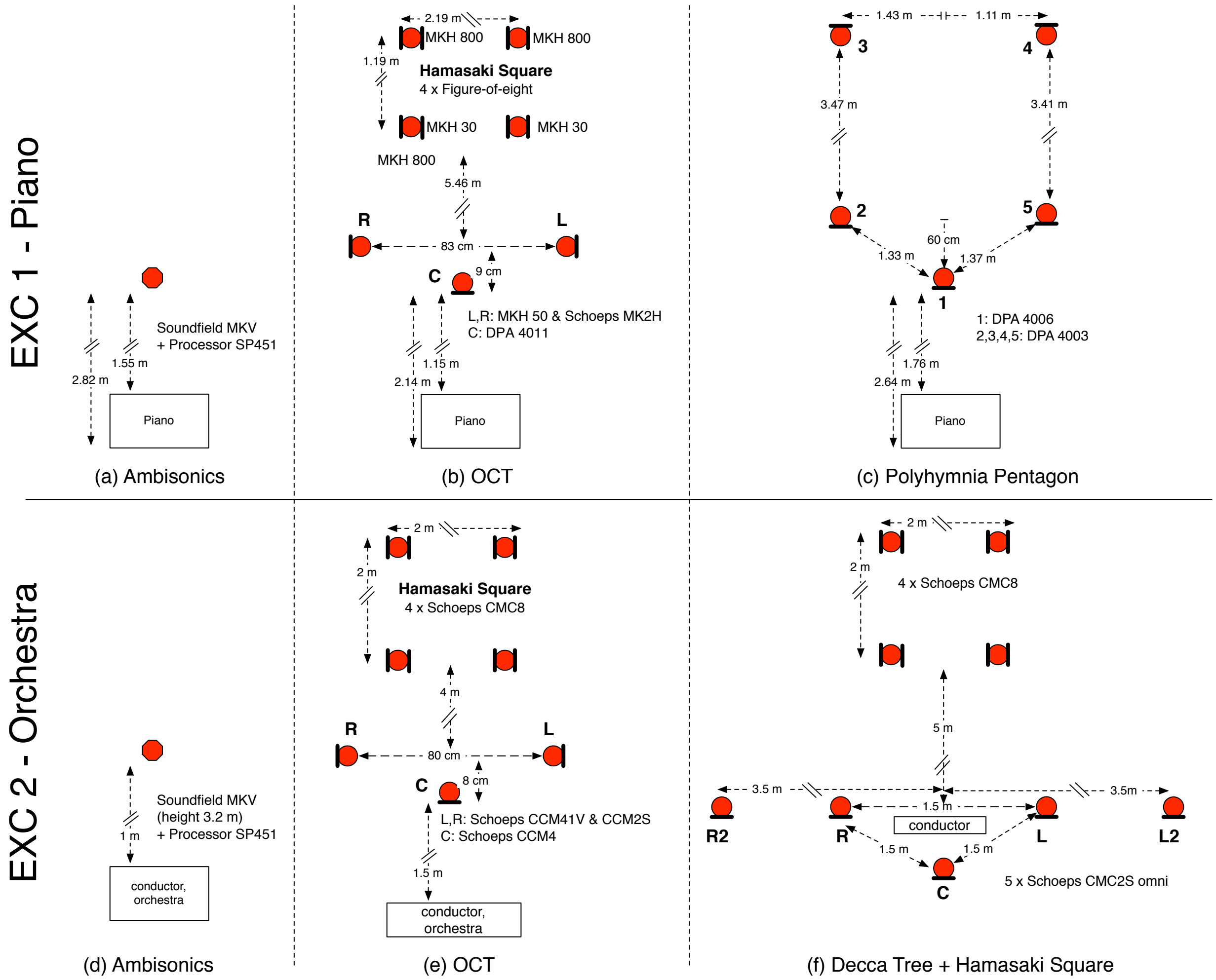


Figure 6.3: Multichannel microphone arrays setups used to record the musical excerpts. (a)-(c) for EXC 1 adapted from Kim et al. (2006) and (d)-(f) for EXC 2 from Camerer and Sodl (2001).

OCP (a hidden anchor). The purpose of the hidden reference and anchor was to set best- and worse-case references for the rating scale, but also to validate participant reliability. In similar experiments, subjects are typically asked to rate the difference between stimuli, but reporting a simple difference does not indicate whether the comparison sample is perceived as better, worse or the same as the reference sample. Therefore we asked explicitly for a sound degradation rating.

The trials were presented in random order. A graphical user interface was designed and the ratings were made with a computer mouse on a slider with a continuous scale from 0 to 100, where 0 corresponds to the bottom (total degradation) and 100 to the top of the scale (no degradation). The slider of the scale was also marked by descriptors in the following order: very strong degradation - strong degradation - moderate degradation - slight degradation - very slight degradation. This corresponds roughly to an analogical-categorical scale and has been shown in psychophysical research to increase response reliability (Weber 1990). Within the presented pair, subjects could switch between sounds A and B at will and could listen as often as necessary to make their ratings.

The experiments took place under similar conditions at the McGill University and Banff Centre facilities and consisted of a training phase, a familiarization phase, and the experimental phase. The subject read the experimental instructions and asked any questions necessary for clarification. In the first phase, five trials with musical excerpts that were different from those presented in the experiment were presented for interface training. Subjects were informed that these ratings would not be recorded. In the second phase, a representative collection of five binaural stimuli of each group (2 EXC $\cdot$ 3 MRT) were used to familiarize them with the musical material. They were told that the familiarization phase would give them the range of variation in sound degradation so they could subsequently use the full scale for their judgments in the experimental phase, which lasted about 60 min. Each pair of binaural stimuli was presented twice to increase the reliability of the data over Sennheiser HD 600 headphones at a normal listening level [70 dB SPL (A-weighted) for the CLP recording]. Besides diffuse-field equalization, no additional headphone equalization was applied. Although in many classic psycho-acoustic experiments, the loudness between stimuli is kept constant, it was decided not to artificially correct for loudness differences between the listening positions, because a loudness change is part of the off-center listening experience. The subjects were told to face the frontal direction and to keep their heads steady. Breaks were permitted whenever they wanted. After the experiment, subjects completed a questionnaire concerning musical practice, headphone listening, sound recording experience, and their rating strategy.

6.2.5 Discussion of Experimental Method

The design of an *in situ* listening test would make it very difficult and almost impossible to allow for real-time, double-blind, comparative and repeatable evaluations. Bregman (1993, 16) reminds us that "simple rules for spatial perception that classical psychophysics has discovered by testing listeners in simple, quiet environments cannot be applied without modification in acoustically complex ones." Hence, our method allowed subjects to select between two binaural stimuli in real time, and therefore had the advantage that listening positions could be compared quickly and repeatedly in a double-blind test. To eliminate the cognitive challenge of participants memorizing the perceived sound field while physically changing listening positions, the subject remained seated while the listening position changed virtually. Furthermore, isolating and presenting the binaural stimuli via headphones, the potential for sound quality biases based on visual cues on the part of the participants was also circumvented.

Our method relies on the assumption that the presentation of the binaural stimuli can evoke all perceptually important elements of the captured sound field as they would have been perceived by a subject directly. Toole (1991) discussed the potential and challenge of using a binaural reproduction system in listening experiments. In particular, the absence of head movements in static binaural recordings and non-individual HRTF cues may cause localization errors mainly in the median plane and in the region of the cone-of-confusion. Møller et al. (1999) measured localization errors in artificial head recordings by using a female speech stimulus emitted from 19 different locations in a standard room. Compared to natural listening, the localization performance decreased significantly and was lowest in the median plane. Therefore it has to be recognized that not all perceptual dimensions may be perfectly reproduced by the reproduction system. However, because the static reproduction process was equal for all stimuli in the listening experiment, it can be assumed that the effect generates a constant bias for all stimuli.

Despite these constraints, several related studies have successfully used similar methods. For instance, to compare listeners preferences of concert hall acoustics, Schroeder et al. (1974) binaurally captured musical excerpts in 33 concert halls and presented them in an anechoic chamber through a transaural loudspeaker setup.

Pfanzagl-Cardone and Höldrich (2008) studied different surround microphone arrays by two presentation techniques: a) subjects listened to the sound field generated by five loudspeakers; and b) to a binaurally re-recorded version of the loudspeaker's sound field presented over headphones. They found that the results of these two presentation techniques differed less than expected and that "the transformation process [through the binaural re-recording] may have 'amplified' the perceived differences between the surround-techniques".

For Olive et al. (1995), subjects rated loudspeakers *in situ* in different rooms. In another test, ratings were given using binaural recordings of the loudspeakers captured in each room. Although some differences in the ratings between the two experiments occurred, the findings were essentially the same. However, the more important finding was that through randomizing the presentation order of the binaural versions, Olive et al. found that there was an effect of human adaption to the room acoustics, which suppresses the acoustical influence of the room on the perception of reproduced sounds (see also Schuck et al. 1993). This effect seems to occur after less than a few minutes of being in a room. Because we virtually relocate the subject in the room for a short period of time (the duration a subject listens to a stimulus), room acoustics effects might be audible that would otherwise be suppressed when listening for an extended period of time. According to Olive (2008), room adaptation is not fully investigated, and therefore one cannot predict the extent to which room adaptation affects the behavioral data of this and other experiments in the field of surround-sound reproduction and room acoustics.

As an alternative to static binaural recordings, a binaural room-scanning system (BRS) could have been used (Olive et al. 2007; Olive 2008). BRS allows head movements through head tracking in the binaural reproduction system, reduces localization errors and increases out-of-head localization. The advantages of BRS displays over static binaural reproduction appear to diminish when room reflections are included in the capturing process (Begault et al. 2001), as is the case in the study presented here. Moreover, BRS signals create a challenge in extracting sound features from a binaural time-variant signal for further analysis as presented in Section 6.5. Also, the front-back confusion in our stimuli may be less severe, because the musical instruments were recorded from the perspective of a concert listener, and it was clear that the instruments appeared only from ahead. Additionally, because the static reproduction process was equal for all stimuli in the listening experiment, it can be assumed that the effect generates an individual, yet constant bias for all stimuli per subject, which may even be limited through the use of high-quality, diffuse-field equalized headphones and instructions to the subject to avoid head movements.

6.3 Experiment A–Telus Studio

6.3.1 Method

Preparation

The Telus Studio at the Banff Centre for the Arts is used as a recording room for medium-large ensembles, lectures, and film presentations. It has a floor-space of ca. 140 m^2 and a volume of ca. 800 m^3. The measured Signal-to-Noise Ratio (SNR) was ca. 45 dB SPL (A-weighted) with respect to the room in quiet. For the reverberation times (Table 6.4) and SNR, the Telus Studio marginally meets ITU BS.1116-1 recommendation for multichannel loudspeaker setups for larger listening rooms. The Schroeder frequency, below which the modal density distribution dominates (Kuttruff 2009, 84), is about 53 Hz. In the recommended bounds of BS.1116-1, the five loudspeakers (Dynaudio BM15A, see Figure 6.7(a) for off-axis directivity) at a height of 1.2 m were placed on an arc with a radius of 4.2 m (Figure 6.5). To capture the binaural stimuli, omnidirectional probe microphones (DPA 4060) were placed at the entrance of the first author's ear canals. To avoid uncontrolled head movements that cause artifacts, a neck-brace was used. The 10 tested positions were chosen as depicted in Figure 6.4 and included the best- and worst-case positions shown in Figure 6.1. The captured listening positions were limited to one side of the listening area because it can be expected that a quasi-symmetrical sound field occurs due to the symmetrical shape of the room and the speaker setup. A monaural recording of POS 10 was chosen as the hidden anchor. The SPL during the recordings varied between 73.5 and 79 dB SPL (A-weighted) depending on position. The level at the CLP was calibrated to 75 dB SPL (A-weighted). In total, 72 pairwise comparisons were prepared for the listening experiment (2 EXC $\times$ 3 MRT $\times$ 12 POS), see Table 6.1.

Participants

Ten listeners (8 male) were tested. They were students from the Sound Recording Program at McGill University, as well as students and staff from the work-study audio program of the Banff Centre for the Arts. Their age varied between 24 and 44 (Median=30) and work experience in sound recording was between 1 and 23 years (Median=9).

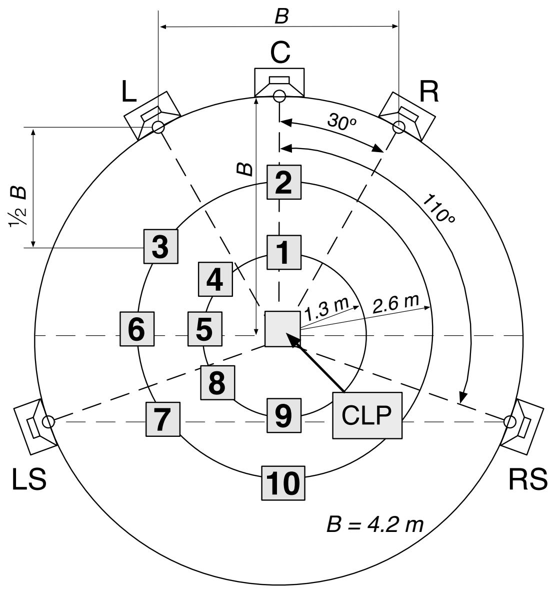


Figure 6.4: Listening positions (POS) in the Telus Studio (Experiment A).

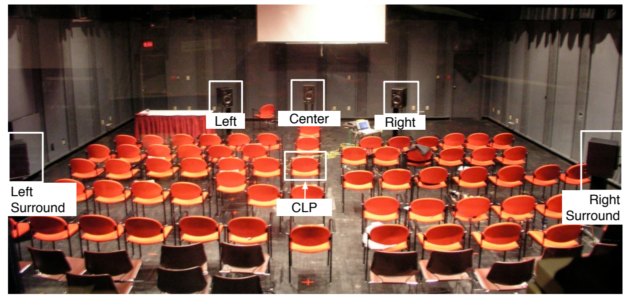


Figure 6.5: Loudspeaker settings.

Table 6.1: Independent variables of Exp. A.
* +2 represents hidden reference and hidden anchor

Independent Variables		Info in Section	Levels
Musical Excerpt	EXC	6.2.1	2
Recording Technique	MRT	6.2.2	3
Listening Positions	POS	6.2.3	10+2*

6.3.2 Results

An EXC(2)×MRT(3)×POS(10) repeated-measures analysis of variance (ANOVA) was performed on the sound degradation ratings. For several stimuli, the ANOVA assumption of normality was violated, which is typical for bounded rating scales. A non-linear arcsin transformation (Snedecor and Cochran 1980) successfully corrected the data and ANOVAs on both sets of data were performed. Because the ANOVA results were similar, we present the uncorrected data analysis which shows that all effects are significant ($p < 0.001$), except for the EXC main effect and the EXC×MRT interaction (Table 6.2). The effect size measure η_P^2 shows that MRT and POS produce the largest effects.

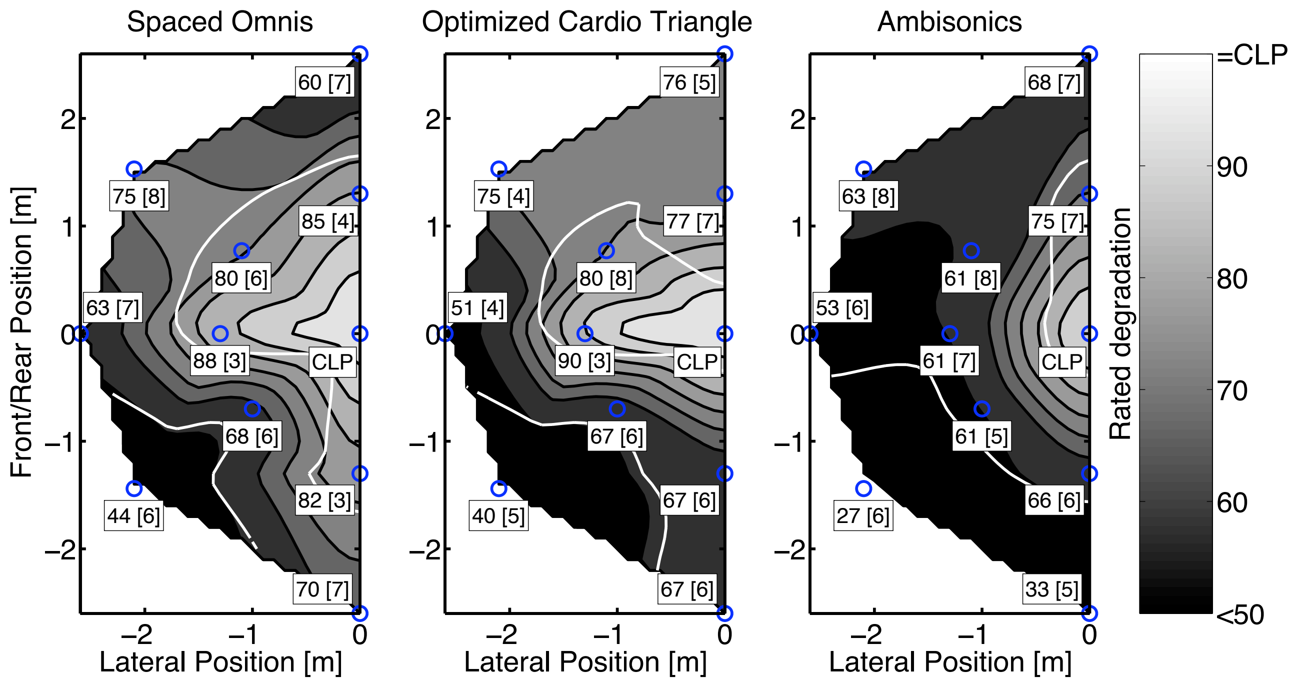
The mean ratings for each position were used to create sound-quality maps of the listening area through three-dimensional contour-plots, defined by the position of the OCPs measured in x and y coordinates from the CLP with mean ratings along the z dimension. A spatial cubic interpolation was used to predict the sound degradation between tested listening positions (Figure 6.6). Furthermore, a Tukey-Kramer HSD post-hoc test was performed to determine which pairs of means are significantly different. Hence, the listening area that was rated similar to the CLP could be determined and is encoded with white lines in Figure 6.6. A Bonferroni post-hoc comparison provided the same results. As illustrated in Figure 6.6, a radial sound degradation from the CLP occurs, but with varying slope across the MRTs. Although the EXC main effect was not found to be significant in the ANOVA (Table 6.2), for the recording techniques OCT and Ambisonics, the listening area of EXC 2 (symphony recording) seems to be slightly wider than for EXC 1 (solo piano). The Spaced Omnis technique for EXC 1 and OCT for EXC 2 created the largest reference listening area. For both EXCs, Ambisonics produced the smallest area. The largest difference between the different recording techniques can be found at the listening position 5 and 10 for EXC 1 and at the listening positions 1 and 2 for EXC 2.

Table 6.2: ANOVA results for Experiment A.
* indicates Greenhouse-Geisser Correction for violation of sphericity

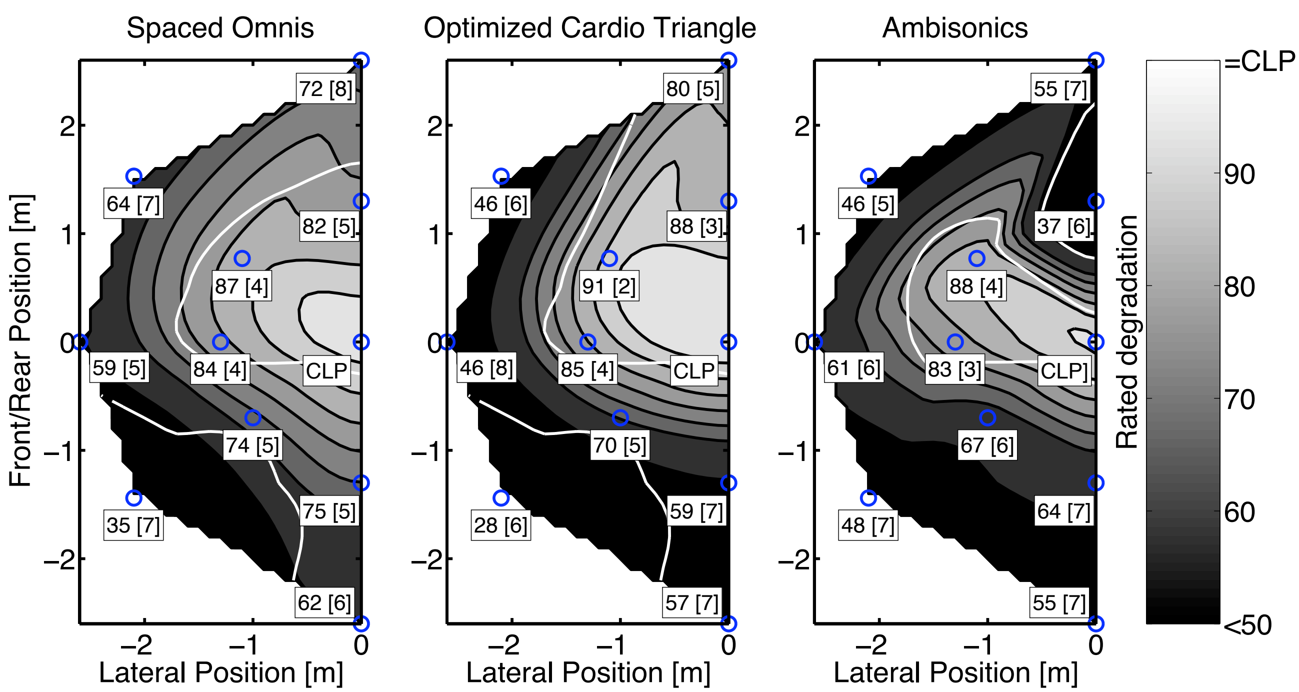
Effect	df	F	ϵ_{GG}	p	η_P^2	η_P^2-Rank
EXC	$1, 9$	0.7	-	.794	.01	7
MRT	$2, 18$	34.6	-	$<$ **.001**	.80	1
POS	$9, 81$	26.9	-	$<$ **.001**	.75	2
EXC×MRT	$2, 18$	3.5	-	.054	.28	6
EXC×POS	$9, 81$	7.1	-	$<$ **.001**	.44	3
MRT×POS*	$18, 162$	5.4	.29	$<$ **.001**	.37	5
EXC×MRT×POS*	$18, 162$	5.2	.30	$<$ **.001**	.37	4

Table 6.3: Mean ratings of MRT and EXC in Experiment A.
mean ratings ± standard error

MRT	EXC 1	EXC 2
Spaced Omnis	71.6 ± 3.7	69.2 ± 4.1
OCT	68.9 ± 4.0	65.0 ± 3.8
Ambisonics	56.8 ± 4.4	60.4 ± 4.0



(a) EXC 1 - Bach piano performance



(b) EXC 2 - Mozart symphony

Figure 6.6: Experiment A: Referring to Figure 6.4 the listening positions (POS) are marked with circles, CLP at (0,0). At each POS, the mean rating and [standard error] are shown. The grouping of similarly rated POS through Tukey-Kramer HSD are shown by white contours.

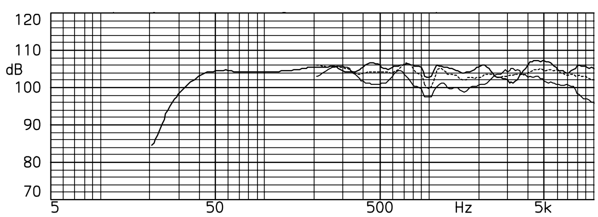
6.3.3 Discussion

The results of the ANOVA suggest that MRT and POS have the biggest effect on the behavioral data. The significant effect of POS is expected and confirms the consensus among listeners and audio engineers concerning the limited listening area of sound reproduction within a ITU five-loudspeaker system. The largest effect size was found to be related to the recording technique (MRT). As can be seen in the illustration of the microphone arrays setups in Figure 6.3, there are uncontrolled variables within the MRT variable. For instance, the type and brand of microphones differs between MRT, and between the two musical excerpts, which can introduce uncontrolled differences due to the microphone on and off-axis frequency response. However, note that for ensuring a fair comparison, renown Tonmeisters have selected the type and placement of the microphones and made an effort to optimize each recording technique.

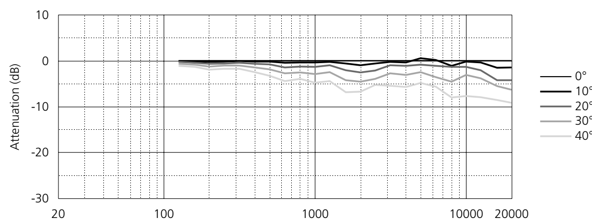
The third-largest effect size was found for the interaction effect EXC$\times$POS, which is visually confirmed in Figure 6.6, e.g., comparing the mean rating of the listening position 1 between both excerpts and the recording techniques shows such an interaction: Spaced Omnis (EXC 1: 85; EXC 2: 82), OCT (EXC 1: 77; EXC 2: 88) and Ambisonics (EXC 1: 75; EXC 2: 37). The ratings for listening position 3 and 7 are particularly interesting, because both positions are classified in the ITU (1997) recommendation as worst-case positions. For both musical excerpts, position 7 received the lowest ratings of all tested positions with an average degradation value of 37 (Figure 6.6). In comparison, position 3 received generally higher ratings (average degradation is 61); for solo piano excerpt, this position was even one of the best-rated seats.

Table 6.4: Reverberation time T_{60} in Telus Studio (Exp. A) and Tanna Schulich Hall (Exp. B).

	[Hz]	125	250	500	1k	2k	4k	8k
T_{60} Experiment A	[sec]	.74	.67	.59	.54	.53	.50	.43
ITU-R 1116-1 recommendation	[sec]	.50	.50	.50	.50	.50	.50	.50
for Telus Studio (Exp. A)		±.15	±.05	±.05	±.05	±.05	±.10	±.10
T_{60} Experiment B	[sec]	.87	.82	.86	.77	.79	.76	.66



(a) Experiment A, Dynaudio BM15A, curves show frequency response for 0°, 30°, and 60° off-axis



(b) Experiment B, Kling & Freitag CA1515-9

Figure 6.7: Horizontal off-axis frequency response of the loudspeakers, provided by the manufacturers.

6.4 Experiment B–Tanna Schulich Hall

6.4.1 Method

To investigate sound degradation as an effect of the reproduction environment, we captured a new set of binaural stimuli. In order to create a very different reproduction environment from the one in Experiment A, several variables in the reproduction chain were changed. These variables are related to the room, loudspeaker type, loudspeaker arrangement, loudspeaker room interaction and the technology to capture the binaural stimuli. We acknowledge that each variable is important and worthy of investigation in itself. However, practical reasons led us to create a *most-different* scenario, in which all of these variables were changed as explained in the following Section.

Preparation

Tanna Schulich Hall (Figure 6.9, McGill University), mainly used for chamber music, jazz- and contemporary music ensembles as well as lectures, has a floor space of ca. 240 m^2 with 188 seats and a volume of ca. 1400 m^3. It is known for its intimacy and the short reverberation time (Table 6.4), resulting in a Schroeder frequency of about 47 Hz. The 5-channel loudspeaker system installed in the hall was used and calibrated for optimal CLP sound quality (Kling & Freitag CA 1515 for the front and CA 1001 for the surround). The directivity of the frontal loudspeakers is illustrated in Figure 6.7(b). Due to the rectangular shape of the room, the positions of the loudspeakers differ from the ITU BS.1116-1 recommendation for the azimuthal angles; instead, the surround speakers are placed at $\pm 150°$ with an arc of ca. 8.2 m, measured from the CLP. It is known that due to this displacement, the expected effect of the surround loudspeaker (to enhance listener envelopment, LEV, Section 6.1.3) is reduced. Furthermore the center speaker is noticeably elevated. Due to the raked seating in the hall, the listening perspective relative to the elevated speakers varies, suggesting that this layout can be considered as a non-standard, yet ecologically valid real-world setup.

A B&K Head and Torso Simulator (HATS) with shoulder damping fabric was placed at 12 positions in the hall (Figure 6.8). The hidden anchor was a monaural recording (see “anchor” in Figure 6.8). The level during the recordings varied within 74 – 77 dB SPL (A-weighted) depending on the position in the hall. In concordance with Experiment A, level at the CLP was calibrated to 75 dB SPL (A-weighted). The measured SNR was ca. 50 dB SPL (A-weighted) with respect to the silent room. The independent variables for the experiment (Table 6.5) yield 78 conditions.

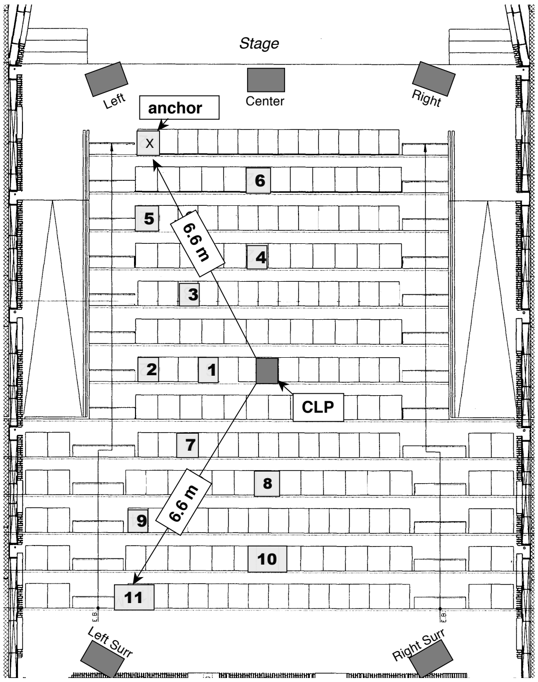


Figure 6.8: Listening positions in Tanna Schulich Hall.

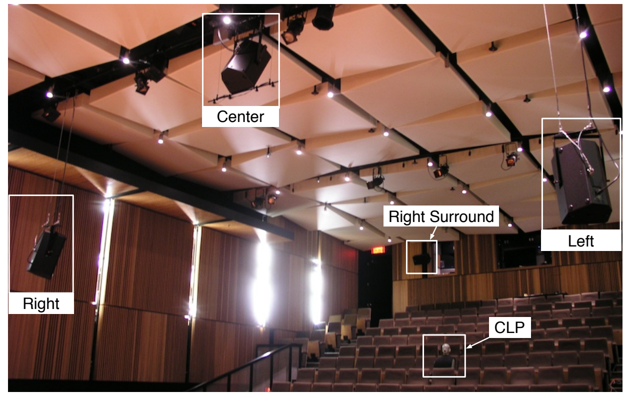


Figure 6.9: Non-ideal loudspeaker settings.

Table 6.5: Independent variables of Exp. B.
* +2 represents hidden reference and hidden anchor

Independent Variables		Info in Section	Levels
Musical Excerpt	EXC	6.2.1	2
Recording Technique	MRT	6.2.2	3
Listening Positions	POS	6.2.3	11+2*

Participants

Nineteen subjects (16 male) with normal hearing and trained in critical listening participated in the experiment. They were students from the Sound Recording Program at McGill University, as well as students and staff from the work-study audio program of the Banff Centre for the Arts. All of the subjects of experiment A took part in this test. Ages ranged from 23 to 44 (Median=27) and work experience from 1 to 23 years (Median=7). On average, the subjects reported headphone usage of about 1.7 hours per day. Because of technical problems, the hearing of nine subjects was not properly tested according to ISO 389-8 (2004). None of the subjects reported any current or previously experienced hearing damage in a dedicated section of the post-experimental questionnaire. Further, using a t-test, the behavioral data of the tested and non-tested subjects were not found to be significantly different.

6.4.2 Results

Because of the large size of the listening area, an EXC(2)×MRT(3)×POS repeated-measures ANOVA was performed on the sound degradation ratings for two conditions: 1) accounting for all 11 positions (allPOS), and 2) accounting for the 6 central positions close to the CLP (1-4, 7, 8 in Figure 6.8).

Results are shown in Table 6.6. All main effects (EXC, MRT, POS) and interactions were found to be significant. The POS main effect and the EXC×MRT interaction have the biggest effect size followed by MRT, whereas the MRT×POS interaction was least important.

For centerPOS all main effects and interactions were found to be significant too, but the order of importance indicated by effect size η_P^2 changed (rightmost column in Table 6.6). MRT has the biggest effect, followed by EXC×MRT, POS, and EXC. The effect of POS in centerPOS is less dominant than in allPOS. As for allPOS, the interactions EXC×POS, MRT×POS, EXC×MRT×POS are significant, but less effective.

The sound quality maps based on the average ratings of the listening area are depicted in Figure 6.11. The white contour lines indicate the grouping of similar POSs through Tukey-Kramer HSD post-hoc tests. For EXC 1 (piano), the biggest reference listening area across recording techniques (Figure 6.11(a)) was created by the Spaced Omnis technique. For the other two recording techniques, the post-hoc analysis revealed reference listening areas that were similar in size. Figure 6.11(b) for EXC 2 shows equivalent sound degradation for Spaced Omnis and OCT that are different from the area of the Ambisonics technique. Generally for all three recording techniques, in EXC 2 the reference listening area around the CLP is bigger than in EXC 1.

Table 6.6: ANOVA results for Experiment B.
* indicates Greenhouse-Geisser correction for violation of sphericity

Effect	df	F	ϵ_{GG}	p	η_P^2	η_P^2-Rank
allPOS						
EXC	$1, 18$	10.5	-	**.004**	.37	4
MRT	$2, 36$	15.2	-	$<$ **.001**	.46	3
POS*	$10, 180$	69.0	.37	$<$ **.001**	.79	1
EXC × MRT	$2, 36$	28.2	-	$<$ **.001**	.61	2
EXC × POS*	$10, 180$	5.6	.55	$<$ **.001**	.23	5
MRT × POS*	$20, 360$	4.2	.43	$<$ **.001**	.19	7
EXC×MRT×POS*	$20, 360$	5.1	.41	$<$ **.001**	.22	6
centerPOS						
EXC	$1, 18$	11.4	-	**.003**	.39	4
MRT	$2, 36$	27.7	-	$<$ **.001**	.61	1
POS*	$5, 90$	13.9	.54	$<$ **.001**	.44	3
EXC × MRT	$2, 36$	19.3	-	$<$ **.001**	.52	2
EXC × POS	$5, 90$	2.5	-	**.037**	.12	7
MRT×POS*	$10, 180$	3.3	.59	**.001**	.16	6
EXC×MRT×POS	$10, 180$	5.0	-	$<$ **.001**	.22	5

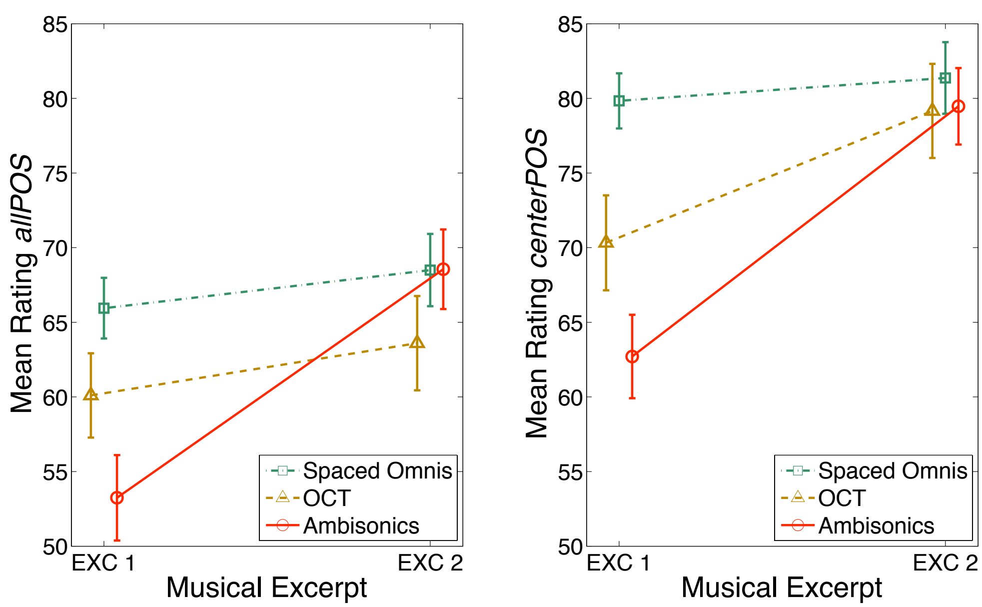
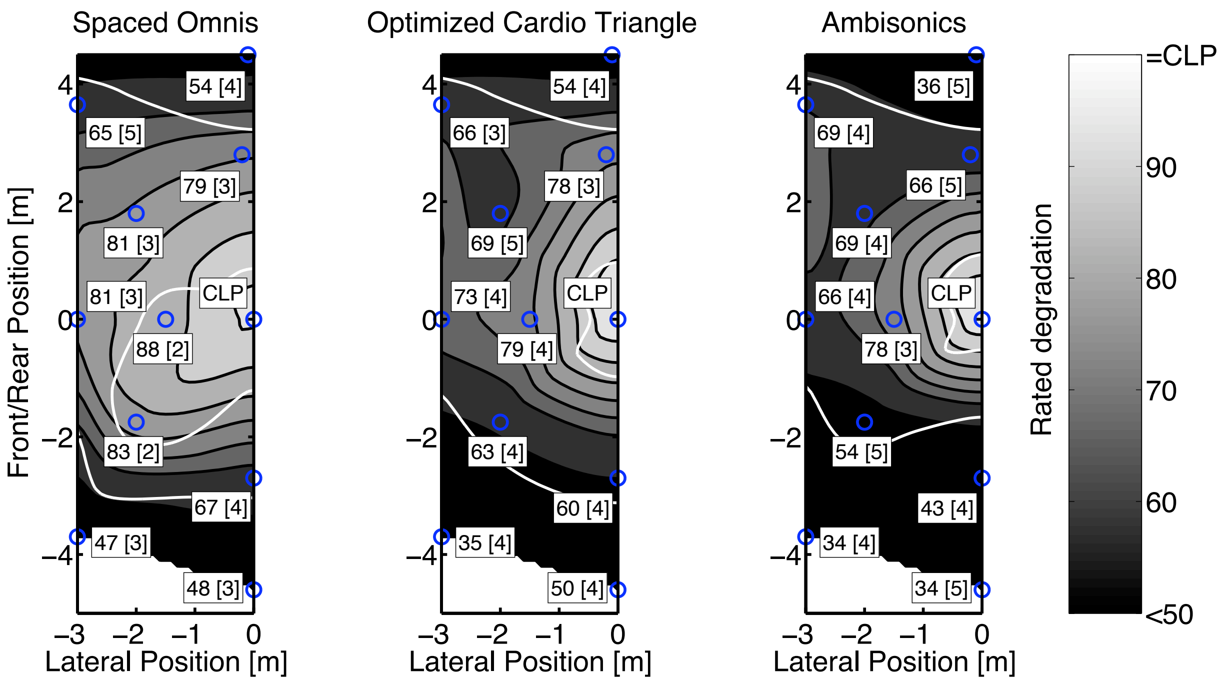
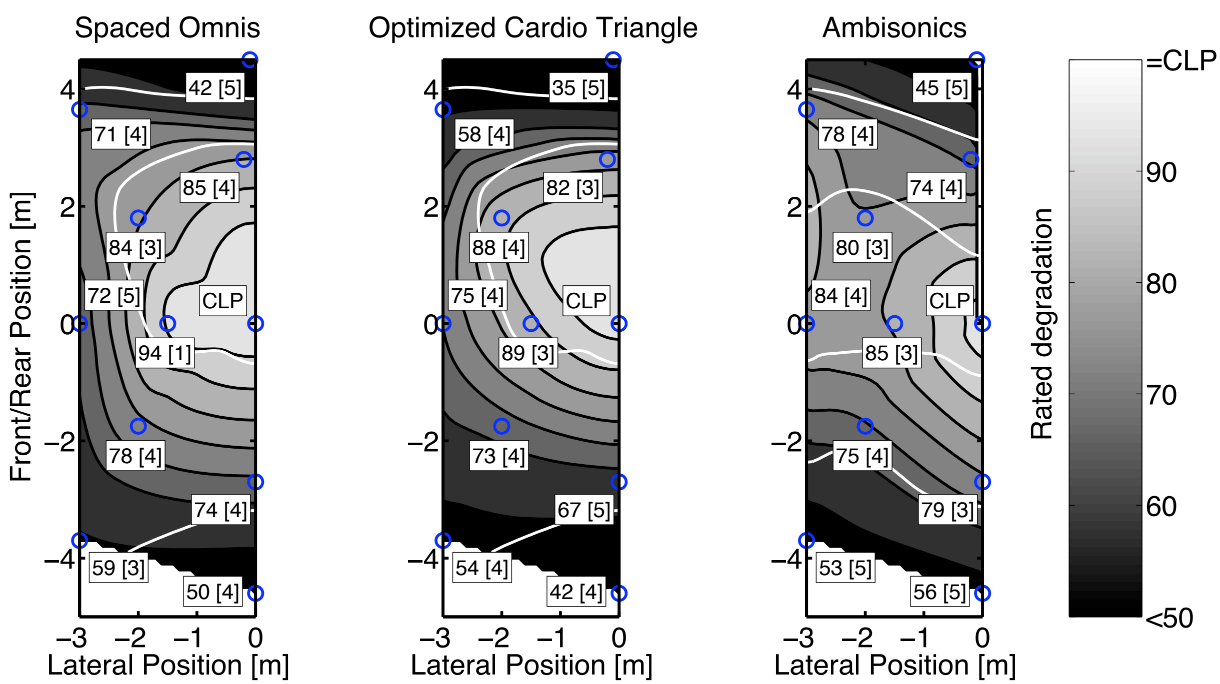


Figure 6.10: Mean sound degradation ratings relative to the CLP as a function of MRT and EXC. Error bars show the standard error across subjects.



(a) EXC 1 - Bach piano performance



(b) EXC 2 - Mozart symphony

Figure 6.11: Experiment B: Referring to Figure 6.8 the listening positions (POS) are marked with circles, CLP at (0,0). At each POS, the mean rating and [standard error] are shown. The grouping of similarly rated POS through Tukey-Kramer HSD are shown by white contours.

6.4.3 Discussion and Comparison with Experiment A

The big effect size of POS in allPOS shows that the listening position has the most influence on the perceived sound degradation. This is expected and is similar to the results of Experiment A. Unexpectedly and contrary to Experiment A, the EXC×MRT interaction, also visible in the mean sound degradation ratings relative to the CLP as a function of MRT and EXC (Figure 6.10), is significant and fairly large in effect size.

This interaction suggests that in this non-ideal listening environment, none of the tested recording techniques is superior and depends on the musical excerpt. We have to acknowledge, in any experimental situation that approaches the real world, many variables between the musical excerpts are uncontrolled. Here, the identified variables ranges from the recording room (early reflections, reverberation) and the recording configurations (e.g., microphone type, distance to the performers) to musical parameters (e.g., instrumentation, tempo, spectrum, complexity, style). As illustrated in Figure 6.10 and also in contrast to Experiment A (Table 6.3), the mean ratings of the perceived sound degradation are higher for EXC 2 than for EXC 1, suggesting that in a more reverberant listening room, a more uncorrelated sound material (EXC 2) is likely to be better reproduced at OCPs than a more highly correlated (solo piano, EXC 1) signal. One could also argue that under non-ideal reproduction conditions, the task of interpreting a multi-instrumental symphonic excerpt (EXC 2) is more complex than it is for a solo piano performance (EXC 1), and that consequently the cognitive load for detailed processing is much higher. This might have caused subjects to pick up on more holistic sound features, i.e., the overall brilliance of the orchestra, rather than the reproduction quality of each individual instrument, thereby influencing the decision-making process.

The increase of MRT effect size and the decrease of POS effect size in centerPOS compared to allPOS suggest that the MRT effect is more pronounced around the CLP and that outside of centerPOS, the MRTs are perceived similarly. Moreover, the EXC×POS interaction is only weakly significant in centerPOS. The mean ratings across MRTs for allPOS and centerPOS in Figure 6.10 show that for EXC 2 the mean ratings become more similar. An ANOVA performed on the sound degradation ratings in the centerPOS for EXC 2 (not displayed) confirmed that the MRT main effect is not significant there.

Figure 6.11(a) demonstrates similarly to Figure 6.6(a) of Experiment A, how the sound quality in EXC 1 degrades radially from the CLP, but with varying slope across the different recording techniques. The plots show least degradation for Spaced Omnis and the most degradation for Ambisonics at OCPs. The largest differences for EXC 1 between recording techniques can be found at the listening positions 7 and 8. For EXC 2, the contour plots in Figure 6.11(b) are less uniform and show less pronounced differences between recording techniques with the biggest difference at position 5. Ambisonics was rated much better than in EXC 1, in particular for position 8. Furthermore, at the positions 2 and 5, the Ambisonics recording produced the least sound degradation of all three recording techniques for EXC 2.

Because the subject's task did not involve directly comparing the sound degradation between Experiment A and Experiment B, one cannot compare the behavioral data of these two experiments directly. However, one can compare the relative performance of each recording technique and each musical excerpt. Based on the sound degradation ratings for each musical excerpt, a Bonferroni pairwise comparison between recording techniques indicates the similarity between ratings as a function of the recording techniques (Table 6.7). For EXC 1, the recording techniques are mostly dissimilar, whereas for EXC 2, the techniques tend towards creating similar sound degradation. For Experiment B, the centerPOS results show that all three recording techniques were rated differently for the Piano recording, but equally for the Symphony recording. Further, the results of the comparison for the Symphony recording are inverted between Experiment A and Experiment B allPOS, indicating another effect of the listening environment.

Table 6.7: Bonferroni Comparison of MRTs.
"≠" indicates a significant dissimilarity; "=" indicates similarity based on a 95% confidence interval.

Compared MRT	OCT vs. Spaced Omnis	Ambisonics vs. Spaced Omnis	OCT vs. Ambisonics
EXC 1 (Piano)			
Experiment A	=	≠	≠
Experiment B centerPos	≠	≠	≠
Experiment B allPos	≠	≠	≠
EXC 2 (Symphony)			
Experiment A	=	≠	=
Experiment B centerPos	=	=	=
Experiment B allPos	≠	=	≠

6.5 Modeling the Behavioral Data from the Binaural Stimuli

A goal in sound quality research is to relate behavioral data to sound features extracted from the presented stimuli. The following presents some related studies.

George et al. (2006) used a Ridge Regression Model to predict preferences of "Frontal Spatial Quality" and "Surround Spatial Fidelity" for different spectrally and spatially degraded (bandwidth limitation and down-mixing) five-channel program items. The spectral sound features were measured from a mono mix-down, summing all loudspeaker channels together. For measuring spatial sound features, binaural recordings were synthesized using convolution of the loudspeaker signals with KEMAR dummy-head HRTFs (Gardner and Martin 1994) for different head positions. From all 22 extracted features, measures based on the Interaural cross-correlation coefficient (IACC) as well as the spectral features "Centroid of the spectral coherence" and "Spectral Roll-Off" had the most impact on the rating. The final models showed a regression coefficients of $R(135) = .91$ for the "Frontal Spatial

Quality" and $R(135) = .95$ for "Surround Spatial Fidelity" with the subjective ratings.

In a pairwise-comparison listening experiment Choisel and Wickelmaier (2006) retrieved auditory attributes from pop music and classical excerpts, which were presented through eight different multichannel reproduction formats (from mono to 5-channel surround). The behavioral ratings of seven auditory attributes were correlated with seven spatial and spectral sound features. The strength of the correlation varied between the musical genres in general, but the spatial measures based on "IACC" and "Lateral Fraction" accounted well for the variance in the spatial auditory attributes "Width" and "Spaciousness". In contrast, no significant correlation was found between any auditory attribute and the spectral features "Spectral Centroid" and "Sharpness".

Kim et al. (2006) analyzed overall preference ratings between surround microphone techniques in several piano recordings. Eighteen features were extracted from a dummy-head re-recording captured at the CLP of a 5.0 surround loudspeaker setup. A stepwise multiple regression model revealed that the measures "Ear Signal Incoherence" and "Side Bass Ratio" predicted the preference ratings reasonably well ($R(16) > .91$).

6.5.1 Method

Sound features from the binaural stimuli were extracted according to a feature set created from the subjects' post-experimental responses regarding their rating strategies and from common descriptors that relate to perceptual artifacts described in Section 6.1.3. These features were extracted in Matlab through existing toolboxes and specifically developed code according to references provided in Table 6.8. The reader is invited to seek more detailed information through these references.

To account for time-varying sound features, binaural stimuli were analyzed across their entire duration using a sliding window of 50 ms duration and a step-size of 25 ms. Final descriptors were extracted using the statistical functions:

xMean: Mean value of data vector x

xMax: 95% percentile of data vector x

xStd: Unbiased standard deviation of data vector x

Spectral and energy features were analyzed for each ear signal separately. We also computed their average and the difference between the ear signals.

Interaural features that are based on the IACC are usually calculated from a binaural room impulse response (BRIR) within certain temporal boundaries (Okano et al. 1998). For continuous musical signals, this procedure is not applicable. Hence, the IACC-based features were calculated from the binaural signals within the sliding window duration of 50 ms and a step-size of 25 ms. According to Griesinger (1997), IACC measured from musical

Table 6.8: Sound features extracted from the binaural stimuli.

Feature	Reference
Spectral Features	
Spectrum in 1/3-octave bands from 50 - 8000 Hz	
Spectral Centroid	Peeters (2004)
Spectral Flux	Peeters (2004)
Spectral RollOff	Peeters (2004)
Narrow Band Deviation (NBD)	Olive (2004)
Absolute Average Deviation (AAC)	Olive (2004)
Sharpness	Fastl and Zwicker (2007)
Roughness	Daniel and Weber (1997)
Attack Strength	Buchholz and Mourjopoulos (2004)
Autocorrelation Function (ACF)	Ando (1998)
Energy Features	
Loudness	Fastl and Zwicker (2007)
dB SPL (A-weighted, 100 ms integration time)	
Interaural Features	
Broadband IACC	
IACC in 1/3-octave bands from 50 - 8000 Hz	
Modified IACC	Mason et al. (2005)

signals, rather than BRIRs, usually gives lower values and depends on the average length of the notes and on the presence of vibrato.

Feature Rescaling

Because each OCP recording was rated against the reference recording taken at the CLP in these experiments, the sound features may be scaled in order to indicate the relation to the CLP. Besides the difference between sound feature X_i from OCP i and X_{CLP} from the CLP (X^a, Eq. 6.2) and their absolute value (X^b, Eq. 6.4), a rescaling method adapted from Choi et al. (2007) was used (X^c and X^d, Eq. 6.3 and 6.5). All rescaling methods were applied to each feature. The version that best correlated with the behavioral data (Spearman rank correlation) was used as a candidate for the final model. Most sound features showed best correlation in a non-scaled version.

$$X_i^a = X_{CLP} - X_i \tag{6.2}$$

$$X_i^c = X_{CLP} \cdot (X_{CLP} - X_i) \tag{6.3}$$

$$X_i^b = |X_i^a| \tag{6.4}$$

$$X_i^d = |X_i^c| \tag{6.5}$$

Prediction Strategy

For relating sound features to behavioral data, a regression approach is commonly chosen, aiming to create a best-fitting model while minimizing collinearity between selected sound features. Consequently, representative sound features may be discarded from the model due to collinearity. To increase the model's meaning while avoiding collinearity, a data reduction method guided by a cluster analysis was performed (Giordano et al. 2010) to combine strongly correlated features before a regression model was created. The initial stage of this procedure requires computing a measure of the distance between sound features, given by the absolute value of their Spearman correlation. The distances are then analyzed with a clustering method (agglomerative hierarchical cluster analysis, average linkage). Finally, each of the clusters of strongly correlated sound features is merged independently into a single variable by means of Principal Component Analysis (PCA). The first Principal Component for each of the clusters is retained as the final reduced variable. Starting from the condition where each feature is in an isolated cluster, the number of clusters can be progressively decreased, thus yielding an increasingly lower number of reduced variables. As the number of clusters decreases, the correlation between reduced sound features decreases whereas the extent to which the original sound features are well represented in the remaining clusters decrease. The final number of clusters is chosen by trading off these two factors.

The mean behavioral data for each tested OCP and all three recording techniques were modeled for each musical excerpt and experiment, resulting in four prediction models. It was decided to create a separate model for each musical excerpt in order to account for the different musical properties that may have influenced the behavioral data. All models compensate for tied ranks, freeing the model from potentially nonlinear monotone relations among variables.

6.5.2 Results

The final prediction models for both experiments are presented in Tables 6.9 and 6.10. The chosen predictors are sorted according to the order in which they were selected by the stepwise regression model. These predictors represent spectral and interaural sound properties. Energy sound features were not selected by the procedure.

The rightmost column of the tables shows how the accuracy of the prediction increases with each regression step to an adjusted R^2 value between 0.80 and 0.94. No collinearity was observed between the successful predictors. This is a promising result, considering that each model predicts the behavioral data of all tested off-center listening positions and all three microphone techniques.

Table 6.9: Prediction model of Experiment A - Telus Studio.

Draft No.	Predictor successively entered	Partial R	Std. coeff. B	Accum. $R^2_{adj.}$
EXC 1 (Piano)				
1.	Attack strength (Right ear)	-.81	-.58	.47
2.	IACC at 400 Hz (Std)	-.64	-.35	.63
3.	Spectr. Energy at 800 Hz (left ear)	-.57	-.29	.73
4.	IACC at 400 Hz (Max)	-.53	.27	.80
EXC 2 (Symphony)				
1.	IACC at 5 kHz	-.78	-.54	.50
2.	Roughness (right ear)	-.67	-.38	.70
3.	ACF τ_E (left ear)	-.64	-.35	.78
4.	Energy difference in ear signals at 6.3 kHz	-.44	-.22	.82

Table 6.10: Prediction model of Experiment B - Tanna Schulich Hall.

Draft No.	Predictor successively entered	Partial R	Std. coeff. B	Accum. $R^2_{adj.}$
EXC 1 (Piano)				
1.	Spectr. Energy at 2.5 kHz (right ear)	-.74	-.55	.68
2.	Spectral RollOff (averaged ear)	-.57	-.30	.75
3.	IACC at 5 kHz (std.)	-.49	-.26	.80
4.	IACC at 2 kHz (std.)	-.39	-.18	.82
EXC 2 (Symphony)				
1.	Spectral Energy above 2.5 kHz, Narrow Band Deviation above 4.3 kHz	-.82	-.55	.78
2.	ACF ϕ (right ear)	-.80	-.34	.89
3.	Spectr. Energy at 2 kHz (left ear)	-.56	-.25	.92
4.	IACC 500 - 2000 Hz	-.55	-.18	.94

6.5.3 Discussion

The behavioral data from both experiments could be reasonably well modeled with four parameter clusters. Because stimuli were intentionally not equalized for loudness across listening positions, it is particularly interesting to see that neither Zwicker's loudness nor the A-weighted overall SPL measure contribute to any of the four models, suggesting that loudness differences in these experiments are unrelated to sound degradation.

The data of Experiment B were better predicted than the data of Experiment A. For both experimental datasets, the symphony excerpt (EXC 2) is slightly better modeled than the piano excerpt. The subjects' post-experimental responses suggested that for certain stimuli, particularly in Experiment A, specific unpleasant reflection patterns contributed to the ratings. It is quite likely that these specific reflections are not accounted for by the extracted sound features and therefore created some prediction errors. The prediction models differ across experiments and across musical excerpts and show the complex nature and the different characteristics of perceived sound degradation, suggesting an interaction of musical material with the reproduction environment. For instance, for both excerpts in Experiment B, a strong contribution to the perceived degradation of the spectral energy above 2.5 kHz was observed. In contrast, these features did not appear in the final prediction models of Experiment A and it is quite likely that the higher off-axis directivity of the loudspeakers used in Experiment B contribute to this predictor (Figure 6.7). For Experiment A, the strongest predictor for the piano excerpt (EXC 1) was Attack Strength and for the symphony excerpt (EXC 2), the IACC in the 5 kHz band. Furthermore, the prediction models for one musical excerpt across the two experiments show only a few commonly shared features. Only properties of the Autocorrelation function (ACF), which was included to indicate spectra-temporal sound modification (e.g., due to reverberation of the listening room), was found in the symphony excerpt for both experiments. Many other studies have also found that sound quality preference judgments depend on the audio material, (e.g., Ando 1998; Olive 2008) and on the acoustical conditions (Lee and Rumsey 2005). Increasing variability of sound quality judgments due to room reverberation, which could have contributed to the differences in behavioral data between the experiments, was reported by Gilford (1959).

However, the final predictors are plausible and reflect the perceptual artifacts outlined in Section 6.1.2 and 6.1.3. For instance, the loss in Attack Strength, found as the strongest predictor for the piano excerpts in Experiment A, can be caused by the time-of-arrival differences and resulting comb-filters that smeared out the attacks of the piano sound, a percept that would be masked by the natural *chorus* sound of the string section of EXC B. The IACC in the range of 500 to 2000 Hz, found for EXC 2 in Experiment B relates to distortion of the spatial impression and is also used to calculate the Binaural Quality Index, a binaural measure in concert hall acoustics (Beranek 2003). In all four prediction models, timbral features (i.e., Spectral RollOff, Roughness, Attack Strength) have a major contribution, whereas binaural attributes that relate to spatial descriptors are of secondary importance. These findings can be supported by Rumsey et al. (2005a). Further, Gabrielsson et al. (1974) assumed that limited contributions of the binaural features may result from stronger intra-individual variability in perceiving spatial descriptors than in perceiving timbral attributes. Consequently timbral features correlate more strongly with mean behavioral data.

6.6 Final Discussion and Further Studies

The results show that when reverberant, classical, multichannel recordings are reproduced in a medium-sized, moderately reverberant space, the useful listening area is larger than it is in a smaller, less reverberant space. It seems possible that the space itself is adding credible reflected sounds to the mix of sounds arriving at the listeners' ears and that the space favors sound sources which are reproduced by relatively uncorrelated loudspeaker feeds.

Ambisonics provides a restricted listening area in both rooms, compared to the two other MRTs. Based on the relatively good performances of the two Spaced Microphone

Techniques (Section 6.2.2 and 6.2.2) compared to the coincident microphone technique, this suggests that time-delay based decorrelation among the loudspeaker feeds (a.k.a. Interchannel Time Differences) supports better off-center listening. To investigate this suggestion in detail, one has to examine the instrument/microphone/room interaction at the recording site through impulse responses generated at the position of the instruments and capture with the tested microphone arrays. Unfortunately, Kim et al. (2006) and Camerer and Sodl (2001) did not provide these data.

The quality of reproduced sound at the different listening positions will vary at low frequencies as a function of the spatial relationship of the loudspeakers, listener and the acoustical properties of the listening room. The use of full-range loudspeakers (without bass-managed subwoofers) further complicates how these variables interact with each other because the loudspeakers will produce interference patterns at low frequencies depending on the correlation of the recorded signals among the different channels. Although we did not find any low-frequency contribution in our prediction models, one cannot reject the possibility that for other sound material the bass perception at different listening positions can vary.

In Experiment B, we showed how a non-ideal reproduction environment affects off-center sound degradation by emphasizing the effect of the musical excerpts. For further understanding, future experiments need to uncouple all those variables that were combined here, i.e., the room, loudspeaker type, loudspeaker arrangement. Moreover, as a trade-off between less controlled, yet real-world, microphone recordings and more controlled ones, the Virtual Microphone Control system (ViMiC, Braasch et al. 2008) could be employed to generate fully controlled musical excerpts.

Also, modeling behavioral data based on sound features from the original 5.0 multichannel recordings could provide further knowledge about off-center sound degradation.

According to signal theory, a binaural stimulus results from convolving the 5.0 surround recording with the listening room, captured via BRIRs. As illustrated in Figure 6.12, sound features measured in the binaural stimuli must consequently be based on features from the BRIR and the original 5.0 multichannel recordings. Future work is needed to fully understand how to model these indirect relationships.

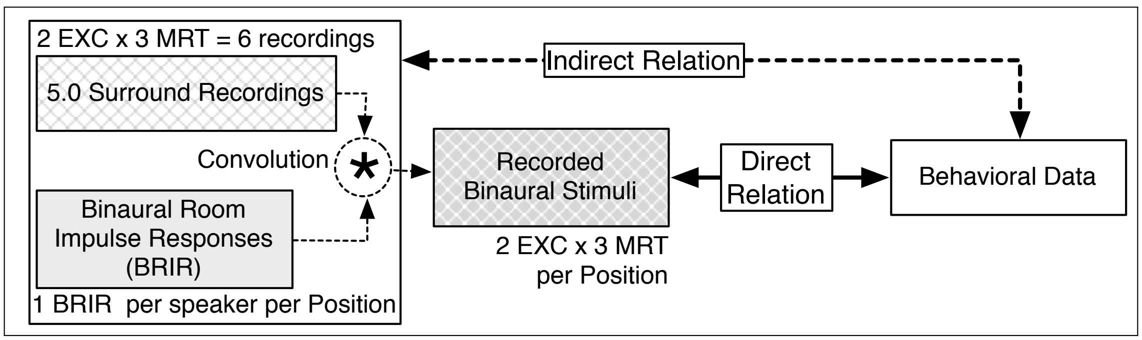


Figure 6.12: Relations between audio material and behavioral data.

6.7 Conclusion

In two perceptual experiment, the off-center sound degradation in a multichannel reproduction setup was determined with respect to three recording techniques (MRT), two classical musical excerpts (EXC), and several off-center listening positions (POS) within two listening environments. It was found that the tested recording techniques affect the sound degradation strength at off-center listening positions and the size of the reference listening area. For all techniques, a somewhat radial sound degradation from the CLP occurs, but with varying slope across the MRTs. Spaced microphone techniques create less sound degradation for off-center listening than the coincident Ambisonics technique.

In a non-ideal listening environment (non-ideal loudspeaker configuration and listening room acoustics), none of the tested recording techniques was superior. Their off-center performance was found to interact with the reproduced audio material. These data suggest, that in a more reverberant listening room, a more uncorrelated sound material such as a symphony recording is likely to be better reproduced at off-center listening positions than a more highly correlated (i.e., Piano) recording.

Furthermore, the behavioral data were modeled from the stimuli's sound features related to energetic, spectral, and interaural sound properties with a regression approach using clusters of similarly grouped sound features. Although the models differ across audio material and listening environment, the primary predictors are spectral and the secondary ones are based on spatial aspects, suggesting that off-center sound degradation is primarily perceived due to sound modifications in the higher frequency range, which can also relates to the off-axis loudspeaker frequency response. The findings can help to inform the design of surround-sound reproduction environments and to inform the production of dedicated audio material which can be appreciated in medium sized rooms by a larger audiences, for instance, the emergence of opera broadcasting into cinemas (Heyer 2008).

6.8 Acknowledgments

This work was funded by a grant from the Canadian Natural Sciences and Engineering Research Council and the Canada Council for the Arts (NSERC, CCA) to Stephen McAdams and by a student award from the Centre for Interdisciplinary Research in Music Media and Technology (CIRMMT) to Nils Peters. Many thanks to Bruno L. Giordano, Georgios Marentakis, Sungyoung Kim and William L. Martens for fruitful advice.

Chapter 7

Off-center Sound Degradation Based on Geometrical Properties

The following Chapter is an extended version of the manuscript:

Peters N., McAdams S.: Off-center Sound Degradation Based on Geometrical Properties in 5.1 loudspeaker setups. prepared for submission.

To improve the reproduction quality at off-center listening positions (OCPs) in surround loudspeaker setups, one needs to understand the nature of the perceived artifacts. Building upon the research of Chapter 6 and based on the geometrical relationships of a listener to the loudspeaker in a surround setup, an OCP can be defined by three attributes: time-of-arrival differences (ToA), sound-pressure-level differences between the signal feeds, and the direction of the arriving wavefronts. Two listening experiments were designed to elict the perceptual effects of the off-center sound degradation of each of these three attributes in qualitative and quantitative terms.

The five most often qualitatively described artifacts between the tested listening positions are related to the position of sound sources; the distance and depth of sound sources; reverberation and envelopment; spread and width of sound sources; and sound coloration. The quantitative study revealed that off-center sound degradation is primarily caused by the level differences of the loudspeaker feeds and that the time-of-flight differences have a stronger perceptual effect on percussive sound material than on sustained sound material. In two out of three musical excerpts, off-center sound degradation was primarily correlated with artifacts related to the reproduction quality of reverb and envelopment.

7.1 Introduction

As outlined in Section 6.1.2, off-center sound degradation is rooted in a physical change of the relation between a listener and the loudspeaker setup geometry. More precisely, the three identified factors caused by geometrical relations are time-of-arrival differences, sound-pressure-level differences between the signal feeds, and the direction of the arriving wavefronts (see Section 6.1.3 for details). The following research questions were addressed in this study:

1. What perceptual descriptors can be used to describe off-center sound degradation?
2. What is the individual contribution of the three geometrical factors to this percept?
3. Does the musical material affect this sound degradation?
4. Do listening room acoustics affect off-center sound degradation?

First, question 1 is explored in a qualitative study (Section 7.3) to provide the research groundwork for approaching questions 2–4, which are investigated in a quantitative experiment as described in Section 7.4.

7.2 Methodology

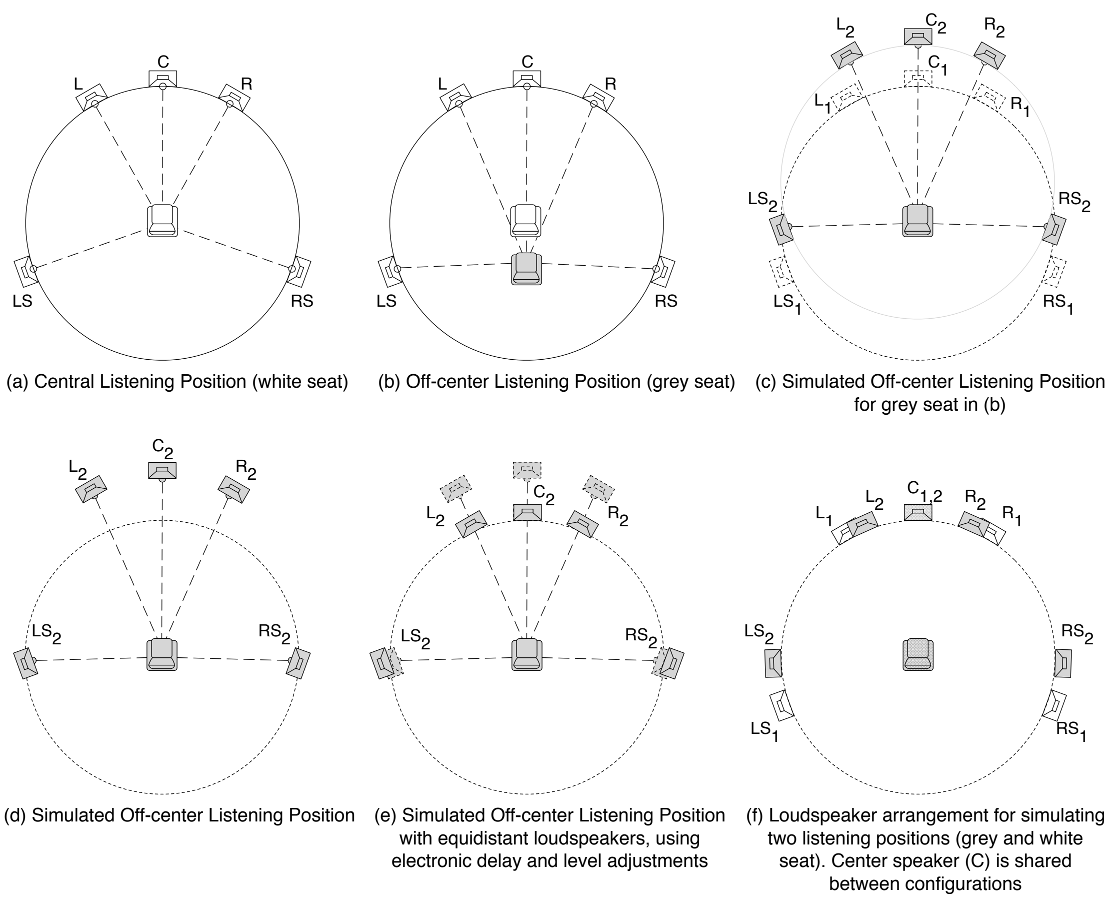
To test different listening positions in a double-blind listening experiment, an apparatus to simulate different surround-sound off-center listening positions was developed by the first author.

7.2.1 Simulation Method

Instead of using the same experimental apparatus from the experiments of Chapter 6, we propose a new method that also allows for virtually repositioning a listener in a double-blind listening test, but avoids the challenges of the binaural recording and reproduction process described in Section 6.2.5.

This simulation method is based on geometrical principles and superposition: instead of placing a listener at different OCPs in a loudspeaker setup, the entire loudspeaker setup is moved to achieve a similar geometrical listener-loudspeaker relationship. Figure 7.1 illustrates this concept within an ITU 5-channel loudspeaker setup. Figure 7.1(a) shows the Central Listening Position (CLP), whereas Figure 7.1(b) adds an off-center position (grey seat). Figure 7.1(c) shows a shifted loudspeaker setup to create the same off-center loudspeaker-listener relationship as in Figure 7.1(b), while keeping the listener at the same physical position as in Figure 7.1(a). By having these two loudspeaker setups simultaneously arranged as in Figure 7.1(c) and routing the 5-channel audio material to one or the other setup, one can virtually switch between the two associated listening positions. To compare multiple positions, multiple loudspeaker setups need to be arranged using the same principle.

These multiple loudspeaker setups can be simplified as illustrated in Figure 7.1(d,e): the physical time-of-flight and level differences in the loudspeaker-listener paths (see Section 6.1.2) can be simulated through electronic delays and level adjustments of the loudspeaker signals while positioning all loudspeakers equidistantly from the subject's seat. The individual loudspeaker directions for each tested off-center position cannot be simplified by electronic means. Therefore the direction in the simplified loudspeaker setup needs to be maintained as shown in Figure 7.1(f). This arrangement also circumvents acoustical shadowing effects amongst the loudspeakers (e.g., loudspeakers C_1 and C_2 in Figure 7.1(c)). Further, in the case of loudspeaker C_1, the direction of different virtual loudspeaker positions is coherent with loudspeaker C_2, so that the same physical loudspeaker can be used to simulate different listening positions (see Figure 7.1(f)). As a consequence of using



(a) Central Listening Position (white seat) (b) Off-center Listening Position (grey seat) (c) Simulated Off-center Listening Position for grey seat in (b)

(d) Simulated Off-center Listening Position (e) Simulated Off-center Listening Position with equidistant loudspeakers, using electronic delay and level adjustments (f) Loudspeaker arrangement for simulating two listening positions (grey and white seat). Center speaker (C) is shared between configurations

Figure 7.1: Principles of the algorithm that simulates different listening positions in a ITU 5-channel surround loudspeaker setup.

electronic delay and level adjustments, this simulation method decouples time-of-flight differences, level differences, and the direction of the arriving wavefronts from each other. This makes it possible to investigate the individual contribution of each factor individually. For instance, one can simulate the OCP of Figure 7.1(b) without the naturally occurring time-of-flight differences between the loudspeakers. This method relies on the assumption that all used loudspeakers are technically equal. For the experiments, 20 Genelec 8020A loudspeakers were deployed and calibrated for equal level and phase relations in order to simulate the five listening positions as depicted in Figure 7.2.

This simulation method is related to that of Merchel and Groth (2009), but in a slightly different context. Merchel and Groth proposed a similar method to virtually reposition the CLP of a stereophonic system according to a single moving listener but without compensating the direction from where the two wavefronts arrive.

Loudspeaker Directivity

Typical loudspeakers have an angular directional dependence of radiation that increases at high frequencies. The loudspeaker directivity has often been found to be a factor in the perceived quality in sound reproduction systems. Further, the sound features that were discovered to predict successfully the behavioral data in Chapter 6.5.2 (spectral and spatial alterations in the higher frequency range) could be related to loudspeaker directivity. However, preferences and scientific viewpoints on loudspeaker directivity vary considerably among researchers, loudspeaker designers, and audiophiles. Consequently, the frequency-dependent loudspeaker directivity also varies accordingly across loudspeaker brands and models. For OCPs, loudspeakers are off-axis with respect to the listener, and this radiation pattern becomes relevant[1].

[1] Bauer (1960) mentioned that in a stereophonic setup, the level difference between the loudspeakers in a listening point is dependent on the room characteristics and the directivity of loudspeakers.

The study of different loudspeaker directivity patterns is beyond the scope of this particular study[2]. Therefore to keep the experimental complexity reasonable, the loudspeaker directivity is a fixed controlled variable and considered to be equal for all listening angles. Practically, all loudspeakers face the subject's listening position (Figure 7.1(f)).

7.2.2 Musical Excerpts - EXC

To investigate whether the musical excerpt affect the sound degradation (research question 3), three musical excerpts were chosen to reflect a variety of complex musical materials and differ in the ratio of transient to sustained sound components, known to be an important factor for sound localization due to the precedence effect (see Section 6.1.3).

EXC 1 - Solo Piano

A 5-channel surround recording of J.S. Bach's *Goldberg Variation No.13*, recorded with a Fukada tree setup. This is the same musical material that was used in Chapter 6. In a controlled double-blind comparison of six different surround microphone techniques by Kim et al. (2006), this Fukada tree version was most preferred. For further details about the recording process, including the microphone configuration, see Kim et al. (2006).

EXC 2 - Contemporary Percussion Ensemble

A 5-channel surround recording of Roger Reynold's *Sanctuary*, 2nd movement, provided by the composer. The excerpt includes cymbals, chimes and a circular trajectory of a drum sequence, and finishes with a long reverb tail.

[2] The influence of loudspeaker directivity should be subject of a follow-up experiments. To simulate frequency-dependent loudspeaker directivities of the direct sound component, the described simulation algorithm is already capable to filter the loudspeaker feeds as a function of the simulated listening position.

EXC 3 - Classical Guitar Ensemble plus Female Singer

A dry multitrack recording of a tango quartet[3]. Because each instrument is recorded on a separate track, the ViMiC software (Peters et al. 2008; Braasch et al. 2008) could be used to create a spatialized auditory scene, using a Fukada tree virtual microphone setting. To legitimize the quality of the final mix, Grammy Award winner and McGill Sound Recording Professor Richard King fine-tuned and approved the mix.

7.2.3 Listening Position

As depicted in Figure 7.2, five listening positions were tested, including the CLP. The positions *1mF* and *1mB* are one meter in front and in back of the CLP, respectively.

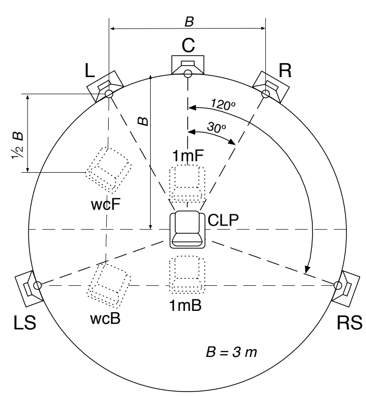


Figure 7.2: Listening positions.

	Position	Piano	Percussion	Tango
dry room	CLP	69.7	70.1	70.1
	wcF	71.9	70.0	72.0
	wcB	68.9	70.3	66.0
	1mF	73.2	70.3	73.0
	1mB	67.3	67.9	67.6
wet room	CLP	70.5	70.7	70.0
	wcF	72.4	69.4	72.0
	wcB	71.8	69.7	66.4
	1mF	74.0	70.1	73.0
	1mB	68.2	68.0	66.5

Table 7.1: Level at listening positions.

The positions *wcF* and *wcB* are further away from the CLP. They were chosen because according to ITU BS.1116-1 (ITU 1997), they are the worst-case listening positions. Note that these two seats were rotated, so that the center loudspeaker appears from the front. This rotation of off-center seats can be also found in cinemas, theatres and modern concert halls (Figure 7.3) and is therefore an ecologically valid adaptation for off-center listeners.

[3]The recording is available at http://www.csl.sony.fr/MusicSpace/MSDemos.rar, accessed May 2010.

Table 7.2: Time-of-flight delays, level differences, and direction of the arriving wavefronts of the five tested listening positions. Note that the time-of-flight delay values are relative to the first arriving wavefront.

Position	Time-of-flight [ms]					Level [dB]					Speaker direction [°]				
	L	C	R	Ls	Rs	L	C	R	Ls	Rs	L	C	R	Ls	Rs
CLP	.00	.00	.00	.00	.00	-9.54	-9.54	-9.54	-9.54	-9.54	-30.0	0.0	30.0	-120.0	120.0
1mF	.56	.00	.56	4.67	4.67	-6.37	-6.02	-6.37	-11.14	-11.14	-43.2	0.0	43.2	-133.9	133.9
1mB	3.64	3.94	3.64	.00	.00	-11.82	-12.04	-11.82	-8.45	-8.45	-22.6	0.0	22.6	-100.9	100.9
wcF	.00	2.68	5.39	3.84	9.75	-3.52	-7.69	-10.51	-9.01	-13.72	-38.3	0.0	25.2	80.5	59.0
wcB	8.73	10.60	11.58	.00	8.73	-12.25	-13.52	-14.12	0.81	-12.25	-18.4	0.0	17.8	-108.4	71.6

The level of all three musical excerpts was calibrated to 70 dBA for the CLP using a B&K 2250 sound level meter. The level across simulated listening positions varied slightly across EXC (Table 7.1) due to the individual signal distribution to the loudspeakers as part of the 5-channel surround recording, the simulated Inter-Channel Level Differences (Table 7.2) and the audio content. Although in many classic psychoacoustic experiments, the loudness of the stimuli is kept constant, it was decided not to compensate for loudness differences across listening positions, because a sound-pressure-level change is part of the off-center listening experience. Instead, perceived loudness differences should be reported/rated by the subjects and used for analyzing the results.

(a) Philharmonie of Cologne, Germany (b) Walt Disney Concert Hall, Los Angeles, US

Figure 7.3: Examples of rotated off-center seats in modern concert halls

7.3 Exploratory Qualitative Study

The goal of this qualitative study was to identify the perceptual descriptors that are commonly used for describing off-center sound degradation. A literature review on this topic suggested that descriptors for off-center sound degradation have not been elicted in a formal qualitative listening experiment before.

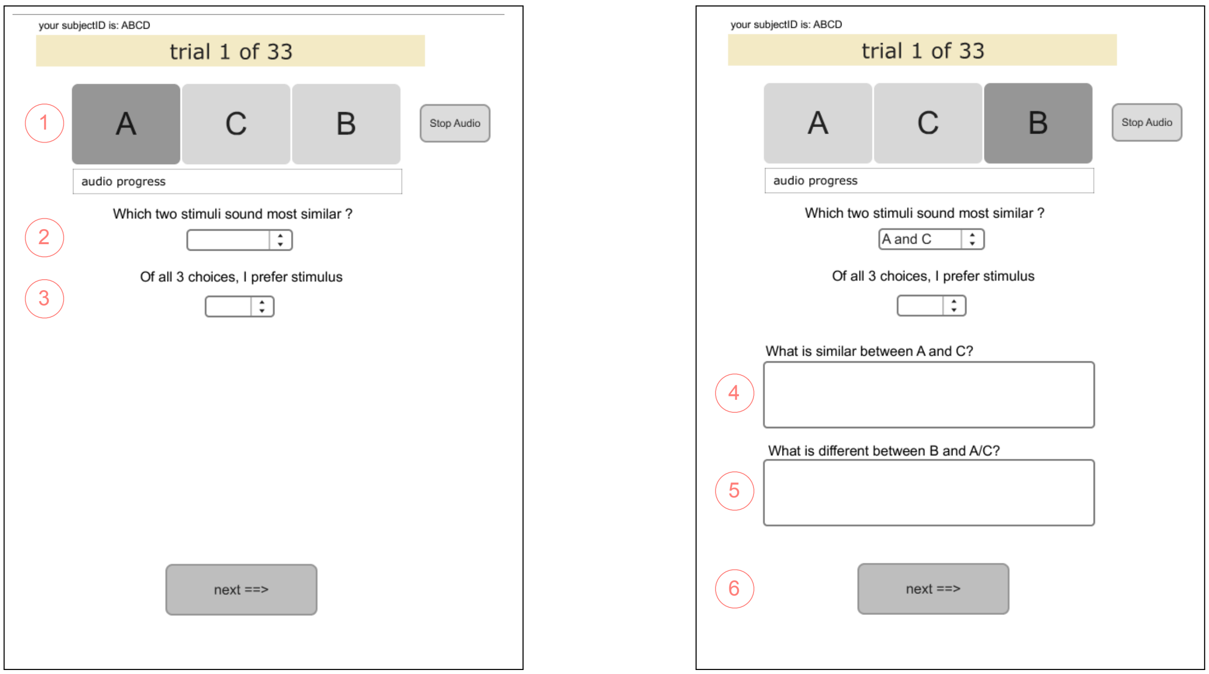
7.3.1 Procedure and Apparatus

In a double-blind listening experiment, 33 triplets of sounds were presented to subjects. These triplets consisted of one randomly selected musical excerpt (Section 7.2.2) reproduced at three out of the five tested listening positions (Section 7.2.3). The subjects worked with the graphical user interface shown in Figure 7.4.

All excerpts lasted about 15 seconds, and participants were able to play them as often as necessary during each trial. To help participants focus on the acoustical sensation and to avoid potential conflicting visual cues, room lighting was dimmed and the green status LEDs of all of the Genelec 8020A loudspeakers were covered. Breaks were allowed at any time. The experimental phase took approximately 60 minutes and was carried out in CIRMMT's Spatial Audio Lab, a hemi-anechoic room (see Figure 7.7(a) and Table 7.5 for reverberation times), thus minimizing the acoustical influence of the room.

Inspired by the Repertory Grid Technique (RGT, Berg and Rumsey 2006), the subject's task was the following: After listening to the three stimuli (Sound A, Sound B, Sound C), subjects chose from the pop-up menu (③ in Fig.7.4(a)) the most preferred sound and from the menu (② in Figure 7.4(a)) the two sounds that were most similar. Then, two text fields appeared, as displayed in Figure 7.4(b), and the participant had to report how the paired sounds were similar (④) and how they differed from the remaining (third) sound (⑤).

Participants were instructed to focus on all dimensions of the sound reproduction quality. For the preference task, participants were asked to imagine a listening condition over an extended period of time at home over loudspeakers.



(a) Phase 1 (b) Phase 2

Figure 7.4: GUI of the exploratory qualitative study.

7.3.2 Subjects

Ten native English-speaking subjects (9 male), with ages ranging from 24 to 34 (median 29.5) took part in this paid experiment. Self-classified as professional musician (2), amateur musician (7) and music lover (1), half of the subjects were McGill Sound Recording students, with formal ear training experience and can therefore be considered as trained listeners. All subjects had to pass a pure-tone ISO 389-8 (2004) audiometric hearing test before the experiment.

7.3.3 Analysis and Results

In total, 660 responses from 10 subjects were analyzed and categorized using the constant comparison method of analysis (Glaser and Strauss 1967) and facilitated by the SPSS Text Analysis for Surveys software: In an iterative process, response categories were created by the first author according to his interpretation of the responses, until a reasonable number of unique categories were achieved. Multiple response categories per single response were possible. As an example, the participant's response "the vocal *[sic]* is 'wider' and seems 'closer' like a giant head is singing to me, and again it is slightly left." was classified with the categories: *Spread and Width of Sound Source*; *Distance and Depth of Sound Sources*; *Artificial vs. Natural Impression*, and *Sound Source Position*. The category *Artificial vs. Natural Impression* was discarded, because the phrases associated with this category vary considerably from subject to subject and are therefore difficult to generalize. Because analyzing verbal data bears the risk of misinterpretation by the researcher, in some cases, the participants were contacted for clarifications.

The eight most frequent categories of sound attributes suggested by the analysis are:

1. Sound Source Position
2. Distance and Depth of Sound Sources
3. Reverberation, Perceived Room Sound (Ambient, Envelopment)
4. Spread and Width of Sound Sources
5. Timbre, Sound Coloration
6. Clarity, Diffusion, and Phasing
7. Spatial Coverage of the Scene
8. Loudness

Coding examples for the eight most frequent categories are listed in Appendix D. The other, less addressed categories are: *Artificial vs. Natural*; *Quality of movements and trajectories*; *Balanced Image*; *Spaciousness*; *Focus*; and *Degree of Immersiveness*. Figure 7.5 shows the frequency distribution of the responses according to the musical excerpts and the individual contributions for naive and trained listeners. In the mandatory post-experimental questionnaire, each participant had to describe what sound parameters were important when choosing the preferred sound. These answers were reviewed after finishing the response analysis and show a great overlap with the main categories, thus confirming the correctness of the identified descriptors from Figure 7.5. With a small number of participants (n=10), the binary preference choice data have insufficiently low statistical power. Further analysis on those data was therefore not carried out, also because the following quantitative experiment (see Section 7.4) focuses more on this aspect.

7.3.4 Discussion and Further Steps

This purpose of this qualitative exploratory study was to inform the first research question of this study (see Page 160) in examining how people describe perceived differences in the sound image due to a (simulated) change in the listening position. The results give insight into the multidimensionality of perceived off-center sound degradation. One can see that responses differ across musical excerpts, suggesting that the presented musical content influences the perception of artifacts and that the identified categories may differ for other stimuli. Therefore, the same stimuli are used for the following quantitative experiment discussed in Section 7.4.

Although a larger population would provide further data and knowledge, through the analysis of the responses in relation to the three stimuli presented, this exploratory study served the intended purposes of providing a lexical framework of sound attributes describing

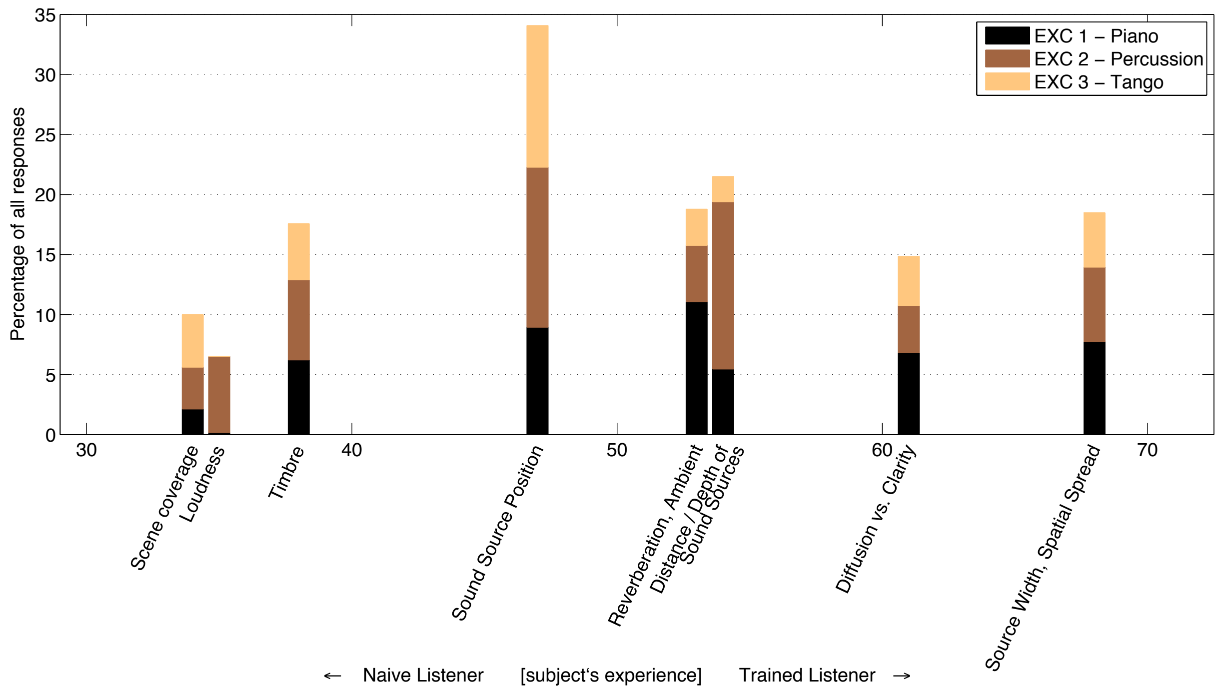


Figure 7.5: Exploratory qualitative study: The eight main categories that were identified for perceived sound degradation at off-center listening positions. Color-coding in the bars represent the contribution of each tested musical excerpt. The position of the category on the horizontal axis indicates the percentage of trained listeners (Sound Recording students) producing responses in each category. The further to the left the categories, the greater the percentage of naive listener responses. The responses do not add up to 100% due to the possibility to address multiple categories in a single response.

off-center sound degradation.

These results were used to design the quantitative listening experiment, taking into consideration that listening experiments have shown that when people have to rate a sound stimulus on several attributes simultaneously, ratings are often highly correlated. For instance, comparing Wave Field Synthesis and Higher Order Ambisonics systems, Frank (2009) used nine timbral and spatial attributes, which could be later reduced to four principle components. The little knowledge thus gained from multiple rated sound attributes does not warrant the increased experimentation time and potential confusion on the part of the subjects. For the quantitative experiment, categories were therefore merged into the four sound attributes, shown to be the most comprehensible and meaningful for naive listeners:

1. *Sound Source Position*, the spatial perception of the sound sources/instruments, including their position, distance, source width, and their relation to each other (ensemble width).
2. *Reverb*, the perception of the reflections, reverberation, room impression, envelopment within the recording.
3. *Loudness*
4. *Timbre*, the auditory sensation that allows a listener to discriminate two sounds when they are equated for pitch, loudness and duration, and when they have the same spatial location and are produced in environments with the same reverberant properties, following the definition by McAdams and Giordano (2009).

These four final attributes are also in consensus with other related studies; Zielinski et al. (2005) used three parameters (Timbre; Frontal spatial fidelity; Surround spatial fidelity) to analyze the Basic Audio Quality (BAQ), an overall measure of sound quality in 5.1 recordings.

7.4 Quantitative Experiment

In this quantitative experiment, the four sound attributes derived from the previous qualitative study were used as dependent variables in a pairwise rating experiment. Here, the primary goal was to investigate how the three factors that characterize an off-center listening position (time-of-arrival differences, sound-pressure-level differences between the signal feeds, direction of the arriving wavefronts), contribute to the sound degradation under two different room conditions.

7.4.1 Procedure and Apparatus

The simulation method described in Section 7.2.1 was also used here. In a double-blind listening experiment, 135 pairs of sounds were presented to subjects. A pair of sounds (Sound A and Sound B) consisted of one of the three randomly chosen musical excerpts (EXC, Section 7.2.2) reproduced at two of the simulated listening positions (POS, Table 7.4).

Prior to the experimental phase, subjects were trained to operate the computer interface as described below. Also, a representative collection of twelve stimuli (3 EXC $\cdot$ 4 POS) were presented to familiarize subjects with the musical material. They were told that these stimuli would give them the full range of sound quality variation so they could subsequently use the entire scale for their judgments in the experimental phase.

After listening to a pair of sounds, the subjects rated similarity in terms of *Timbre, Loudness, Position*, and *Reverb* on separate scales (① – ④ in Figure 7.6). These attributes reflect the main categories identified in the previous exploratory study (Section 7.3).

- *Timbre*, the overall similarity in timbre between the stimuli (aka *Sound Coloration*).
- *Loudness*, the overall similarity in loudness between the stimuli.
- *Positioning*, all spatial differences between the stimuli that are directly related to the

sound sources (e.g., position, distance, source width).

- *Reverb*, all spatial differences between the stimuli that are related to the reverberation and the room impression.

The responses were given on continuous scales, where 0 corresponds to the bottom (very dissimilar) and 100 to the top of the scale (very similar). No additional labels were attached. The order of the four scales on the GUI was randomized per trial to reduce order effects between the attribute ratings.

Further, subjects also had to provide a preference judgment between these two sounds on a six-point scale, ranging from a strong preference for Sound A to a strong preference for Sound B (⑤ in Figure 7.6). The context for the preference rating was explicitly provided: the most appreciated stimulus should be chosen, according to individual subject's standard and based on imagining listening to these excerpts for an extended period of time at home over loudspeakers. Deliberately, the user interface elements differ between the similarity and preference rating task (slider vs. buttons) to heighten subject's awareness of the different concepts implied in similarity and preference ratings. Intentionally, there was no option to express neutral (no) preference, but, the participants were told that if the sliders of the four sound attributes were all set to the top of the scale, a neutral preference would be assumed. After completing the similarity- and preference ratings, the subjects could proceed to the next trial.

Subjects were allowed to listen to the pair of sounds as often as necessary, triggered by pressing the space bar of the computer keyboard. The rating task was executed using the mouse, whereas the audio playback was controlled via the computer keyboard. Subjects reported the interface operation to be convenient and efficient. To minimize fatigue, participants were encouraged to take breaks whenever needed. Furthermore, after 45 and 90 trials, a reminder for taking a break appeared on the computer screen. The duration of

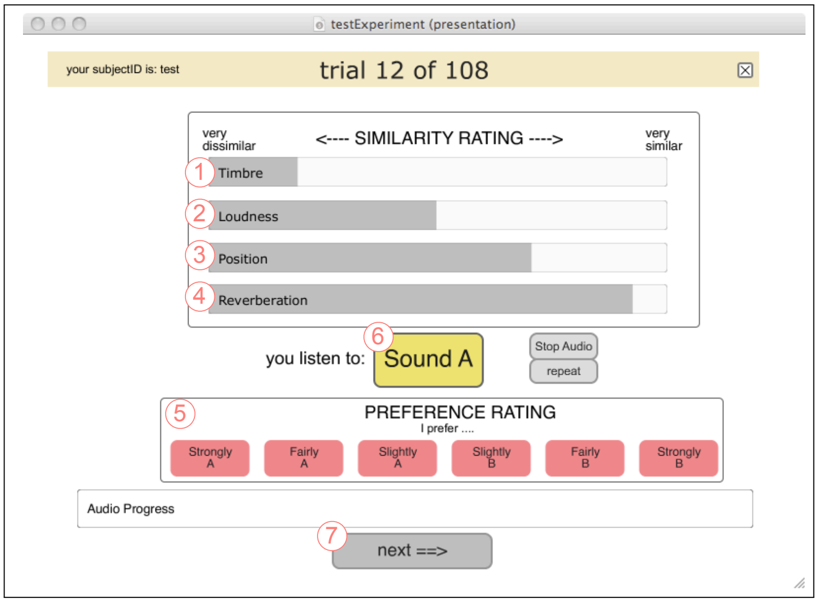


Figure 7.6: GUI of the quantitative listening experiment.

the experimental phase varied greatly across participants, between 45 minutes and 2 hours. Lighting was dimmed to help participants focus on the acoustical sensation.

To test the influence of the room acoustics on the off-center sound degradation, all subjects performed the experiment in two different conditions of listening-room acoustics, as further described in Section 7.4.4, thus resulting in an overall experiment of two listening sessions. All independent and dependent variables are summarized in Table 7.3.

Table 7.3: Experimental variables.

Independent Variable Name & Abbreviation		Described in Section	Levels	Dependent Variable	Described in Section	Scale Type
Within-Subject						
Listening Positions	POS	7.4.2	45	Loudness	7.4.1	100-point continuous
Musical Excerpts	EXC	7.2.2	3	Positioning	7.4.1	100-point continuous
Room Condition	CON	7.4.3	2	Reverb	7.4.1	100-point continuous
Between-Subject				Timbre	7.4.1	100-point continuous
Subject Group	group	7.4.4	2	Preference	7.4.1	6-point discrete

7.4.2 Simulated Geometrical Effects at the Listening Positions - POS

The four listening positions *CLP, wcF, wcB*, and *1mB* were simulated and tested as described in Figure 7.2 and Table 7.2. Additionally, for the positions *wcF* and *wcB*, all three geometric factors were uncoupled from each other and tested individually. All tested combinations are shown in Table 7.4. For instance, the level differences between the loudspeakers of the position *wcF* were tested while maintaining the Time-of-flight differences and the Direction of the arriving wavefronts from the ideal CLP (Stimulus No. 6 in Table 7.2). Consequently, each of the three factors can be observed in isolation to determine to which extent each factor causes a degradation in sound quality (Research Question 2, see Page 160).

Table 7.4: Simulated listening positions.
'*Off*' indicates that the geometrical effect is similar to what one would perceive at the CLP. '**On**' indicates that the simulation applies the parameter values as reported in Table 7.2.

Stimuli No.	ID	Position	Time-of-Flight differences	Level differences	Direction of arriving wavefronts
1	CLP	CLP	**on**	**on**	**on**
2	1mB	1mB	**on**	**on**	**on**
3	wcF	wcF	**on**	**on**	**on**
4	wcF(dir)	wcF	*off*	*off*	**on**
5	wcF(gain)	wcF	*off*	**on**	*off*
6	wcF(dly)	wcF	**on**	*off*	*off*
7	wcB	wcB	**on**	**on**	**on**
8	wcB(dir)	wcB	*off*	*off*	**on**
9	wcB(gain)	wcB	*off*	**on**	*off*
10	wcB(dly)	wcB	**on**	*off*	*off*

7.4.3 Room Acoustics Condition - CON

CIRMMT's Spatial Audio Lab (114 m^3), where the experiments took place, is by default equipped with RPG ProFoam Wedge 14" absorbers which can be completely removed or replaced with other surface materials. To investigate the influence of the listening room acoustics on the off-center listening perception (research question 4, see Page 160), the lab was used in two configurations:

- **Dry Room Condition:** The absorbent setting, using RPG ProFoam Wedge 14” absorbers (Figure 7.7(a)).
- **Wet Room Condition:** The diffuse setting, replacing all wall absorbers with RPG Skyline 4”, 8”, and 12” diffusors in a random order (Figure 7.7(b)).

We hypothesized that a more reverberant room condition would perceptually reduce off-center sound degradation, as suggested from the data of Chapter 6. The resulting reverberation times of the two room conditions are shown in Table 7.5. For the Wet Room Condition, rather than using hard reflecting surfaces[4], we intentionally used diffusor panels to creating diffuse reflection pattern while avoiding strong, discrete early reflections.

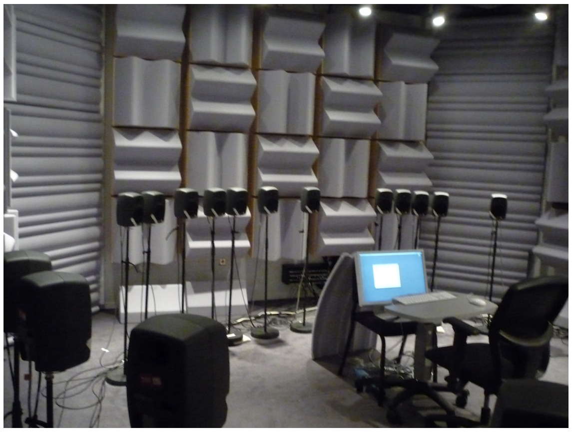
Table 7.5: Reverberation times T_{60} for the different room conditions.

T_{60}	[Hz]	125	250	500	1k	2k	4k	8k
Dry	[sec]	0.17	0.12	0.07	0.02	< 0.01	< 0.01	< 0.01
Wet	[sec]	0.28	0.17	0.17	0.18	0.24	0.22	0.36
ITU-R 1116-1	[sec]	0.24	0.24	0.24	0.24	0.24	0.24	0.24
recommendation		± 1.50	± 0.05	± 0.05	± 0.05	± 0.05	± 0.10	± 0.10

7.4.4 Participants and Subject Groups

In this paid study, 30 subjects were recruited through university-run email newsletters. None of them had participated in the previous experiment. Prior to the experiment, a pure-tone audiometric test was used to verify the hearing of all subjects (ISO 389-8 2004). To test the influence of room conditions and to counterbalance possible carry-over effects, two groups of 15 participants each were randomly formed: Group A (5 female), aged

[4] Hard reflecting surfaces increase the reverberation time, see Sabine’s formula (Kuttruff 2009, 131).



(a) Absorbent setting (dry)

(b) Diffuse setting (wet)

Figure 7.7: CIRMMT's Spatial Audio Lab setup.

between 22 and 54 years (median 26), was tested in the Dry Room Condition first and on the Wet Room Condition a week later. Group B (7 females), aged between 20 and 42 years (median 28), was first tested with the Wet Room Condition and then with the Dry Room Condition a week later. Table 7.6 shows the self-reported musicianship of all subjects.

Table 7.6: Self-reported musicianship of the participants.

Group	Professional Musician	Amateur Musician	Non-musician "music lover"	Non-musician Non-"music lover"	Sound Recording students
A	2	9	2	2	4
B	6	7	0	2	5
Total	8	16	2	4	9

7.5 Analysis and Results

The following section describes the statistical analysis of the behavioral data, using the experimental variables summarized in Table 7.3. In Section 7.5.1, the similarity ratings of the four sound attributes Timbre, Loudness, Position, and Reverb are analyzed, and then in Section 7.5.2 the preference ratings are studied in detail. Using a multiple regression approach, in Section 7.5.3 the preference data are predicted with the four sound attributes.

7.5.1 Similarity Ratings

MANOVA

A multivariate analyses of variance (MANOVA) is an extension of the univariate analysis of variance (ANOVA) and is used when there are two or more dependent variables. The advantage of using MANOVA over multiple ANOVAs is that main effects and interactions are assessed on a combination of all dependent variables.

Two repeated-measures MANOVAs were performed on the similarity measures of the sound attributes *Timbre, Loudness, Positioning*, and *Reverb* (dependent variables in Table 7.3), first for all 135 pairwise comparisons, and second for the subset of comparisons including the central listening position (CLP). The between-subject variable *group* represents differences between Group A and Group B, including the presentation order of room conditions.

These analyses will help to investigate the extend to which each of the three geometrical factors contribute to each of the four previously defined sound attributes. Furthermore, due to the independent variables EXC and CON, we can study the effect of musical excerpts and listening room acoustics.

All Pairwise Comparisons Table 7.7 presents the multivariate results for the POS(45)×EXC(3)×CON(2)×group(2) MANOVA. These results show a significant main effect for EXC and POS, and two significant interaction effects EXC×CON×group and EXC×POS. After a closer look at the individual test results from each of the four dependent measures (univariate results, Table 7.8), the EXC×CON×group interaction is not significant ($p > .067$).

Table 7.8 also provides η_P^2, a measure of effect-size, where a higher number indicates a larger effect size. For the EXC main effect, the attributes Position and Loudness, and for POS main effect, the attributes Position, Reverb, and Loudness make the strongest contribution. The attribute Timbre contributes primarily to the significant EXC×POS interaction.

Table 7.7: MANOVA, multivariate results; Pillai's trace statistic.
Significant interactions are displayed with **bold** numbers.

Effect	df	F	p	η_P^2
EXC	8, 104	28.5	< **.001**	.69
POS	176, 4752	16.9	< **.001**	.39
CON	4, 24	0.9	.464	.13
EXC×CON	8, 104	1.0	.072	.07
EXC×POS	352, 9504	6.9	< **.001**	.20
CON×POS	176, 4752	1.1	.555	.03
CON×POS×EXC	352, 9504	.95	.771	.03
EXC×group	8, 104	1.0	.482	.07
POS×group	176, 4752	.8	.970	.03
CON×group	4, 24	.4	.808	.06
EXC×CON×group	8, 104	2.1	**.041**	.14
EXC×POS×group	352, 9504	1.0	.371	.04
CON×POS×group	176, 4752	1.0	.292	.04
CON×POS×EXC×group	352, 9504	1.0	.760	.03

A Bonferroni post-hoc multiple comparison was carried out to clarify which main effect levels are different from each other. The upper part of Table 7.9 shows the comparisons across EXC for each sound attribute. For the attributes Position and Loudness, all musical excerpts are significantly different. For the Timbre attribute, the Piano vs. Percussion comparison (EXC 1 vs. EXC 2) did not show a significant difference. Also, for the Reverb attribute, the comparison Piano vs. Tango (EXC 1 vs. EXC 3) was not significant.

CLP comparisons Because of special interest in sound degradation relative to the CLP, in the following MANOVA was only performed on the pairwise comparisons related to the CLP, leading to a POS(9)×EXC(3)×CON(2)×group(2) analysis. The multivariate

Table 7.8: Univariate results of significant multivariate effects of Table 7.7.
* indicates Greenhouse-Geisser correction for a violation of sphericity.

Effect	Measure	df	F	ϵ_{GG}	p	η_P^2
EXC	Timbre	$2, 54$	10.3	-	$< .001$	.28
	Position	$2, 54$	62.8	-	$< .001$	.70
	Reverb	$2, 54$	25.7	-	$< .001$	.49
	Loudness	$2, 54$	53.0	-	$< .001$	.66
POS	Timbre*	$44, 1188$	11.5	.215	$< .001$	.29
	Position*	$44, 1188$	49.4	.185	$< .001$	.65
	Reverb*	$44, 1188$	52.0	.148	$< .001$	.66
	Loudness*	$44, 1188$	39.2	.193	$< .001$	.59
EXC×POS	Timbre*	$88, 2376$	2.8	.170	$< .001$	.94
	Position*	$88, 2376$	7.7	.182	$< .001$	.22
	Reverb*	$88, 2376$	12.3	.149	$< .001$	.31
	Loudness*	$88, 2376$	8.9	.178	$< .001$	.25
EXC×CON×group	Timbre	$2, 54$	2.5	-	.668	.08
	Position	$2, 54$	2.7	-	.785	.09
	Reverb	$2, 54$	2.8	-	.281	.10
	Loudness	$2, 54$	0.3	-	.067	.12

Table 7.9: Bonferroni post-hoc comparison of musical excerpts. Numbers show significance values for a detected difference. **Bold numbers** indicate significance for $p < .05$, **gray background** indicate $p < .01$.

Compared EXC	Piano vs. Percussion	Piano vs. Tango	Percussion vs. Tango
All pairwise comparisons			
Timbre	1.000	**.001**	**.008**
Position	**< .001**	**< .001**	**< .001**
Reverb	**< .001**	.253	**< .001**
Loudness	**.047**	**< .001**	**< .001**
CLP comparisons			
Timbre	.453	**.003**	.051
Position	**< .001**	**< .001**	**< .001**
Reverb	**.014**	.364	**< .001**
Loudness	.063	**.014**	**.014**

results (Table 7.10) show significant effects for the two main effects POS and EXC and the interaction effect EXC×POS. Their univariate results show Table 7.11.

The lower part of Table 7.9 shows the results of a Bonferroni post-hoc test across EXC for all pairwise comparisons in the context of the CLP. Similar to the previous MANOVA, a significant difference exists between all three musical excerpts for the attribute Position.

For the other three attributes, one or more Bonferroni comparisons reveal non-significant differences across the musical excerpts. The biggest similarity between the excerpts can be found in the Piano vs. Percussion comparison (EXC 1 vs. EXC 2), where both Timbre and Loudness ratings are not significantly different.

Table 7.10: MANOVA, multivariate results of the *CLP comparison*; Pillai's trace statistic. Significant interactions are displayed with **bold** numbers.

Effect	df	F	p	η_P^2
EXC	$8, 104$	20.6	$<$ **.001**	.61
POS	$32, 864$	17.1	$<$ **.001**	.39
CON	$4, 24$	1.1	.388	.15
EXC×CON	$8, 104$	.9	.492	.06
EXC×POS	$64, 1728$	8.4	$<$ **.001**	.24
CON×POS	$32, 864$	1.0	.555	.03
CON×POS×EXC	$64, 1728$	1.0	.486	.04
EXC×group	$8, 104$	.6	.805	.04
POS×group	$32, 864$	.9	.683	.03
CON×group	$4, 24$	1.9	.143	.24
EXC×CON×group	$8, 104$	1.0	.433	.07
EXC×POS×group	$64, 1728$	1.1	.411	.04
CON×POS×group	$32, 864$	.6	.957	.02
CON×POS×EXC×group	$64, 1728$	1.2	.176	.04

Table 7.11: MANOVA of the *CLP comparison*, univariate results of significant multivariate effects. * indicates Greenhouse-Geisser Correction for violation of sphericity.

Effect	Measure	df	F	ϵ_{GG}	p	η_P^2
EXC	Timbre	$2, 54$	8.0	-	.001	.23
	Position*	$2, 54$	60.4	.82	$< .001$	.69
	Reverb	$2, 54$	12.1	-	$< .001$	.31
	Loudness	$2, 54$	27.0	-	$< .001$	.50
POS	Timbre*	$8, 216$	13.8	.56	$< .001$	.34
	Position*	$8, 216$	81.4	.54	$< .001$	.75
	Reverb*	$8, 216$	59.2	.55	$< .001$	.69
	Loudness*	$8, 216$	34.6	.35	$< .001$	.56
EXC×POS	Timbre*	$16, 432$	13.8	.49	.001	.12
	Position*	$16, 432$	12.4	.52	$< .001$	.31
	Reverb	$16, 432$	16.8	-	$< .001$	.38
	Loudness*	$16, 432$	8.7	.48	$< .001$	.24

A Bonferroni post-hoc test for the POS main effect is illustrated in Table 7.12. Here, one can compare the similarity ratings of all four sound attributes for each listening position relative to the CLP. From 36 pairwise comparisons, only three comparisons show no significant difference across all four sound attributes. These pairs are *wcF(dir)* vs. *wcB(dir)*; *wcF(dly)* vs. *1mB*; and *wcB(dly)* vs. *1mB*. In all other POS comparisons, at least one of the four sound attributes was rated significantly different (illustrated through bold number on grey background in Table 7.12). In eight POS comparisons, all four sound attributes were rated significantly different. In total, the Position attribute was rated significantly different 28 times (Loudness 23, Reverb 21, Timbre 14 times, respectively).

Table 7.12: Result of Bonferoni post-hoc test for POS main effect of the *CLP comparison*. Abbreviations of the listening positions are referring to the nomenclature introduced in Table 7.4. Numbers show significance values for a detected difference **Bold numbers** indicate significance for $p < .05$, **gray background** indicate $p < .01$.

Pair	wcF(dir)	wcF(gain)	wcF(dly)	wcB	wcB(dir)	wcB(gain)	wcB(dly)	1mB	Attribute
wcF	< **.001**	**.023**	1.000	1.000	< **.001**	1.000	.092	**.018**	Timbre
	< **.001**	< **.001**	< **.001**	**.049**	< **.001**	1.00	< **.001**	< **.001**	Position
	< **.001**	**.009**	.953	< **.001**	< **.001**	< **.001**	.505	1.000	Reverb
	< **.001**	.238	< **.001**	1.000	< **.001**	**.038**	< **.001**	< **.001**	Loudness
wcF(dir)	-	**0.001**	**.001**	< **.001**	.075	< **.001**	**.007**	< **.001**	Timbre
	-	**.019**	.060	< **.001**	1.000	< **.001**	< **.001**	**.002**	Position
	-	.868	**.016**	< **.001**	1.000	< **.001**	.414	< **.001**	Reverb
	-	< **.001**	**.005**	< **.001**	1.000	< **.001**	.768	**.001**	Loudness
wcF(gain)		-	1.000	**.008**	.568	.667	1.000	1.000	Timbre
		-	1.000	< **.001**	< **.001**	< **.001**	1.000	1.000	Position
		-	1.000	< **.001**	1.000	< **.001**	1.000	1.000	Reverb
		-	**.001**	1.0	< **.001**	.909	< **.001**	**.012**	Loudness
wcF(dly)			-	.861	.116	1.000	1.000	1.000	Timbre
			-	< **.001**	< **.001**	< **.001**	**.010**	.871	Position
			-	< **.001**	**.048**	< **.001**	1.000	1.000	Reverb
			-	< **.001**	0.160	< **.001**	1.000	1.000	Loudness
wcB				-	< **.001**	1.000	.395	**.032**	Timbre
				-	< **.001**	< **.001**	< **.001**	< **.001**	Position
				-	< **.001**	1.000	< **.001**	< **.001**	Reverb
				-	< **.001**	.067	< **.001**	**.014**	Loudness
wcB(dir)					-	**.005**	.584	.381	Timbre
					-	< **.001**	< **.001**	< **.001**	Position
					-	< **.001**	.454	**.011**	Reverb
					-	< **.001**	1.000	< **.001**	Loudness
wcB(gain)						-	1.000	.991	Timbre
						-	< **.001**	< **.001**	Position
						-	< **.001**	< **.001**	Reverb
						-	< **.001**	1.000	Loudness
wcB(dly)							-	1.000	Timbre
							-	1.000	Position
							-	1.000	Reverb
							-	.100	Loudness

Figure 7.8 illustrates the main effects POS and EXC, found to be significant in the MANOVA, as a function of POS. One can see that for the POS main effect, the sound attributes follow a similar trend and have their largest dissimilarity for the positions *wcB(gain), wcB*, and *wcF*. For the positions *wcF(dir), wcB(dir), wcB(dly)* and *1mB* the similarity ratings are on average higher than or equal to 75 points. In terms of the EXC main effect, one can see in the right subplot of Figure 7.8 that the percussion excerpt created the largest mean difference between the sound attribute ratings. Here, the rating for Position is the lowest of all four attributes and also remarkably lower compared to the other two excerpts. Among all three excerpts, the loudness attribute achieved the highest similarity ratings of all four attributes. For the piano excerpt (EXC 1) and tango excerpt (EXC 3), Reverb achieved the lowest mean ratings.

The EXC main effect is further illustrated in Figure 7.9, displaying for each sound attribute in a separate plot the mean value and 95% confidence interval of the similarity ratings as a function of the musical excerpt and listening position with respect to the CLP.

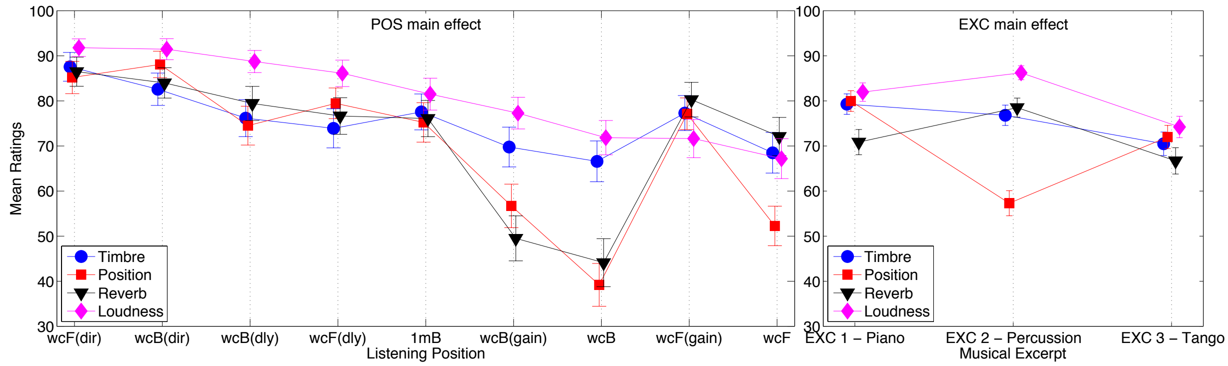


Figure 7.8: Significant POS and EXC main effects for the comparisons against the CLP. Error bars show the 95% confidence interval for the mean. Points are connected for ease of reading.

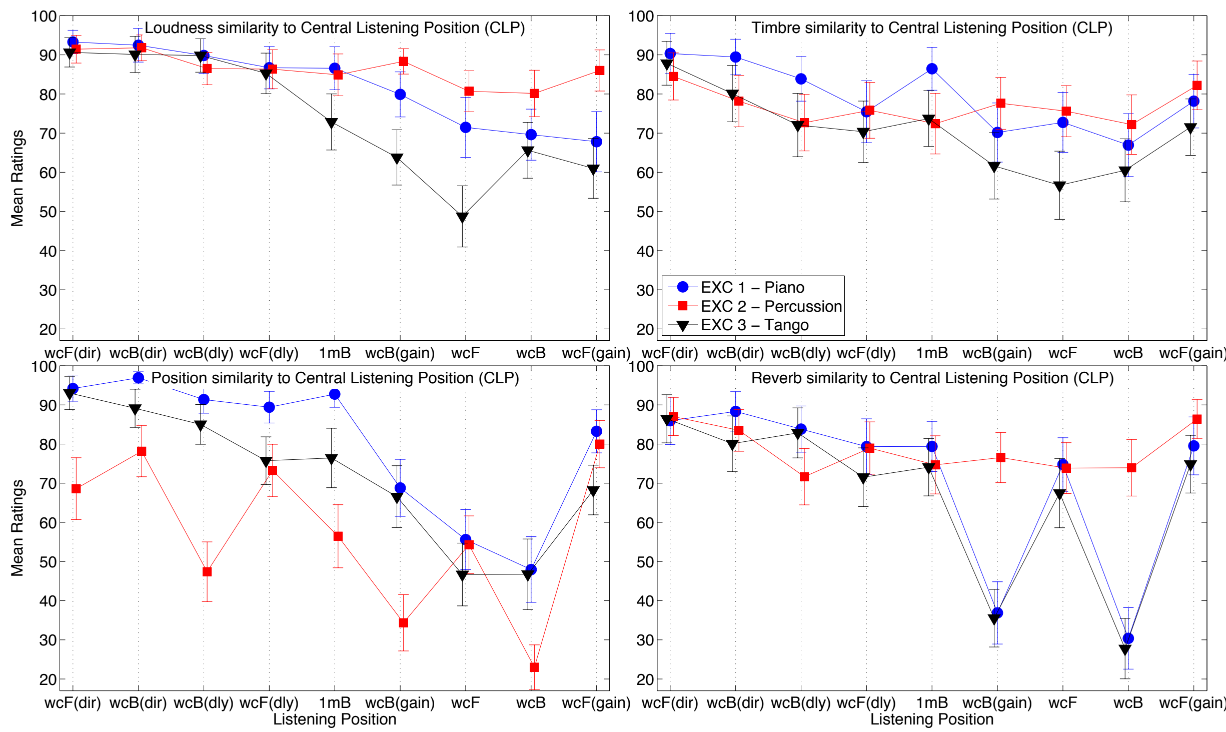


Figure 7.9: Similarity ratings of the four attributes as a function of musical excerpt and listening position for the comparisons against the CLP. Error bars show the 95% confidence interval for the mean. Points are connected for ease of reading.

Correlation of Attributes

To determine the extent to which the four sound attributes *Timbre, Position, Reverb*, and *Loudness* were independently rated, correlations across attributes were analyzed (Figure 7.10). From the similarity ratings by each of the 30 subjects in both room conditions, the Spearman rank correlation values of all attribute combinations were computed. Figure 7.10 shows how these values vary between subjects in a boxplot diagram. The vertical dashed-dotted line indicates the correlation value above which the correlation is significant ($p = .05$). On average, the correlations between all attribute pairs are significant, positive, and below $\rho(135) = .5$. The pairs associated with *Position* show the lowest correlation across all attribute combinations. The attributes *Reverb, Loudness* and *Timbre* are more highly correlated ($\rho(135) \approx .45$) and the long whiskers that mark the 25th and 75th percentiles, indicate a larger variability of those correlation values across participants, ranging to about $\rho(135) = .87$.

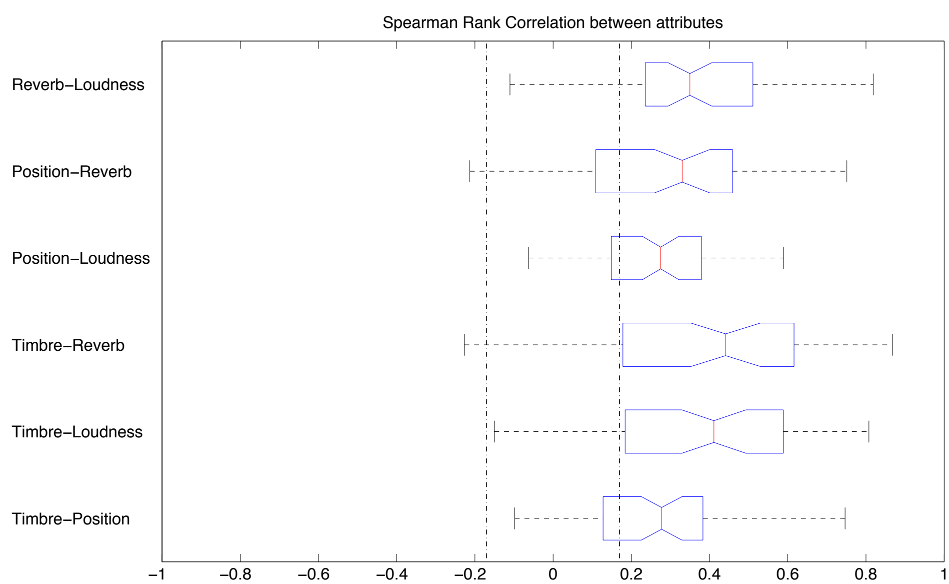


Figure 7.10: Spearman rank correlations between attributes. The boxplot shows the variability among subjects. Vertical dash-dotted lines represent the range of non-significant correlation values around zero.

Multidimensional Scaling

To explore the degree of similarity and dissimilarity of the attribute ratings visually, INDSCAL, a multidimensional scaling (MDS) technique was used. Like any other MDS techniques, INDSCAL (INividual Differences SCALing, described in detail in Carroll and Chang 1970) tries to model (dis)similarity ratings as distances in an R-dimensional space. In the spatial representation of the N stimuli, a large dissimilarity is represented by a large distance and a small dissimilarity (also, a large similarity) is represented by a small distance. As opposed to the generalized MDS technique, INDSCAL accounts for differences in the subjects' perceptions. The distance d_{ij} between stimuli i and j for subject n is calculated by Equation 7.1. Each subject has an individual space in which every dimension of the common space is weighted individually by w_{nr}.

A visual inspection of the individual space of the different INDSCAL solutions revealed a great commonality across the participants, suggesting that a (simpler) generalized MDS model (which assumes commonality between individuals) according to Equation 7.2 would be appropriate. Unlike a generalized MDS, INDSCAL produces unique orientation of the axes of the group space, which limits further comparisons amongst different INDSCAL data sets via Procrustes rotation (e.g., for comparing the similarity ratings between the three musical excerpts, see Section 7.5.1). Therefore, the presented MDS models were carried out using a generalized MDS.

$$d_{ij}^2 = \sum_{r=1}^{R} w_{nr}(x_{ir} - x_{jr})^2 \qquad (7.1)$$

$$d_{ij}^2 = \sum_{r=1}^{R} (x_{ir} - x_{jr})^2 \qquad (7.2)$$

Suggested as a rule of thumb by Kruskal and Wish (1978), the maximum number of dimensions R should be determined by the equation $I - 1 > 4 \cdot R$, where I is the number of stimuli. Because 10 stimuli were tested here, R should not be greater than 2. The MDS solution of the similarity ratings of *Timbre* and *Position* for EXC 1 (Piano) was computed in different dimensionalities and stress and fit measures are plotted as a "scree plot" in Figure 7.11. The scree-plots show a reduction of the stress measures with increasing dimensionality. Although the scree-plots do not show a highly remarkable elbow-point, a reasonably low stress measure is already achieved by a 2D solution. Therefore, only two-dimensional solutions were considered for this analysis, with stress and fit measures summarized in Table 7.13. Here, the S-Stress (a stress measure based on squared distances) for each musical excerpt is consistently highest for the *Timbre* similarity.

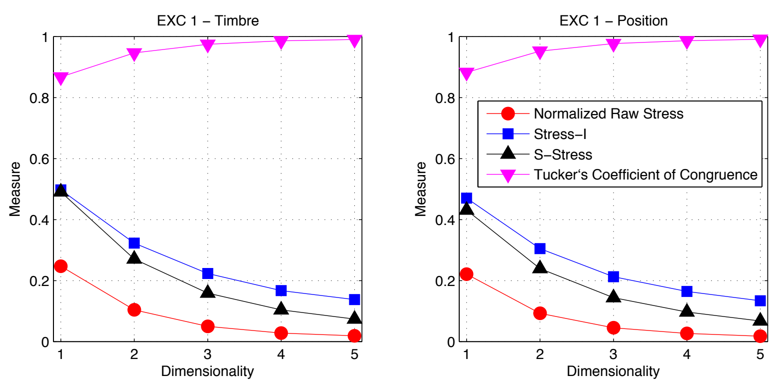


Figure 7.11: Scree plot for the similarity ratings of Timbre (left) and Position (right) of the Piano excerpt (EXC 1).

Table 7.13: Stress and Fit Measures for two-dimensional MDS solutions.
Note that the generalized MDS algorithm aims to minimize the Normalized Raw Stress

	EXC 1 - Piano				EXC 2 - Percussion				EXC 3 - Tango			
	Position	Reverb	Timbre	Loudness	Position	Reverb	Timbre	Loudness	Position	Reverb	Timbre	Loudness
Normalized Raw Stress	.093	.093	.103	.104	.099	.106	.108	.103	.096	.095	.102	.089
Stress-I	.305	.305	.322	.323	.315	.326	.329	.321	.310	.307	.320	.299
Stress-II	.846	.805	.948	.944	.908	.972	.980	.956	.879	.847	.942	.829
S-Stress	.246	.238	.280	.268	.263	.286	.287	.278	.254	.245	.270	.240
Tucker's Coefficient of Congruence	.952	.952	.947	.947	.949	.946	.945	.947	.951	.952	.947	.954

Procrustes Analysis

As a measure of similarity between the different MDS solutions, a Procrustes analysis was performed to estimate the similarity of shapes by applying linear transformations (translation, reflection, rotation and scaling). Results of a Procrustes analysis are expressed in the standardized sum-of-squared distances measure between the points of the transformed shape and the second (to-be-matched) shape. The smaller this measure, the more successful the transformation is, and therefore, the more similar the two compared shapes are.

The results of this analysis are shown in Table 7.14. Most of the sum-of-squared distance measures are high, suggesting that the MDS solutions are independent. Three of the 48 analyzed combinations result in sum-of-squared distance measures are below .2; such MDS solutions can be considered as fairly similar. These three combinations are: the Position attribute of EXC 1 and EXC 3; the Timbre attribute of EXC 3 to the Position attribute of EXC 3; the Position attribute of EXC 1 (bold numbers in Table 7.14).

Table 7.14: Procrustes analysis of similarity measures. Values near zero imply more similar shapes, while values near one imply dissimilarity. Values smaller than 0.2 are indicated in **bold numbers.**

		EXC 1 - Piano			EXC 2 - Percussion				EXC 3 - Tango			
		Reverb	Timbre	Loudness	Position	Reverb	Timbre	Loudness	Position	Reverb	Timbre	Loudness
EXC 1	Position	.77	.30	.70	.68	.69	.77	.69	**.10**	.63	**.15**	55
	Reverb		.75	.68		.63	.82	.59		.62	.68	.60
	Timbre			.74			.90	.83			.49	.70
	Loudness							.88				.59
EXC 2	Position					.91	.64	.82	.68	.83	.55	.85
	Reverb						.64	.59		.62	.70	.43
	Timbre							.80			.71	74
	Loudness											.70
EXC 3	Position									.59	**.14**	.63
	Reverb										.73	.64
	Timbre											.54

The two-dimensional MDS solutions after generalized Procrustes transformation are shown in Figure 7.12 in a subplot for each musical excerpt and in Figure 7.13 in a subplot for each sound attribute. To facilitate the readability of these figures, similar listening positions in close proximity are clustered by way of visual inspection. For the Tango excerpt (EXC 3, Figure 7.12) a clear divide between Back- (South-West sector) and Front listening position (North-East sector) exist with the CLP somehow in between. This divide seems to exist also in the MDS solution for the other two excerpts, but is not so strongly visible because the stimuli points are more scattered. The clustering in the MDS solutions in the lower subplots of Figure 7.13 for the Reverb and for the Position attribute indicates that the MDS solutions of these attributes for several listening positions are fairly similar across musical excerpts. For the Position attribute, the CLP stimuli of all three musical excerpts are in close proximity to each other and, of all listening positions, the *wcF* and *wcB* stimuli are furthest away from the CLP, indicating lowest similarity. For the Reverb attribute, remarkably, the stimuli related to *wcF* and *wcB* are both in close proximity to the related *wcF(gain)* and *wcB(gain)*, suggesting a very similar perception.

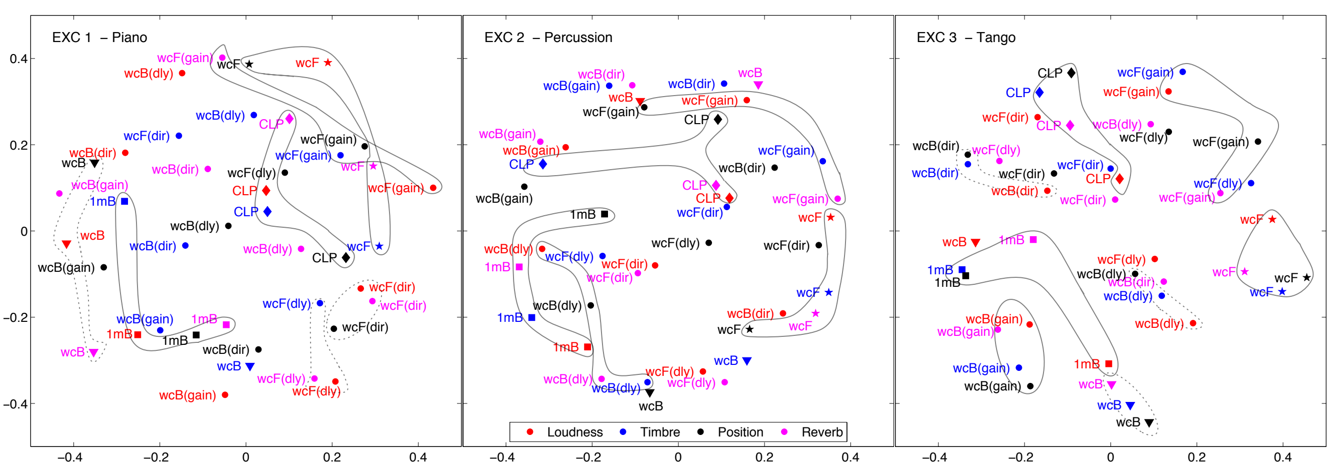


Figure 7.12: MDS solutions grouped for each EXC after generalized Procrustes transformation. Solid lines cluster listening positions with close proximity of all four attributes, dotted lines cluster listening positions with close proximity of three attributes.

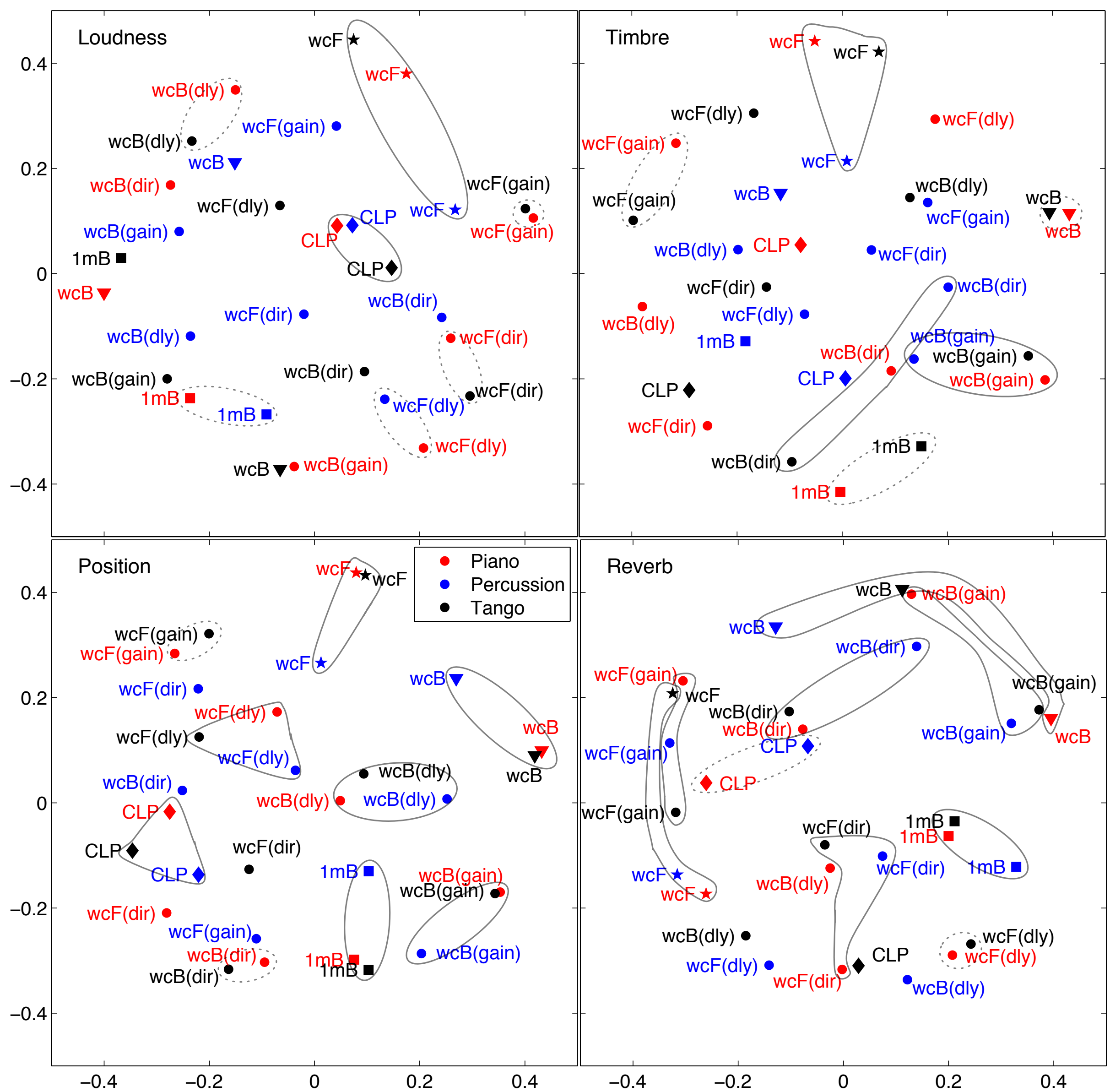


Figure 7.13: MDS solutions for each attribute per musical excerpt after generalized Procrustes transformation. Solid lines cluster listening positions with close proximity of all three musical excerpts, dotted lines cluster listening positions with close proximity of two musical excerpts.

7.5.2 Preference Ratings

The following Section will report on the analysis of the preference ratings, given on a six-point scale, ranging from a strong preference for Sound A to a strong preference of Sound B (see rating scale in Figure 7.6). This analysis enables us to investigate off-center listening perception as a function of preference, rather than as a function of similarity, as reported in the previous Section.

Analysis of Variance

An Analysis of variance (ANOVA) was performed to determine whether the listening position (POS), the musical excerpt (EXC) or the room condition (CON) and their interactions have a significant effect on the preference data. Because a test of normality on the preference data showed that the assumption of normality is violated for ca. 25% of all comparison trials, an ANOVA is inappropriate and a non-parametric test should be carried out instead. Therefore, two tests are performed:

1. A POS(45)×EXC(3)×CON(2) repeated-measures ANOVA on all preference ratings (Table 7.15)
2. Two-way non-parametric Friedman tests on all combinations of the three main effects (Table 7.16)

Both analyses show essentially the same results: The main effects CON and POS, as well as the interaction effects EXC×POS, CON×POS are significant. The main effect EXC and the interaction EXC×CON are not found to have a significant effect. The three-way interaction EXC×CON×POS was found in the ANOVA to be significant. The latter cannot be confirmed with a non-parametric method because no such test exists to the author's knowledge. However, ANOVA has been proven to be fairly robust if a normal distribution cannot be guaranteed, especially when samples are large.

Table 7.15: ANOVA on all pairwise preference ratings.
* indicates Greenhouse-Geisser correction for a violation of sphericity.

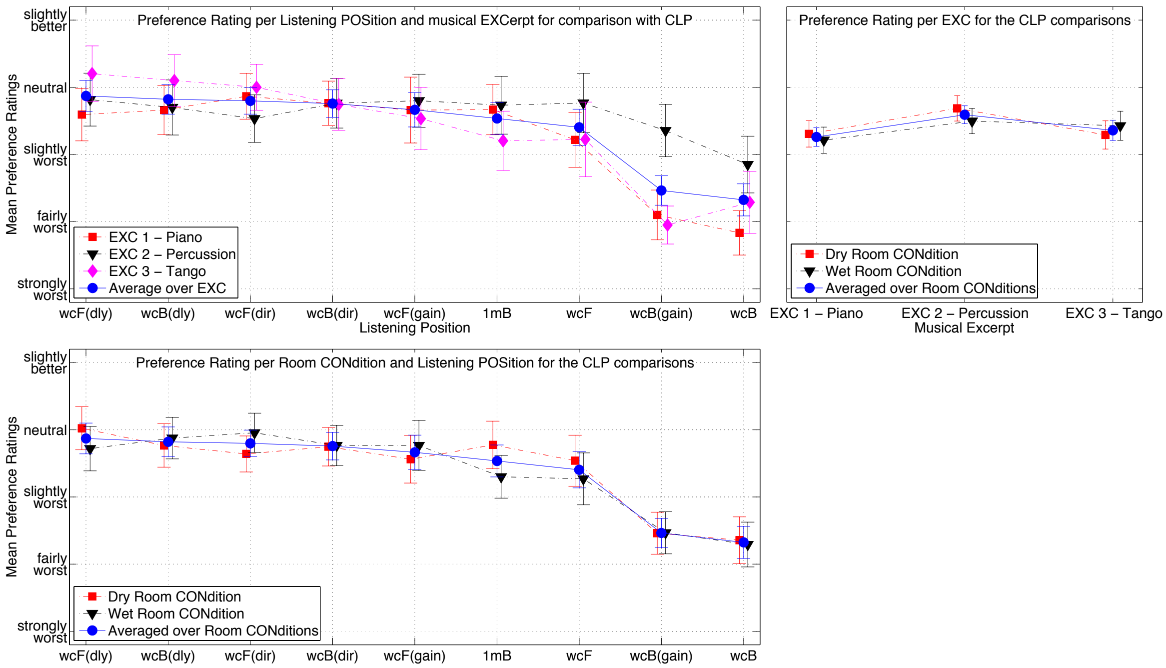
Effect	df	F	ϵ_{GG}	p	η_P^2
EXC	$2, 54$	1.25	-	.294	.04
CON	$1, 27$	25.58	-	**.001**	.32
POS*	$44, 1188$	34.09	.105	$<$ **.001**	.56
EXC×CON	$2, 54$	1.89	-	.162	.07
EXC×POS*	$88, 2376$	5.00	.172	$<$ **.001**	.16
CON×POS*	$44, 1188$	1.42	.312	**.040**	.05
EXC×CON×POS*	$88, 2376$	1.66	.166	$<$ **.001**	.06

Table 7.16: Comparing all main effects using a two-way non-parametric Friedman test.

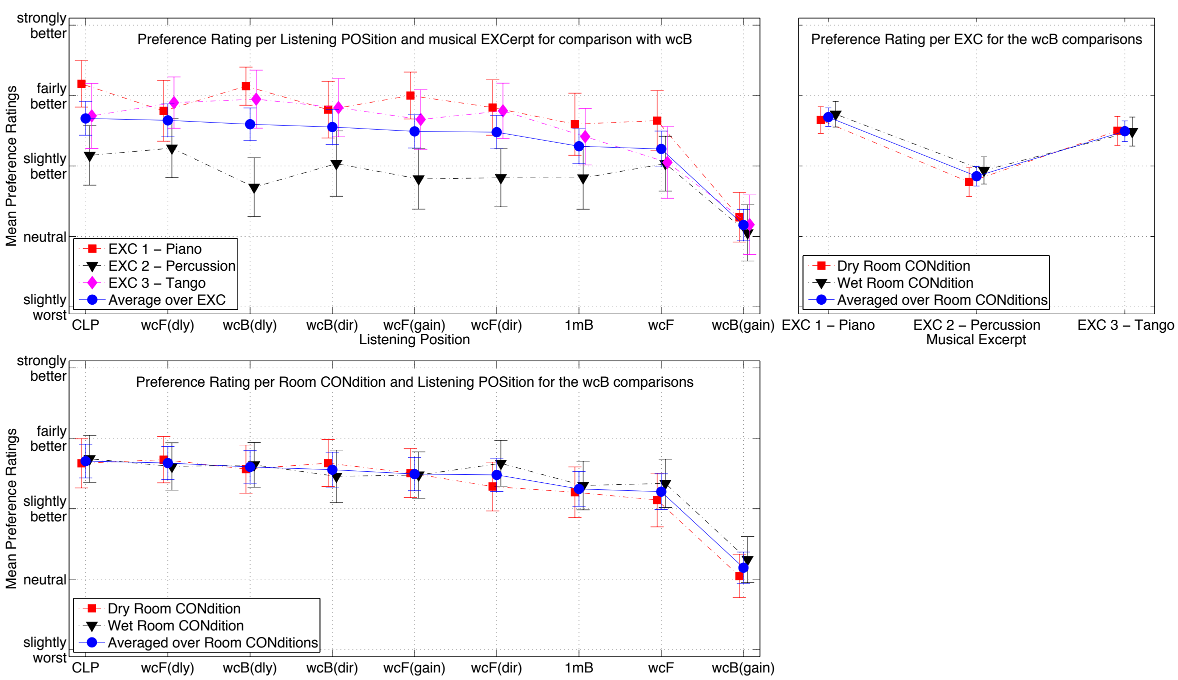
Main effect A	Main effect B	Friedman's Results	
EXC	CON	$\chi^2(2, 168) = 1.46,$	$p = .4830$
	POS	$\chi^2(2, 3780) = 5.14,$	$p = .0765$
CON	EXC	$\chi^2(1, 168) = 6.86,$	$p =$ **.0088**
	POS	$\chi^2(1, 2520) = 9.89,$	$p =$ **.0017**
POS	EXC	$\chi^2(44, 3780) = 1132.18,$	$p <$ **.0001**
	CON	$\chi^2(44, 2520) = 1038.05,$	$p <$ **.0001**

Figure 7.14 provides further insight into the preference data. Figure 7.14(a) depicts the mean preference ratings of the subset of all the preference data from comparisons to the CLP. Figure 7.14(b) shows the mean preference ratings for all pairwise comparisons with the *wcB* listening position.

The upper left subplot of both figures depicts the mean preference ratings across listening positions as a function of EXC, and confirms the significant POS main effect. Remarkably, the upper left subplot of Figure 7.14(b) indicates a difference of the mean preference for some POS caused by EXC, which was not found to be a significant main effect across all comparisons. A Friedman test on this subset of data suggests significant differences between the EXC here ($\chi^2(2, 756) = 10.98, p = .004$).



(a) Mean Preference Ratings for the comparisons with the CLP.



(b) Mean Preference Ratings for the comparisons with the wcB Position.

Figure 7.14: Mean Preferences. Error bars show 95% confidence interval.

The upper right plots of the Figures 7.14(a,b) illustrate the mean preference ratings across listening positions (POS) as a function of the room condition (CON). A significant main effect of the room condition (CON) found in the ANOVA (Table 7.15) and Friedman test (Table 7.16), is not clearly visible in Figure 7.14. However, in the upper right plot of Figure 7.14(a) for the listening position *1mB*, one finds that the room conditions create a remarkable difference in the mean preference data, which may instead indicate the two-way interaction CON×POS. Further, the right plot of both figures shows for both room conditions the same mean preference data for each musical excerpt; this confirms that the EXC×CON interaction is not significant. Here, in comparing the Figures, the right plot of Figure 7.14(b) shows that the mean preference ratings for EXC 2 (Percussion) are lower than those of the other two musical excerpts, suggesting an influence of EXC not apparent in the results of the ANOVA and Friedman tests.

Internal Preference Mapping

For analyzing the complexity of the preference data, a special technique related to MDS, but for preference data, is used. MultiDimensional PREFerence scaling (MDPREF), proposed by Chang and Carroll (1968), is an internal preference mapping procedure that uses a point-vector, or "ideal vector" model to visualize the relation of subject's preference values and stimuli in a common space. Each subject's preference is represented by a preference vector which goes through the center of origin and has the highest preference at the end. Stimulus points are also distributed in the space and the strength of a subject's preference for a specific stimulus can be predicted by its orthogonal projection onto the vector: The closer this projection point is to the tip of the preference vector, the more preferred the stimulus is. Although the sequential approach of MDPREF has been criticized by Borg and Groenen (2005, 340) and Marketo et al. (1994), it is a popular method, especially in marketing research.

For use in MDPREF, all pairwise preference ratings were converted into a first-score matrix where rows represent subjects and columns represent the 10 POS stimuli (Figure 7.15). To compensate for value differences caused by the variation in how subjects used the range of the preference scale, MDPREF was set up to subtract the row means of the first-score matrix. In the final configuration, the ideal preference points of the subjects are positioned on the unit circle, whereas the POS-stimuli points are scattered inside. The further away the POS-stimuli points are from the origin, the more variance within the preference data can be explained by them.

The results of the ANOVA in the previous section guides the MDPREF calculations. Multidimensional preference scaling was performed on each musical excerpt separately (Fig. 7.16– 7.18), combined within one analysis (Figure 7.20), and merged together (Fig. 7.19). Because the room condition was found to be significant in the previous section, it is also considered here. Each subject has an ideal preference point for the Dry and Wet Room Conditions, indicated in the Figures with differently shaped and colored markers.

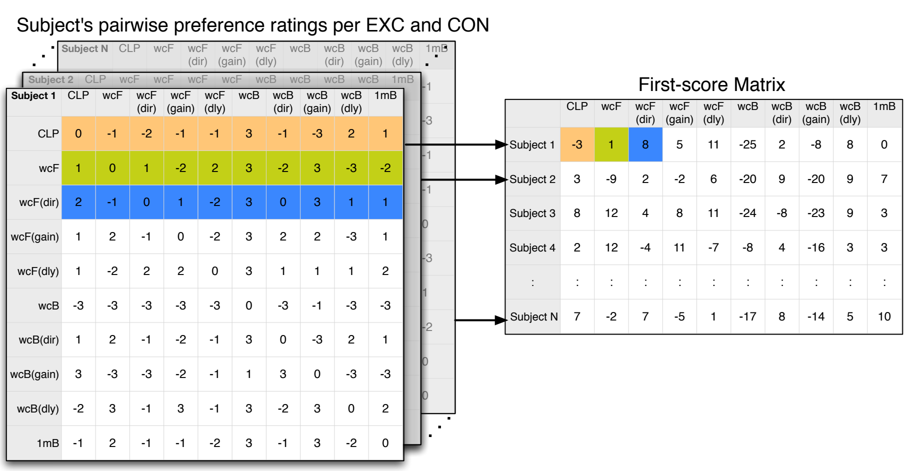


Figure 7.15: MDPREF - Transformation of the pairwise preference ratings into a first-score matrix. Colors in the matrices symbolize how each row of the subject's preference matrix (left) is summed to a value for the first-score preference matrix (right).

It was found that these individual ideal preference points between the room conditions varied greatly for some subjects. The left plots in the Figures 7.16–7.20 show the variability of the subject's preference points between the room conditions in radial angles. When one uses these distance measures and would apply the common definition of outliers, as being further away than 1.5 times the 75% percentiles, several subjects would have been discarded[5] from further analysis (see horizontal dashed line in the left subfigures).

Remarkably, the people with the highest variability were those "recruited" through a general McGill student newslist and can be considered as naive listeners. Subjects that were recruited through CIRMMT's newsletter or from other sources[6] showed less variability. Because the high number of trials did not allow for repetitions in the experimental design, one can only speculate as to the cause of the high variability. There are some reasons to believe that these variabilities are related to the subject's motivation and inter-individual differences in the degree of difficulty in performing the experimental task. Indeed, the two subjects with the highest variability across the experiment (subject A4 and B12) identified themselves in the post-experimental questionnaire as non-musicians and "non music-lovers". Further, Subject B12 also completed the experiment in the least amount of time.

The two-dimensional representation provided by MDPREF accounts for 59% of the variance in the preference ratings for EXC 2 (Percussion). More variance can be accounted for the Piano excerpt (EXC 1, 75%) and for the Tango excerpt (EXC 3, 80%). A third dimension added to the preference representation of EXC 2 improved the model by just 12%, and is not displayed here. For EXC 2 (Percussion), the differences between Wet and

[5] No subject was discarded here. The process of discarding potential outliers is debatable because it may introduce biases on behalf of the analyzing person.

[6] Others are mainly music and composition students from the Université de Montreal who heard of the experiment from their teacher.

Dry Room Conditions were the largest across all three excerpts. The preference vector for the average across subjects is depicted for the Dry and Wet Room Conditions independently, and for the average of both. The resulting mean preference ranking of the listening positions, derived from orthogonal projection onto the preferences vectors for the different Room Conditions, are summarized in Table 7.17 and in Figure 7.21. To avoid the subjective influence of outlier elimination, Table 7.17 presents the preference ratings of the average across subjects with and without potential outliers as discussed above.

For all excerpts, subjects tend to cluster around the first dimension, thus it can be said that for all three musical excerpts, there exists an overall agreement about their listening position preference. Because all subjects' preference points are distributed on a circle, the deviation of the individual preference points from the average preference point can be expressed as the standard deviation using angular degrees. The second column of Table 7.17 presents those values. The standard deviation value is largest for EXC 2, being around 65°. For EXC 1 and EXC 3 this value is about 17° smaller. Also for EXC 1 and EXC 3, the deviation value decreases remarkably from the Dry to the Wet Room Condition, whereas for EXC 2, this value does not change much. As expected, when the potential outliers are eliminated, the deviations reduce to 15° (EXC 1), 48° (EXC 2) and 37° (EXC 1).

Figure 7.20 shows a preference mapping of the listening positions calculated for all three musical excerpts together. Here one can see that the preference values for EXC 2 are closer to the center of origin compared to the preference ratings of the other two excerpts, indicating that the preference ratings between listening positions for EXC 2 were rated less clearly compared to EXC 1 and EXC 3. This can also be observed in the upper plots of Figure 7.14(b), showing that the preference ratings of EXC 2 are below those of the other two excerpts. Figure 7.20 shows a similar distributions over the first dimension (x-axis) for EXC 3 and EXC 1, while EXC 2 has more weight on the second dimension (y-axis).

Listening positions located close to each other are similarly preferred. One can clearly see in Figures 7.16–7.19 the proximity of the listening positions *wcF* to *wcF(gain)* and *wcB* to *wcB(gain)*. In contrast, the *CLP* is closely surrounded by *wcB(dir)* and *wcF(dly)* for all musical excerpts. The configurations of other POS stimuli show similarities and differences between the EXC. EXC 1 and EXC 3 show the most similarities. Remarkably, while the locations of *CLP*, *wcF* and *wcB* are very similar, the location of the position *1mB* changes across the musical excerpts.

Figure 7.19 shows the calculated preference ratings of the listening positions merged over all excerpts. The preference vector of the average subject is pointing along the negative x-axis. In terms of stimulus points, four distinct regions along the two axes can be indicated:

- Positive y-axis: *wcF* clustered with *wcF(gain)*
- Negative y-axis: 1mB
- Positive x-axis: *wcB* clustered with *wcB(gain)*
- Negative x-axis: *CLP* clustered with *wcF(dir)*, *wcF(dly)*, *wcB(dir)*, *wcB(dly)*

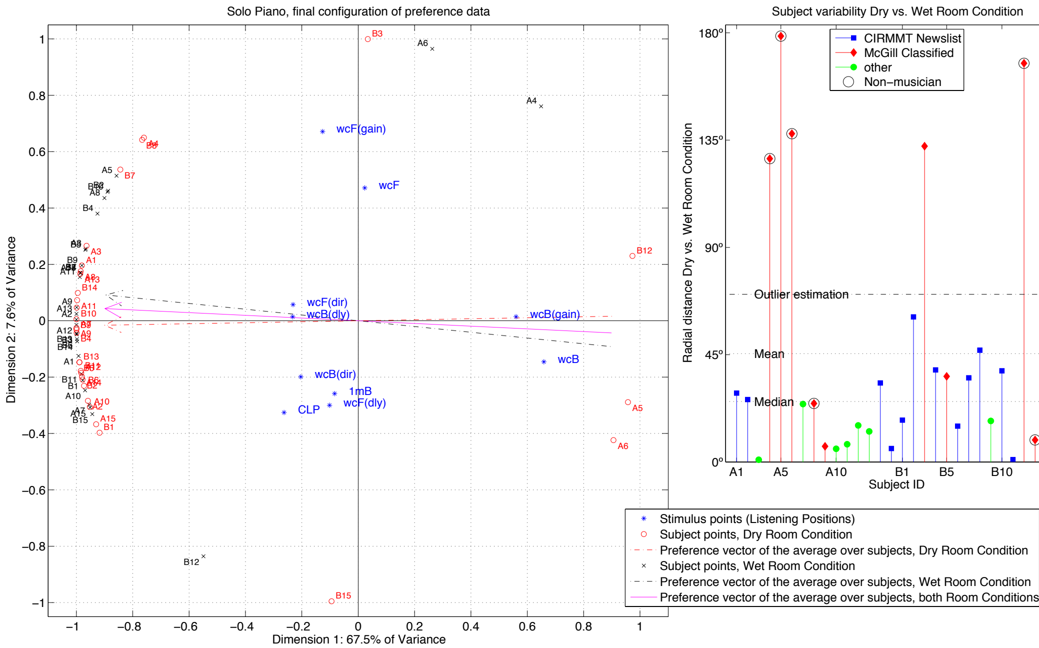


Figure 7.16: Two dimensional MDPREF solution for musical excerpt 1 - Piano, accounting for 75% of the variance.

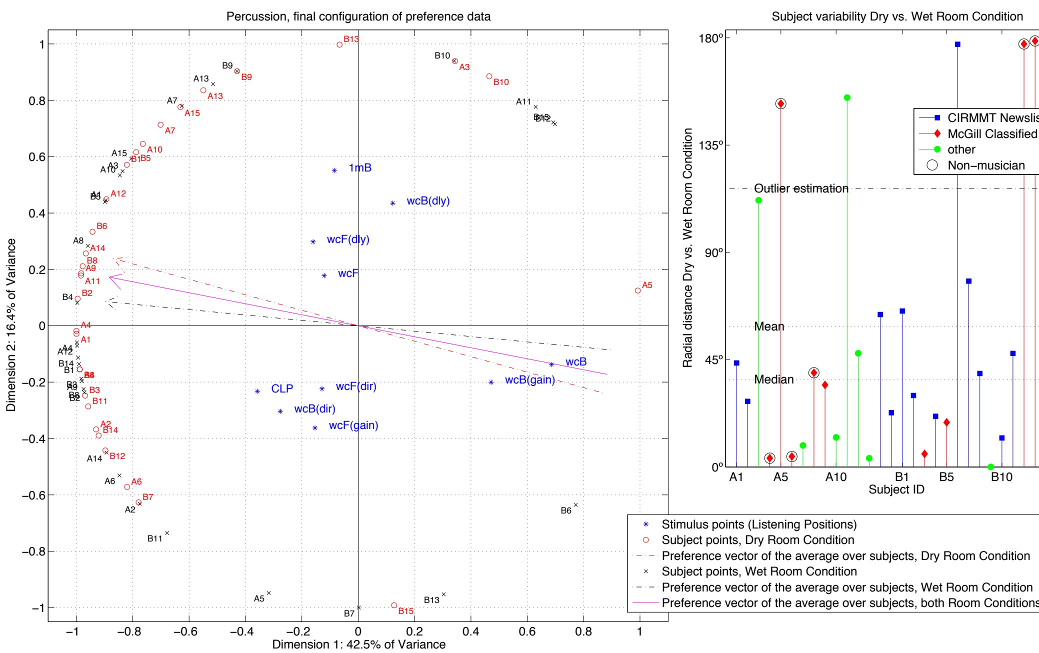


Figure 7.17: Two dimensional MDPREF solution for musical excerpt 2 - Percussion, accounting for 59% of the variance.

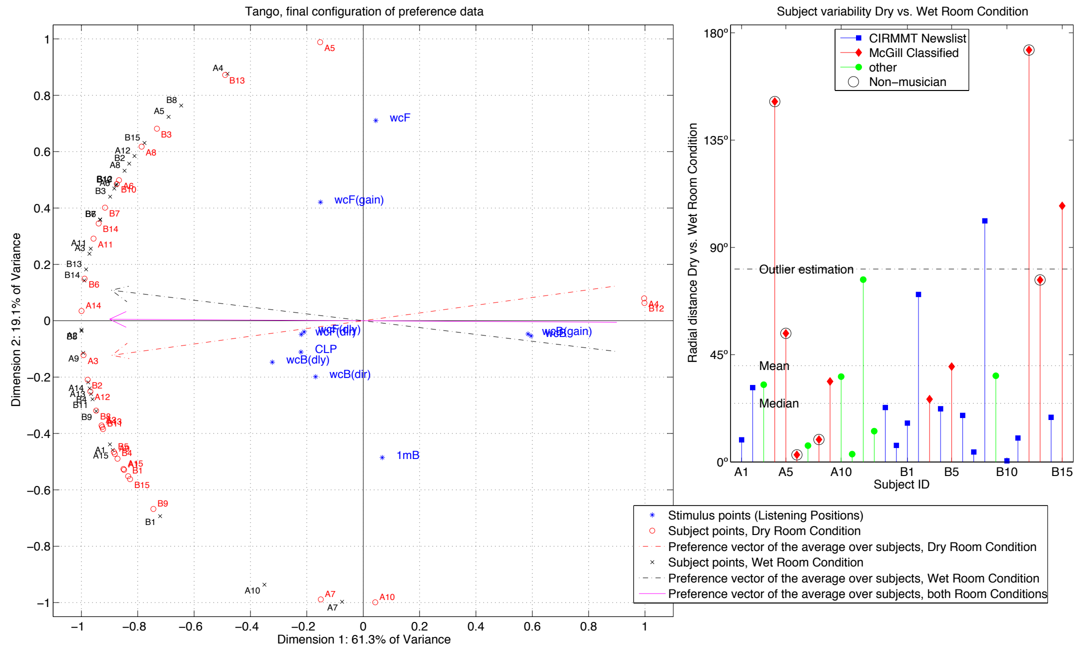


Figure 7.18: Two dimensional MDPREF solution for musical excerpt 3 - Tango, accounting for 80% of the variance.

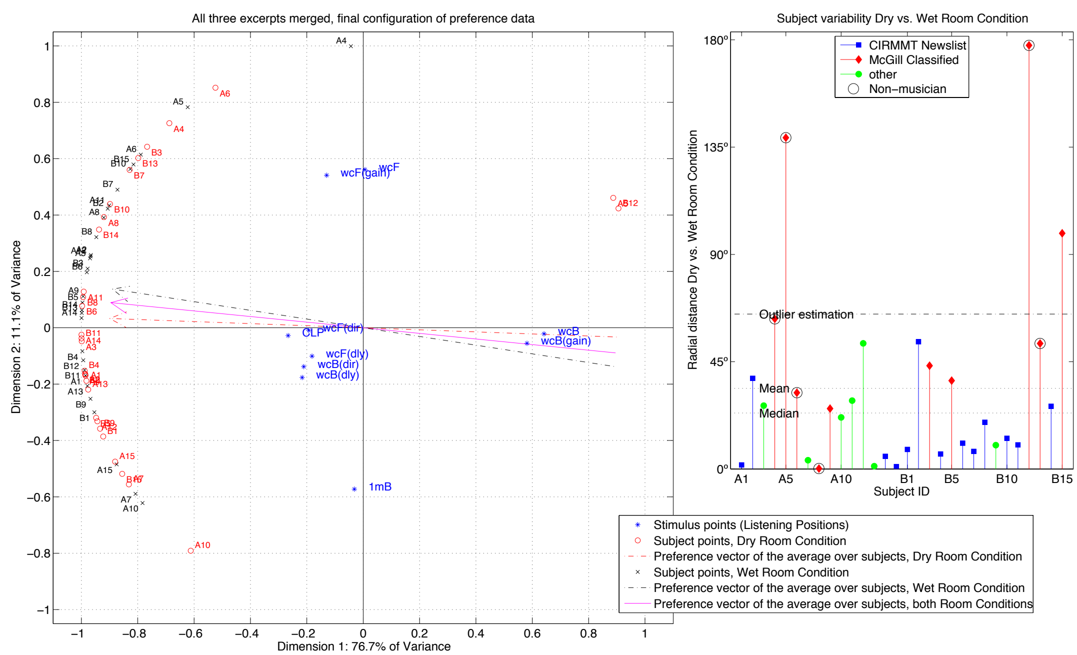


Figure 7.19: Two-dimensional MDPREF result, merging all three musical excerpts together, accounting for 95% of the variance.

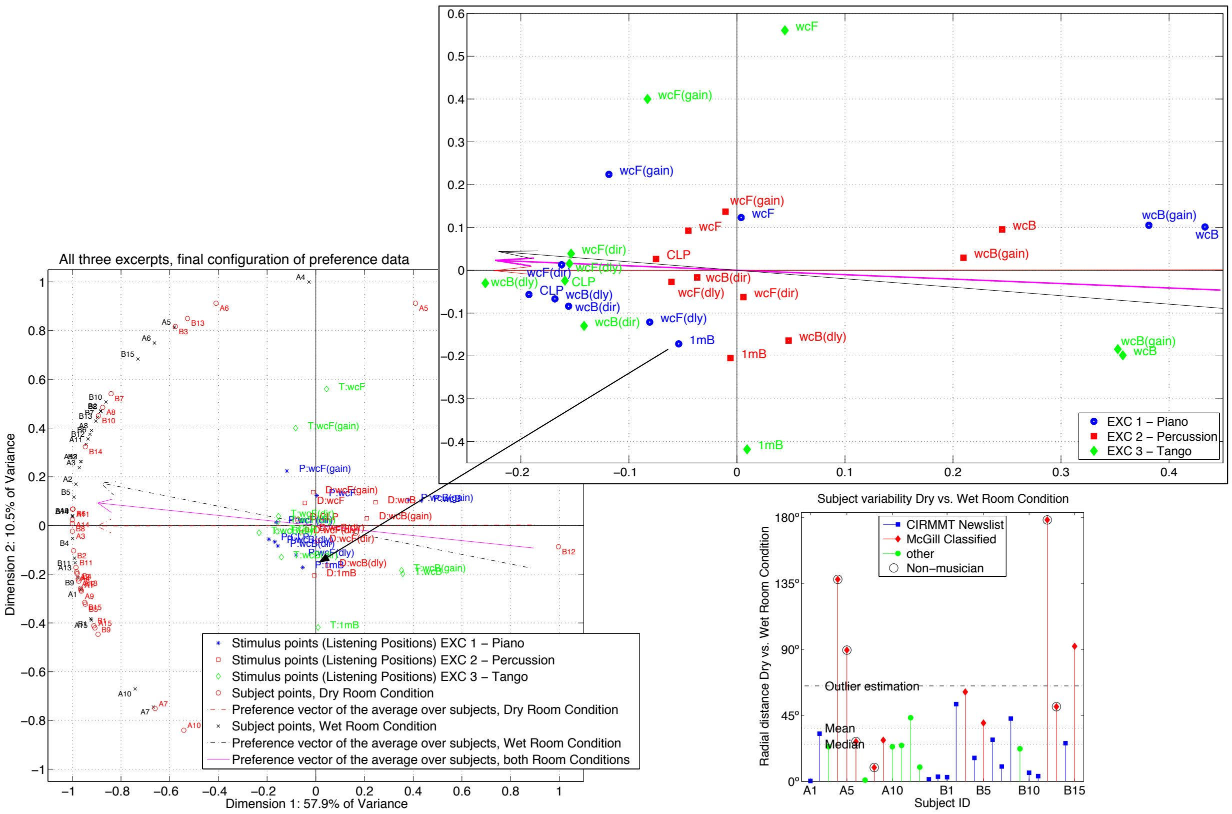


Figure 7.20: Two-dimensional MDPREF result of all three musical excerpts accounting for 68% of the variance, close up plot shows the distribution of the listening positions (POS) for each musical excerpt. The preference vector of the average subject is displayed in magenta.

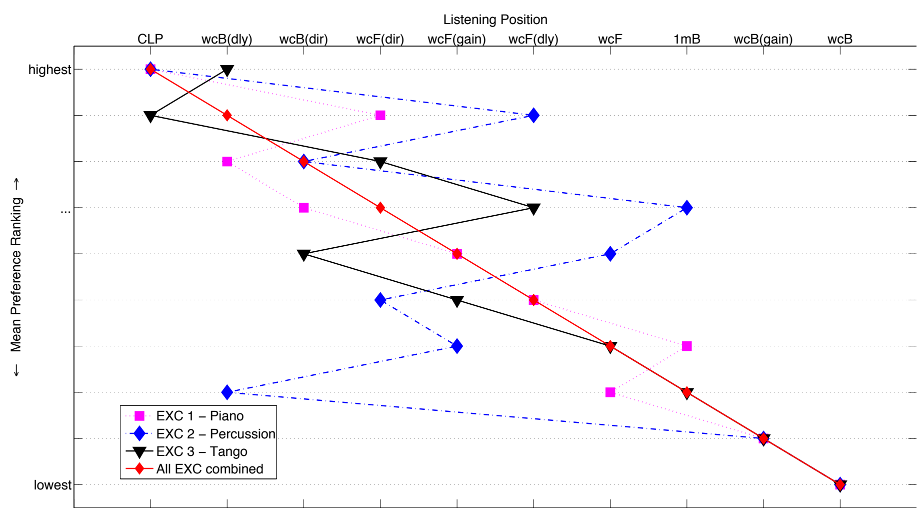


Figure 7.21: Preference ranking of the listening positions of the averaged Subject. Room conditions are combined.

Table 7.17: Average preference ranking. The higher the number, the higher the preference. The second column shows the standard deviation of the subjects' ideal points in degrees: the smaller the standard deviation, the stronger the consensus among the subjects. In **bold numbers:** ranking of listening positions that have changed when outliers are eliminated.

Room condition	Std. Dev. of subject points	CLP	wcF	wcF-dir	wcF-gain	wcF-dly	wcB	wcB-dir	wcB-gain	wcB-dly	1mB
EXC 1 - Piano Fig.7.16											
Dry	59.2°	10	3	8	6	5	1	7	2	9	4
Wet	37.9°	8	3	10	7	5	1	6	2	9	4
Dry & Wet combined	48.7°	10	3	9	6	5	1	7	2	8	4
without potential outliers											
Dry	15.9°	10	3	8	6	5	1	7	2	9	4
Wet	16.0°	**10**	3	**9**	**6**	5	1	**7**	2	**8**	4
Dry & Wet combined	15.1°	**8**	3	**10**	**7**	5	1	**6**	2	**9**	4
EXC 2 - Percussion, Fig.7.17											
Dry	63.3°	9	8	4	5	10	1	7	2	3	6
Wet	68.1°	10	7	5	6	9	1	8	2	3	4
Dry & Wet combined	65.2°	10	8	5	6	9	1	7	2	3	4
without potential outliers											
Dry	54.8°	**10**	**6**	**5**	**4**	**9**	1	7	2	3	**8**
Wet	43.9°	10	**6**	**4**	**5**	**8**	1	**9**	2	3	**7**
Dry & Wet combined	48.1°	10	**6**	5	**4**	9	1	7	2	3	**8**
EXC 3 - Tango, Fig 7.18											
Dry	59.8°	9	3	8	5	7	1	6	2	10	4
Wet	36.4°	8	4	9	6	7	1	5	2	10	3
Dry & Wet combined	48.6°	9	4	8	5	7	1	6	2	10	3
without potential outliers											
Dry	41.7°	9	3	8	5	7	1	6	2	10	4
Wet	32.3°	**9**	4	**8**	**5**	7	1	**6**	2	10	3
Dry & Wet combined	36.6°	9	**3**	8	5	7	1	6	2	10	**4**
All three EXCs combined, Fig. 7.19											
Dry	48.4°	10	4	7	5	6	1	8	2	9	3
Wet	32.1°	10	4	8	9	5	1	7	2	6	3
Dry & Wet combined	40.8°	10	4	7	6	5	1	8	2	9	3
without potential outliers											
Dry	26.7°	10	**3**	7	5	6	1	8	2	9	**4**
Wet	29.5°	10	4	**6**	**7**	5	1	**8**	2	**9**	3
Dry & Wet combined	27.5°	10	4	7	**5**	**6**	1	8	2	9	3

7.5.3 Relating Similarity to Preference

To determine the extent to which preference is related to the four sound attributes, the similarity ratings were correlated with the preference data. We hypothesized that the less similar an attribute was rated in a pairwise comparison, the stronger the absolute preference judgement would be, which leads to a negative correlation coefficients. Figure 7.22 shows that this hypothesis is somehow true, because all attributes are generally negatively correlated to the preference judgments. However, although all combinations show a significant correlation on average, the median values of all four Spearman rank correlation values are $\rho(135) < .5$, indicating only a low correlation. All correlation analyses were also performed on each room condition separately but not displayed here because the results were essentially similar across room conditions.

At least three interpretations are possible: 1. The subjects used sound attributes other than the four that were provided. 2. A more complex model than a simple correlation of dissimilarities is necessary to reveal a relation. 3. The individual strength of the correlation between sound attributes and preference rating differs greatly depending on the musical excerpts, which is not revealed in Figure 7.22.

A stepwise multiple regression was performed to model the preference ratings with the similarity ratings on the four sound attributes *Reverb, Loudness, Position*, and *Timbre*. For this analysis, the behavioral data of all subjects was averaged for each of the 135 pairwise comparisons. A stepwise multiple regression was then performed for each musical excerpt separately (45 behavioral data points) and for all three musical excerpts combined (135 behavioral data points) using the four sound attributes as the predicting variables. The results are shown in Table 7.18 and in Figure 7.23. First- and second-order interactions between the sound attributes (e.g., Reverb $\cdot$ Loudness) were also tested, but did not increase the model's final fit. Figure 7.23 suggests that the higher model fit of the Piano and Tango

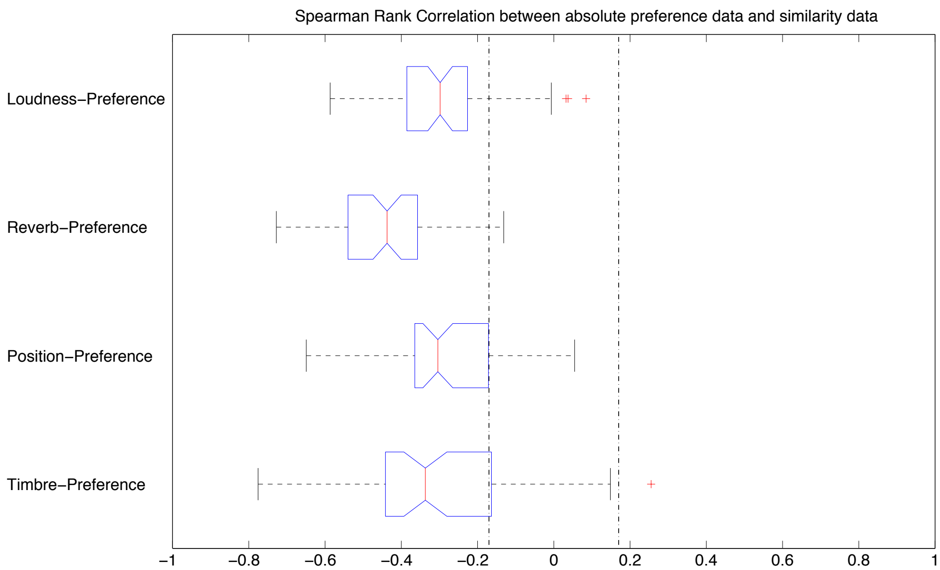


Figure 7.22: Spearman rank correlations between preference data and similarity data. The boxplot shows variability among subjects. Vertical dash-dotted lines represent the range of non-significant correlation values around zero.

excerpt are due to two clusters with very different preference values, whereas the values for the Percussion excerpt are more evenly distributed.

The collinearity measures "Variance Inflation Factor" (VIF) of the successful predictors were all reasonably low (<< 10.0), which confirms stabile regression models. Generally, all four prediction models can reasonably well explain the absolute preference data, R^2 ranging from .53 for EXC 2 to .93 for EXC 1. Remarkably, the four models differ in the number and kind of selected predictors, suggesting that the preference relations depend on the sound stimuli. The number of successful predictors, suggested by the stepwise regression procedure, ranges from one (EXC 2) to three (EXC 1, and all EXC combined). Except for the Percussion excerpt (EXC 2), the strongest predictor for all prediction models is the *Reverb* attribute. For EXC 2, it is *Position*. Further, the first and third predictor always have a negative coefficient B, meaning that the lower the similarity in the pairwise rating, the stronger and clearer the absolute preference. Surprisingly, the second predictor

Loudness for EXC 1 and combined EXC and *Timbre* for EXC 3 have positive coefficients B, meaning that a higher similarity of the Loudness attribute (or Timbre, respectively) lead to a stronger and clearer preference. However, the increase in the model's fit ($R^2_{adj.}$) through the second predictor variable is already fairly low, so the effect is marginal.

Table 7.18: Prediction model of preference data using the similarity ratings.
Final model's fit measure (adjusted R^2) in **bold numbers**.

Predictor successively entered	Partial R	Std. coeff. B	Accum. $R^2_{adj.}$	VIF
EXC 1 (Piano)				
1. Reverb	-.893	-.872	.91	3.09
2. Loudness	.531	.208	.92	1.77
3. Position	-.366	-.204	**.93**	4.33
EXC 2 (Percussion)				
1. Position	-.737	-.737	**.53**	1.00
EXC 3 (Tango)				
1. Reverb	-.895	-1.100	.81	2.05
2. Timbre	.447	.274	**.85**	2.05
All EXCs combined				
1. Reverb	-.891	-.972	.78	1.39
2. Loudness	.425	.231	.81	1.38
3. Position	-.331	-.092	**.82**	1.15

7.6 Discussion and Conclusion

The results of the statistical analyses of the previous sections provide some answers to the initial research questions outlined in Section 7.1.

7.6.1 What Perceptual Parameters Indicate this Sound Degradation?

In the exploratory qualitative study, the most frequently described sound degradation effects found in the tested musical excerpts were identified and merged into four main categories entitled *Loudness, Timbre, Positioning*, and *Reverb*. Using the behavioral data from the quantitative experiment, the absolute preference ratings were reasonably well modelled

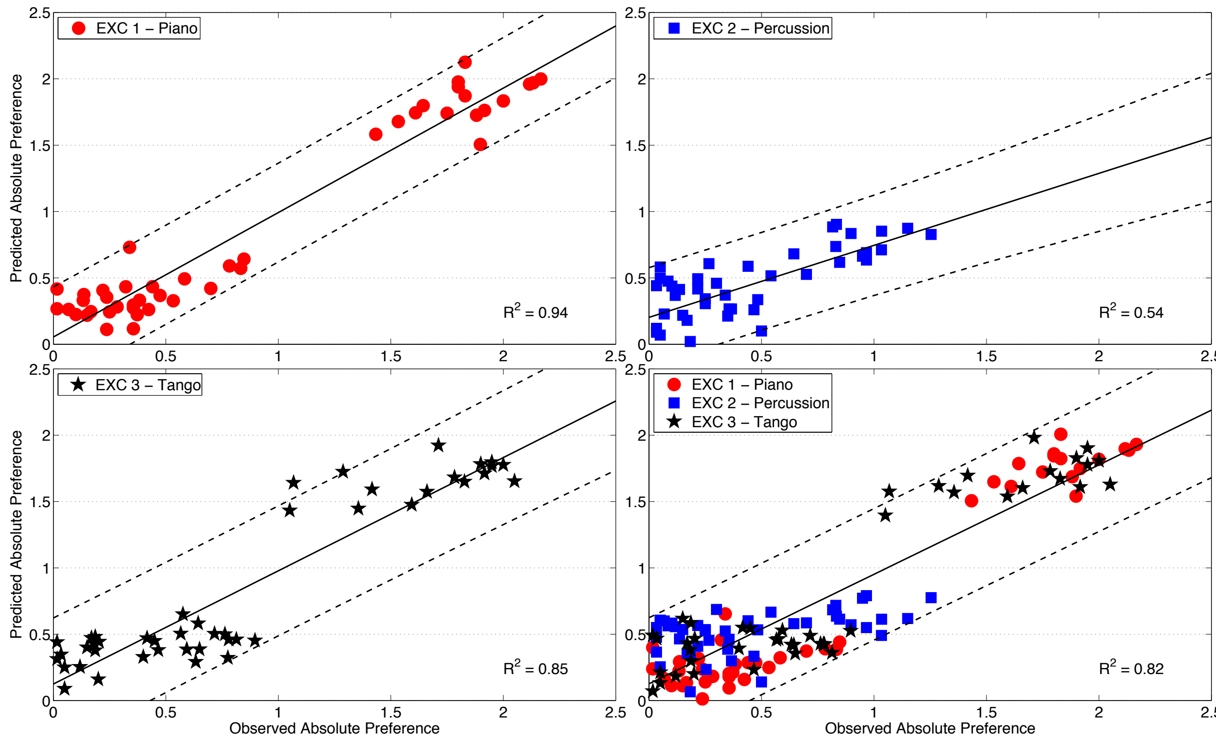


Figure 7.23: Final regression models for each musical excerpt separate, and all excerpt combined (lower right), dashed lines represent 95% confidence bounds of the model.

from the similarity ratings of those four sound attributes with multiple linear regression models (Section 7.5.3). All four sound attributes appear in the individual prediction models and the extent to which they correlate with the preference ratings differs across musical excerpts. In an overall model, combining all three musical excerpts, the sound attributes *Reverb*, *Loudness* and *Position* in this order predict the mean preference very well (Table 7.18).

7.6.2 What Geometrical Factors Contribute to Sound Degradation at OCPs?

As can be seen from the preference analysis in Section 7.5.2, the gain difference stimuli (*wcB(gain), wcF(gain)*) appear close to their associated listening position for each EXC, while the stimuli that simulate the loudspeaker direction (*wcB(dir), wcF(dir)*) appear in close proximity to the central listening position (*CLP*). The level differences due to a change in listening position thus seem to contribute the most strongly to the overall perceived sound degradation. The stimuli that contain only the time-of-flight differences between the loudspeaker feeds *(wcB(dly), wcF(dly))* appear in close proximity to the *CLP* for the musical excerpts 1 and 3, which means that their preference is similar to that of the *CLP*.

In contrast, for the percussion excerpt (EXC 2), the stimuli *wcB(dly)* and *wcF(dly)* are located further away from the CLP, meaning that the time-of-flight differences cause a perceivable degrading effect for this stimulus. The stimulus *wcF(gain)* is even closer to the *CLP* than *wcF(dly)*. This makes sense, because due to their transient nature, percussive sounds are more sensitive to time delay changes than sustained piano and guitar sounds as in EXC 1 and 2 and cause artifacts such as a change in the position of the phantom source and comb-filtering effects.

7.6.3 Does the Musical Material Affect This Sound Degradation?

In our data, significant effects for the main EXC effect and its interaction EXC×POS were found, which indicate that the musical material affects the strength of the perceived off-center sound degradation. The Bonferroni post-hoc comparison of the musical excerpts with regard to the four sound attribute ratings also indicate clearly that in most of these comparisons, the attributes were rated differently. Furthermore, the MDS visualizations of the sound attribute ratings (Figure 7.12–7.13) and the ranking of the preferred listening positions (Figure 7.21) provide a visual confirmation.

Because the three musical excerpts are different in multiple ways, such as musical material and style, instruments, performer, or the recording technique, one could argue that the stimuli are not controlled enough to make valid conclusions. However, these stimuli can be considered as real-world examples and are therefore more ecologically valid and meaningful than are noise bursts or other artificially created stimuli that are usually used in psychophysical experiments. Further, the variety in the excerpts also helped to reduce subjects fatigue.

7.6.4 Does a Listening Room's Acoustics Affect Off-center Sound Degradation?

The analysis of the preference ratings revealed a significant effect of the room condition (see CON in Table 7.15 and 7.16). Also, the individual preference vectors of the internal preference mapping procedure (Section 7.5.2) show certain differences between the two room conditions for some participants. This fits to the comments of several participants, which reported after the second experimental session that the pleasantness and difficulty of the experimental task differed between the room conditions. However, analyzing the variances of the similarity data (Section 7.5.1) did not reveal any significant effect due to the room condition. This result is rather surprising, because a common belief is that room acoustics is a salient factor in sound quality perception. The reason why the room acoustics did not affect the behavioral data might be found in the experimental method. First, the change in the room acoustics, measurable for instance through the reverberation time (Table 7.5), from the *Dry* to the *Wet* condition, is only about 0.2 sec and might have been too subtle to cause a significant difference in the behavioral data. An alternative explanation might be the human ability to suppress the acoustical influence of the room on the perception of reproduced sounds, (human adaption to the room acoustics, see Schuck et al. 1993; Olive et al. 1995; Olive 2008). According to Olive (2008), this effect seems to occur after less than a few minutes of being in a room, but has not been fully investigated. However, it may have been present because of logistical reasons: the two room conditions were tested in a blockwise experimental design.

Therefore to answer this research question, further investigation is needed. Also, the experimental method focuses only on the simulation of the direct sound component and does not exactly account for complex early reflection patterns of the real room caused by a five-channel loudspeaker setup at off-center locations. An effect due to the early reflection pattern is possible. Because the early reflection pattern also depends on the loudspeaker directivity, a follow-up study could focus on loudspeaker directivity. Evans et al. (2009) summarized technical aspects and strategies on integrating loudspeaker directivity in listening experiments, which should be considered in a follow-up experiment.

7.6.5 Implications on Designing Surround Loudspeaker Setups

This study has shown that off-center sound degradation is mainly caused by the level differences of the loudspeaker feeds, and that artifacts related to the reproduction of reflections, reverberation, and room impression of the audio material, and to the overall loudness are primarily perceived (Table 7.18).

Consequently, for development of surround-sound loudspeaker systems with a large listening area, one needs to reduce the level differences between the loudspeaker feeds. This could be practically achieved by increasing the distance between listening area and loudspeaker, so that the inverse square law (see Section 6.1.2) determines a weaker change in the loudspeaker level differences. However, room acoustics conditions practically limit this approach (room size, critical distance and the increase of listening room reverb).

A more promising avenue can be taken by using circular arc-line array loudspeakers which ideally create cylindrical sound sources at each loudspeaker position. Rather than 6 dB attenuation per doubled distance of conventional point-source-like loudspeakers, the sound level of line-array loudspeakers decreases at a lower rate of 3 dB. Consequently, the level differences at OCPs are reduced. Research into these promising types of loudspeaker arrays is underway (Keele and Button 2005, 44).

Alternatively, a Wavefield synthesis system (WFS) can also be exploited to reproduce the loudspeaker feeds as *virtual loudspeakers* at their respective positions using a plane-wave characteristic, which also has a weaker level attenuation than a conventional loudspeaker due to distance. Boone et al. (1999) proposed to virtually reproduce the surround loudspeaker with a WFS array if the listening room is too small to correctly position the surround loudspeakers as shown in Figure 7.2. In a conventional surround audio production, the surround channels are used for ambient sounds and room response from behind. As shown in Table 7.18, the reverberation sound attribute is the most highly correlated with preference. Therefore, our findings support plane-wave reproduction of just those surround loudspeaker feeds to increase the listening area for a variety of audio material, which reproduces a typical front-stage impression.

7.7 Summary

In a series of listening experiments, we investigated the effects that characterize off-center listening positions due to geometrical circumstances. These effects are the direction of arriving wavefronts, time-of-flight differences, and level differences of the loudspeaker feeds. In a qualitative verbalization study, we first categorized common artifacts in off-center listening situations. Findings of the succeeding quantitative experiment show that off-center sound degradation can be primarily associated with the level differences of the loudspeaker feeds and that artifacts related to the reproduction quality of the reverberation (i.e., room impression, reflections, envelopment) in the recording and to the overall loudness can predict preference (Table 7.18). However, significant differences between the three musical excerpts were found in the behavioural data, which indicates that the reproduced audio material affects the nature of a perceived off-center sound degradation. These findings have implications for the design of surround loudspeaker setups as outlined.

7.8 Acknowledgement

This work was funded by a grant from the Canadian Natural Sciences and Engineering Research Council and the Canada Council for the Arts (NSERC, CCA) to Stephen McAdams and by a student award from the Centre for Interdisciplinary Research in Music Media and Technology (CIRMMT) to Nils Peters. Many thanks to Harold Kilianski and Yves Methot for their support during the experiment and to all 40 participants.

Part IV

Conclusion

Chapter 8

Summary and Discussion

This chapter presents a summary of the work of this dissertation. Included is some discussion on the findings, results and relevance of this work, and reflections on the spatial audio research process. Derived from all this are suggestions for future research and development, all of which continue to be motivated by the interdependent relationships among composer, technology and listener as illustrated in the theoretical model (Section 1.3).

8.1 Summary

8.1.1 Use of Spatial Audio Technology

As an engineer, to gain insight into the artistic use of spatialization of sounds and music by composers and sound artists, a web-based questionnaire was developed (Appendix A) to survey the perceptual aspects that are essential in spatialization, the challenges with current tools, and what functionalities future spatial audio technology should feature. The survey's primary goal was to inform the methods of development of spatialization tools and the design of listening experiments in this dissertation. Chapter 2 reported on the findings through quantitative and qualitative analyses of this survey.

The high completion rate of the survey (52 out of 94 surveys were completed) suggests that this kind of questionnaire was well received by artists. Respondents find that spatial aspects are important perceptual attributes of listening and that spatialization can enhance the experience of the listener. Furthermore, spatialization can be an important compositional paradigm for artistic expression and more than just a “stunning effect”. Responses demonstrate the usefulness of spatial audio technology for musical applications and the need for better scientific understanding in order to continue development based on the needs of the primary users. Overall, the users were not fully satisfied with the capability of their current audio tools to produce desired spatial aspects, the most important being “Immersiveness”, “Distance perception of sound sources” and “Localization accuracy of sound sources”.

A gap between importance ratings and satisfaction levels can be an opportunity for further research. The spatial aspects of “Elevated sound sources”, “Distance perception”, and “Virtual sound sources within the audience area” were found to be the most important, but the least satisfying. Interestingly, there is a substantial body of research on these highly rated but mostly dissatisfying spatial aspects. “Virtual sound sources within the audience area” (a.k.a. focused sources) is a paramount feature of the emerging Wave Field Synthesis (WFS) and near-field-compensated Higher Order Ambisonics (NFC-HOA). Because these are fairly young reproduction methods, the research potential has not yet been fully exploited and improvements in the rendering of such focused sources are to be expected. The aspects “Elevated sound sources” and “Distance perception” are well-studied fields (e.g., Blauert 1997), but not every spatial audio rendering concept incorporates these features. For instance, Vector Base Amplitude Panning (VBAP) is capable of rendering elevated sound sources, but does not render distance; for WFS, it is the opposite.

The survey also showed that the three most required technical features are the "Integrability into DAW", "Controllability via external controller", and "Spatial rendering in real time". The rationale for choosing spatialization tools primarily depends on the ease of use, more than the rendering quality or the availability and purchasing costs. Surprisingly, the most often used rendering technologies are the panning functionalities of the DAW or hardware mixing console. Moreover, many available spatialization tools, which often allow the creation of more spatial aspects than the DAW's panning functionalities, are practically unknown. However, respondents are aware of the emerging rendering concepts associated with WFS and Ambisonics and plan to use them.

Surprisingly, of the 52 participants, only two were willing to be contacted for a notation sample of their spatial composition. A more positive response ratio would have allowed me to use notation examples to develop test scenarios for the development process of ViMiC (Virtual Microphone Control) and SpatDIF (Spatial sound Description Interchange Format) or to design stimuli for the listening experiments in Chapter 7.

8.1.2 Experiments

Four experiments were performed to examine listener perceptions at different listening positions. Two investigated the effect of the microphone recording technique and listening room environment (Chapter 6), a third collected lexical descriptions of perceptual artifacts in a qualitative way (Section 7.3), and the fourth investigated the effect of the three primary factors that characterize off-center listening positions based on geometrical listener-loudspeaker relationships in an ITU surround-sound loudspeaker setup using three types of musical excerpts (Section 7.4). The design of the experiments is linked to the results of the survey on the compositional use of spatialization (Chapter 2) in an attempt to create a collaborative research-development paradigm, where the musical creators (composers) and their ideas, as well as their technical needs, directly inform the development of spatial audio technology.

In the first two experiments, I found that all three tested multichannel recording techniques led to a somewhat radial sound degradation from the Central Listening Position (CLP). Moreover, the strength of the sound degradation at off-center listening positions is a function of the recording technique. Spaced microphone techniques create less sound degradation for off-center listening than the coincident Ambisonics technique. Furthermore, when loudspeaker configuration and the listening room acoustics do not follow the recommendations for which the multichannel recording was created (as often mentioned in the survey), none of the tested recording techniques appeared to be significantly superior. Instead, the sound material was found to be the dominant factor. The data suggest that in a more reverberant listening room, a rather uncorrelated sound material (i.e., symphony recording) is likely to be better reproduced at off-center listening positions than a more highly correlated material (i.e., a solo piano recording).

To gain an understanding of perceived sound degradation, a variety of sound features were extracted from the stimuli and used in statistical models to predict the sound degradation ratings of the different off-center listening positions (Section 6.5). In the resulting statistical models, the primary predictors are spectral and the secondary ones are based on spatial aspects. Moreover, the models suggest that off-center sound degradation was primarily due to modifications in the higher frequency range at and above 2.5 kHz (Section 6.5.2). Off-center sound degradation was found to depend on the musical material *and* the listening environment, because the statistical models differ between the two musical excerpts and the two listening environments. These findings have a direct impact on optimizing the production of audio material to be presented in medium-sized rooms to a larger audience and to inform the design of surround-sound reproduction environments. The experiment's findings also support the ViMiC spatialization concept, described in Chapters 4 and 5, by suggesting that a spaced arrangement of virtual microphones creates less off-center sound degradation due to the (additional) employment of Inter-Channel Time Differences (ICTDs), than spatialization concepts solely based on amplitude panning as used by many spatialization tools (see also Figure 1.1). However, as the ViMiC approach uses the principles of sound recording to encode spatial information and simulates the conditions of real recordings to create a coherent sound image, the same precautions have to be considered as in reality when choosing the arrangement of virtual microphones (see e.g., Theile (2001); Griesinger (2001)).

The two experiments in Chapter 6 generated new research questions and an altered methodology for investigating off-center sound degradation in Chapter 7, where a novel loudspeaker-based experimental apparatus was developed, the software for which was partially implemented using Jamoma modules developed in the course of this dissertation and listed in Appendix C. With this new apparatus, the individual effect of the time-of-arrival differences, the sound-pressure-level differences between the signal feeds, and the direction of the arriving wavefronts can now be studied, all of which were identified as contributing to perceptual effects at off-center listening positions.

In an exploratory qualitative experiment using three musical excerpts (Section 7.3), participants were virtually placed at different listening positions and were asked to describe in their own words the perceived sound differences. An interpretation of these responses showed that the five most commonly described artifacts between off-center listening positions are related to the position of sound sources, the distance and depth of sound sources, reverberation and envelopment, spread and width of sound sources, and sound coloration. The results again show the complexity and the multidimensional nature of perceived off-center sound degradation. The responses also differed across musical excerpts, suggesting that the perception of artifacts depends on the sound material.

The succeeding quantitative experiments (Section 7.4) primarily investigated how the three factors that characterize an off-center listening position (time-of-arrival differences, sound-pressure-level differences between the signal feeds, direction of the arriving wavefronts) contribute to sound degradation under two different acoustical room conditions. Off-center sound degradation at the tested seats is primarily caused by the level differences of the loudspeaker feeds. However, the behavioral data show significant differences between the three musical excerpts, which indicates again that the reproduced audio material affects the nature of a perceived off-center sound degradation. We found that the time-of-flight differences from loudspeaker to listener have a stronger perceptual effect on percussive sound material than on sustained sound material. In two out of three musical excerpts, off-center sound degradation was primarily correlated with artifacts related to the reproduction quality of the reverberation (i.e., room impression, reflections, envelopment). In the case of the Percussion excerpt, the preference for a listening position could only be moderately well predicted. Here, the overall loudness was the best predictor for preference.

The latter experiment was conducted in two different acoustical room conditions, one was hemi-anechoic and the other had a reverberation time of about 0.2 sec. One hypothesis was that the more reverberant room condition would affect and reduce perceived sound degradation. For the preference ratings, the room condition was significant (Table 7.15 and 7.16). However, to my surprise, for the similarity ratings of the tested sound attributes, no significant difference between those two room conditions were found. The acoustical difference (e.g., in reverberation time) between the two room conditions might be too small to cause measurable effects in the similarity ratings. This should be further investigated in follow-up experiments, using room acoustics with larger reverberation times, for example.

8.1.3 Developments

The survey results from Part I suggest great individual and context-related differences in the compositional use of spatialization and that there is no one spatialization system that can satisfy every artist. Therefore, Chapter 3 presented a stratified approach for sound spatialization to structure technical aspects of spatialization systems with the aim of finding new ways of artistic expression and of combining spatialization tools according to an artist's individual needs. For instance, an authoring software can be connected to a multi-touch screen, both devices controlling in real time simultaneously an Ambisonics rendering system, configured for binaural reproduction. Spatialization tools by different developers such as Holo-edit, SpatDIF, Jamoma and ICST's Ambisonics tools were reviewed in this context in Section 3.4, and it was demonstrated in Figure 3.2 how those tools can be combined. Artists and researchers would benefit greatly if other spatialization tools could also be interconnected and integrated into existing environments using this approach. It remains to be investigated how this stratified approach can be integrated into the widely used DAW's, known for their static bus architecture, which often implies a limitation in the routing of multichannel audio signals.

Chapter 4 reports on the developments of ViMiC as an external for the Max/MSP computer music environment. To make ViMiC available in Digital Audio Workstations (DAWs), it was developed as an AudioUnit plug-in with the assistance of Tristan Matthews and Tim Place. To embed the ViMiC Max/MSP external, several high-level modules for the open-source project Jamoma were created. Chapter 5 demonstrated how the survey findings were applied in the design process of ViMiC, with the primary concern of flexibility, usability, and integrability into existing environments. The variety of use cases across many different and unexpected scenarios demonstrates ViMiC's current usability and portability and exceeds the original concept of providing a technology for the compositional use of spatialization.

8.2 Reflection

During this dissertation, methods and knowledge from a number of different fields were applied and combined, specifically room acoustics, digital signal processing, human-computer interaction, psychoacoustics and music perception. Besides the interdisciplinary challenge in devoting to each research field the necessary focus, interest, and understanding, there are very real physical barriers that might explain why there is still little research on spatial audio reproduction in performance venues. Performance venues are often not pre-equipped with the necessary loudspeaker equipment and/or configuration. Besides the four listening experiments reported in Part III, together with Georgios Marentakis, three other experiments on sound localization using multichannel loudspeaker setups in three different spaces were carried out. Further, I assisted in the technical setup for a concert by Sean Ferguson employing 24 loudspeakers (see Section 5.3.2), performed ViMiC live spatialization in a telematic music concert at the ICAD 2007 conference (Section 5.3.3), and ran several ViMiC demonstrations using eight and more loudspeakers in public events. If more performance venues would be readily equipped with adequate multichannel loudspeaker setups, preparation time for measurements and listening experiments could be reduced, which would facilitate this research field.

As mentioned before (see Section 1.2), spatial sound properties are often considered to be a secondary parameter in music perception research. The significance of type of musical excerpts and recording technique, for example, on spatial audio perception leads me to speculate on the complex relationship between timbre and spatial aspects. Can timbre perception be studied independently of spatial audio perception, and vice versa?

Another field where spatial perception is constantly neglected is the current practice of hearing tests. When conducting listening experiments, a pure-tone test is usually employed to detect hearing damage among the participants. When I prepared the experiments in Part III, I wanted to test the spatial hearing performance of my participants. After a literature review on hearing tests, I was surprised to find no standardized listening test dedicated to spatial hearing, except for a related proposal for a spatial hearing questionnaire for patients with hearing loss (Tyler et al. 2009). I think a hearing test which includes aspects of spatial hearing is important for this and other related research fields, because nothing concerning a listener's spatial hearing abilities can be derived from a pure-tone hearing test.

For the development of ViMiC and other spatialization tools, I decided to use the Jamoma platform, for several reasons. First, Jamoma allowed me to use existing resources and programming solutions to facilitate e.g., interface design, OSC communication and parameter handling, making ViMiC more flexible and easier to use. Further, in the context of the NSERC/CCA research project *Compositional Applications of Auditory Scene Synthesis in Concert Spaces via Gestural Control*, Jamoma was ideal for exploratory research using different gestural controllers, because many functionalities such as sensor interfaces and mapping solutions were already available in Jamoma, thus avoiding repetitive development efforts. Also, Jamoma has a vibrant development team, always trying to create more usable computer music software, and an international user-group consisting of both artists and researchers, eager to test and comment on ViMiC. Especially valuable (not only) for this research, all major sound rendering algorithms (VBAP, HOA, ViMiC, DBAP)[1] are now included and accessible in a standardized way. This rich palette of spatialization tools may facilitate the design of scientific studies, e.g., in the perceptual evaluation of different

[1] At the time of writing, Wave Field Synthesis is not included, but can be simulated using ViMiC as indicated in Section 4.5.

spatial audio rendering concepts. As a researcher in the field of music technology, using the same set of tools as artists use in their work can make research and development more applicable and meaningful.

8.3 Relevance

The ideas, methods and tools presented in this dissertation may be relevant in a number of different fields. The analysis of the survey responses on the compositional use of spatialization can be a valuable resource for other artists, as well as for researchers and developers in the fields of music technology, musicology, and music preservation. Also, concert organizers and room acousticians can learn from the participants' comments on the most challenging aspects of performance venues, related to the acoustical conditions, technical limitations, and time constraints at the venue. I believe that this survey can advise others and help to establish a communication between artists and researchers, to enhance future artistic practice through better audio technologies. I also believe that this survey had a bidirectional effect, in that the questions caused further thinking, considering and understanding about spatialization and spatial audio technology among the participants.

One of the cornerstones of this dissertation, the ViMiC spatialization concept is now maturely developed and already applied in multiple fields, ranging from musical applications such as sound installations and concert performances, to motion picture systems and research in audio technology and medicine. Potential new fields of application include the use of ViMiC as an educational tool for sound engineers. Further, it would be interesting to test ViMiC with NHK's emerging 22.2 reproduction system (Hamasaki et al. 2007).

Work on the Jamoma project in the form of the development of new modules (Appendix C), improvements on the core-architecture, co-authoring several conference proceedings (e.g., Place et al. 2008; 2010a;b) as well as hosting user and developer workshops, has exposed ViMiC to increased attention and widespread use, both for research and performance throughout the world[2]. Jamoma is also an ideal test-bed and one of several environments in which SpatDIF is currently tested. SpatDIF, initiated by this dissertation, is under development in the community and has drawn attention from different international research institutions and individual composers, as well as from the commercial audio sector.

The newly developed experimental apparatus to simulate off-center listening positions in a controlled environment (see Section 7) has a large potential for other experiments that involve the effect of listening position on sound and music perception. The findings of the experiments contribute to the understanding of music perception at off-center listening positions in surround-sound loudspeaker setups and can inform developers, audiophiles, researchers, engineers and composers.

8.4 Final Conclusion

The principle aim of this dissertation was "to inform and contribute to the research and development of spatial audio technology through a systematic development of sound spatialization tools for musical applications and perceptual studies with regard to loudspeaker reproduction for a larger audience," (page 2). From an engineering viewpoint, this dissertation provided great insight into different aspects of production and reproduction of spatial audio content, in particular for off-center listening.

The experiments on the effect of the listening position indicate that the spatial perception of reproduced sound is multidimensional, depending on many technical, acoustical and musical details. Although each detail could not be fully investigated in the course of

[2] Google Scholar currently lists 46 publications mentioning Jamoma `http://scholar.google.com/scholar?q=jamoma`. Examples of artistic work facilitated by Jamoma are listed at `http://redmine.jamoma.org/projects/jamoma/wiki/Art_Projects_Using_Jamoma`, accessed July 2010.

this dissertation, it proves that perceptual research is an important component in guiding the development of spatial audio technology.

Technically, one has to acknowledge the high complexity implied in emerging high-quality spatialization techniques such as WFS and HOA and in large-scale loudspeaker setups. To create spatial audio technology that is accepted, understood, and used by artists, researchers need to include both creators and listeners in the development process to lower the high conceptual and technical barriers and to account for the artists' needs.

A comprehensive research approach, including sound production and reproduction aspects, is beneficial for improving spatial audio technology and for gaining scientific understanding of spatial sound perception.

8.5 Future Work

The experiments in this dissertation revealed one of the first insights into off-center listening perception, and it is clear that there are more aspects that need to be understood.

Due to practical reasons, some of the experimental variables were not independently controlled. For instance, in the experiments in Chapter 6, both the loudspeaker type and the listening room acoustics were different between the two experiments. Future experiments controlling these variables would provide more information on how they interact. Further, it was found that the musical materials have an effect on the perception of sound degradation at off-center positions, but controlled musical materials, recorded simultaneously with different microphone techniques as used in the experiments in Chapter 6, are logistically demanding and expensive to produce. A pool of such (simultaneously captured) recordings shared across researcher would be ideal to divide costs and production efforts, to foster future research, and to increase transparency and repeatability of the experiments.

The experimental method to investigate off-center listening positions by virtually repositioning a listener as presented in Section 7.2.1 enables a variety of new perceptual studies. An interesting task would also be the comparison of this presented method with an BRS (Binaural Room Scanning, Olive et al. 2007) approach. The effect of the room environment, which was not found to be significant in the similarity ratings of the experiment in Section 7.4, is puzzling and could relate to the human ability to adapt to an acoustic environment, a research topic that was investigated by Olive (2008). In the context of off-center listening positions, this effect needs further investigation.

All four described experiments focused on the auditory percept and excluded potential biases due to other sensations such as visuals. Given that in a live performance, auditory perception is part of the multimodal experience, the question of the importance of visual cues (e.g., position, room interiors, fellow audience members) is a field for future research.

The results in the listening experiments of Part III show that the off-center sound degradation perception depends on the sound material, on the loudspeaker setup and on the combination of sound material and listening environment. Further, the analysis of the questionnaire in Chapter 2 showed that composers organize their use of spatial properties, for instance, according to timbre, texture, and musical function. These two findings, from different parts of this dissertation, have a direct impact on future developments. In a music production process, often only one spatial audio rendering concept is used *exclusively*. Because rendering concepts typically have advantages and disadvantages in terms of technical features and perceptual qualities, their *character* should be considered in combination with the sound source, analogous to how a sound engineer choses a specific microphone to record a specific instrument. Further, the combination of rendering concept could be beneficial for the reproduction quality of the piece. Future musical application should therefore allow or even promote the combined use of different rendering algorithms within a composition, similar to a painter who uses different brushes, or a chef who uses different knives. Currently, a number of tools (e.g., Spatialisateur (Jot 1999), SoundScape Renderer (SSR, Geier et al. 2008), see Figure 1.1) include a variety of sound rendering techniques, but using multiple rendering techniques simultaneously for different sound sources is usually not contemplated.

Theories in spatialized audio are mostly based on the understanding of signals that are not necessarily musical. Therefore, future research needs to address the spatial perception of musical signals, for instance through employing techniques and algorithms from the field of musical information retrieval and machine learning. As a result of this research, an application could be developed that recommends the optimal rendering algorithm to a composer working with a particular sound material and a listening room acoustics. This would facilitate the use of the growing variety and complexity of spatial sound rendering techniques. The technical structure for such systems has already been laid out through the stratified approach for sound spatialization described in Section 3. An early prototype application which enables the user to manually combine common spatial audio rendering concepts (VBAP, DBAP, ViMiC and HOA) according to compositional aspiration and/or technical possibilities is currently under development (see Place et al. 2010b, Figure 6). Current investigations in exploiting multi-core computer systems for real-time audio processing supports this future research, because high-quality spatial audio rendering for large-scale loudspeaker setups is a potential resource-demanding DSP process.

Another step could address the porting of the ViMiC system to other computer music environments, such as SuperCollider or IRCAM's computer-aided composition environment OpenMusic which renders sounds off-line, rather than in real time (deferred time concept). ViMiC with OpenMusic therefore may enable the creation of complex multichannel sound structures that are not yet possible in real time with today's computer technology.

Part V

Appendices

Appendix A

CIRMMT Survey

CIRMMT Spatialization Questionnaire

1. Welcome!

Welcome to the CIRMMT Spatialization Questionnaire!

Thank you for contributing to this questionnaire.

Your participation is of great value in our goal to improve composer's working conditions through better audio technology.
By getting to know how you think about spatialization and what tools and functionalities spatial audio systems should strive to include, we can focus our research to meet your needs.

All the data collected will be kept confidential and your name will not be included as part of any aggregated data.
If you do not wish to complete the questionnaire at any time you are free to exit.
You can also re-enter at any time to update or complete your responses.

If you have any questions about the purpose of this questionnaire, or problems in undertaking it, please contact Nils Peters, nils.peters@mcgill.ca.

Thank you for your time!

CIRMMT Spatialization Questionnaire

2. General Section

1. Your age in years

2. Your gender

- Female
- Male

3. How many years have you been composing?

overall

with the help of computer

with spatialization techniques

4. Please list the universities/conservatories of your musical education. If you are not formally trained, please fill in "self taught".

One

Two

Three

Four

Five

5. Please list your current affiliation(s). If you are currently not affiliated, please provide your country of residence.

One

Two

Three

Four

6. Do you associate yourself with a compositional concept/school (e.g. l'école spectral, acousmatique) ?

- No
- Yes, pelase specifiy

CIRMMT Spatialization Questionnaire

3. Compositional Section

*** 7. For which "orchestration(s)" do you primarily use spatialization in your music?**

- [ ] instrumental
- [ ] live electronics
- [ ] mixed music
- [ ] prepared electronics
- [ ] Other (please specify)

*** 8. In what musical context do you primarily use spatialization?**

- [ ] theater
- [ ] film, video
- [ ] internet applications
- [ ] concert
- [ ] dance
- [ ] sound installations
- [ ] Other (please specify)

CIRMMT Spatialization Questionnaire

4. Compositional Section

*** 9. Rate the importance you attach to the following spatial features:**

	not important	slightly	fairly	very	extremely important	N/A
creating of fast movements of sound sources	◯	◯	◯	◯	◯	◯
creating of slow, subtile movements of sound sources	◯	◯	◯	◯	◯	◯
creating of movements along complex trajectories	◯	◯	◯	◯	◯	◯
simulation of Doppler effect	◯	◯	◯	◯	◯	◯
avoiding Doppler effect when moving sound sources	◯	◯	◯	◯	◯	◯
localization accuracy of sound sources	◯	◯	◯	◯	◯	◯
elevated sound sources (height information)	◯	◯	◯	◯	◯	◯
virtual sound sources within the audience	◯	◯	◯	◯	◯	◯
distance perception of sound sources	◯	◯	◯	◯	◯	◯
adaptable apparent source width	◯	◯	◯	◯	◯	◯
simulating of a specific room acoustic	◯	◯	◯	◯	◯	◯
creating of artificial/unreal room impressions	◯	◯	◯	◯	◯	◯
immersiveness	◯	◯	◯	◯	◯	◯
large listening area	◯	◯	◯	◯	◯	◯
simulation of an instrument's directivity	◯	◯	◯	◯	◯	◯

Other (please specify and rate)

CIRMMT Spatialization Questionnaire

5. Compositional Section

*** 10. At what point in the compositional process do you usually decide to integrate spatialization?**

- ◯ Spatialization is the initial thought
- ◯ I create all musical aspects (incl. spatialization) together
- ◯ Spatialization is the final step
- ◯ None of these fit my composition style (please specify)

11. Why do you use spatial aspects in your music?

CIRMMT Spatialization Questionnaire

6. Compositional Section

*** 12. Would your music work without spatialization?**

- ◯ Yes
- ◯ No
- ◯ I don't know

Please elaborate

13. How do you configure sound elements in space to achieve spatialized musical expression (i.e. dynamic (moving), static (fixed to one point), multiple sources, etc.) ?

14. What are the most important perceptual aspects or effects you take into consideration when spatializing a specific instrument/sound?

CIRMMT Spatialization Questionnaire

7. Compositional Section

*** 15. In what kind of venues are your spatial compositions performed?**

- [ ] specialized venues for electro-acoustic music
- [ ] outdoors
- [ ] art galleries
- [ ] cinemas
- [ ] domestic rooms
- [ ] churches
- [ ] traditional concert halls
- [ ] theaters
- [ ] Other (please specify)

16. At the venue, how much time do you need to adapt/refine your composition to the acoustical conditions?

17. Regarding the venues, what are the main challenges you have faced so far with respect to your compositional aspirations?

CIRMMT Spatialization Questionnaire

8. Compositional Section

18. How do you notate spatialization in your scores? Do your notation methods vary for pre-composed electronics (tape music) and live spatialization? Please specifiy.

19. To improve performer-composer interaction, we are interested in how spatialization is notated in scores and how this varies between pre-composed tape music or live music contexts.

Would you like to provide a notation sample? (i.e. as a .jpg or .pdf file)

- ○ No
- ○ Yes, I'd like to provide a notation sample, please contact me at the following email address

CIRMMT Spatialization Questionnaire

9. Technical Section

20. What software and hardware tools have you used for spatial compositions?

	never heard of it	heard about it but never used it	no longer use it	currently in use	planning to try it
-no software used-	◯	◯	◯	◯	◯
IRCAM Spatialisateur	◯	◯	◯	◯	◯
Vector Based Amplitude Panning (VBAP)	◯	◯	◯	◯	◯
Distance Based Amplitude Panning (DBAP)	◯	◯	◯	◯	◯
First Order Ambisonics, **	◯	◯	◯	◯	◯
Higher Order Ambisonics, **	◯	◯	◯	◯	◯
Wave Field Synthesis, **	◯	◯	◯	◯	◯
Moore's Space Unit Generator	◯	◯	◯	◯	◯
Virtual Microphone Control (ViMiC)	◯	◯	◯	◯	◯
ZKM Zirkonium	◯	◯	◯	◯	◯
GMEM Holophone tools	◯	◯	◯	◯	◯
Waves 360º Surround tools	◯	◯	◯	◯	◯
Vortex Surround tools	◯	◯	◯	◯	◯
Panning within the audio sequencer, **	◯	◯	◯	◯	◯
-no hardware used-	◯	◯	◯	◯	◯
conventional Pan-Pots in mixing consoles	◯	◯	◯	◯	◯
Virtual Surround Panning (VSP) in Studer digital mixer	◯	◯	◯	◯	◯
TC-electronics System 6000	◯	◯	◯	◯	◯
TiMax Audio Imaging systems	◯	◯	◯	◯	◯
Fraunhofer ISONO	◯	◯	◯	◯	◯
sonic emotion M3S	◯	◯	◯	◯	◯

**: Please specify implementation here. If applicable, please also specify other tools not mentioned above and classify.

CIRMMT Spatialization Questionnaire

10. Technical Section

21. What is your motivation for working with your current spatialization equipment (versus using other tools)?

*** 22. Please answer the following questions according to your current choice of spatialization tools.**

	yes	no	N/A
Did you seek professional training to operate your tools?	○	○	○
Do you work with a technical assistant?	○	○	○
Do your tools have more spatialization features than you use?	○	○	○
Do you think the time you need to spend with the tools can be reduced by optimizing them?	○	○	○
Are your compositional goals accomplished by your tools?	○	○	○
Does your software constrain you in the number of sounds you can spatialize simultaneously?	○	○	○

Please comment your answers here:

CIRMMT Spatialization Questionnaire

11. Technical Section

*** 23. For composing spatialized music in your studio, how often do you use the following:**

	never	seldom	sometimes	usually	always
headphones	◯	◯	◯	◯	◯
stereo loudspeakers	◯	◯	◯	◯	◯
Quadrophonic	◯	◯	◯	◯	◯
5.1 or 5.0	◯	◯	◯	◯	◯
7.1 or 7.0	◯	◯	◯	◯	◯
subwoofer	◯	◯	◯	◯	◯
other	◯	◯	◯	◯	◯

24. Further to the previous question, please specify the details of your studio sound system (brand, model, distance of the speaker):

headphones

stereo speaker

Quadrophonic

5.0

7.0

subwoofer

other

*** 25. What is your loudspeaker setup for the performance situation?**

◯ same setting as in my studio

◯ different setting than in my studio; please explain how it differs.

CIRMMT Spatialization Questionnaire

12. Technical Section

*** 26. Rate the importance you attach to the following technical features:**

	not important	slightly	fairly	very	extremely important	N/A
visual 3D representation of a sound scene	◯	◯	◯	◯	◯	◯
spatial rendering in real-time	◯	◯	◯	◯	◯	◯
controllability via a graphical user interface	◯	◯	◯	◯	◯	◯
controllability via external controller	◯	◯	◯	◯	◯	◯
spatial rendering off-line (non-real-time)	◯	◯	◯	◯	◯	◯
recreating the spatial attributes under different room-acoustical conditions	◯	◯	◯	◯	◯	◯
binaural encoding	◯	◯	◯	◯	◯	◯
spatial rendering for non-standardized speaker configurations	◯	◯	◯	◯	◯	◯
integration into audio sequencer as plug-ins	◯	◯	◯	◯	◯	◯
synchronization with movie files	◯	◯	◯	◯	◯	◯

Other (please specify and rate):

CIRMMT Spatialization Questionnaire

13. Technical Section

*** 27. Rate how satisfied you are with the ability of your preferred spatialization tools to produce the following spatial aspects:**

	1 (not satisfied)	2	3	4	5 (extremely satisfied)	N/A
creating of fast movements of sound sources	○	○	○	○	○	○
creating of slow, subtle movements of sound sources	○	○	○	○	○	○
creating of movements along complex trajectories	○	○	○	○	○	○
simulation of Doppler effect	○	○	○	○	○	○
avoiding Doppler effect when moving sound sources	○	○	○	○	○	○
localization accuracy of sound sources	○	○	○	○	○	○
elevated sound sources (height information)	○	○	○	○	○	○
virtual sound sources within the audience	○	○	○	○	○	○
distance perception of sound sources	○	○	○	○	○	○
adaptable apparent source width	○	○	○	○	○	○
simulating of a specific room acoustic	○	○	○	○	○	○
creating of artificial/unreal room impressions	○	○	○	○	○	○
immersiveness	○	○	○	○	○	○
large listening area	○	○	○	○	○	○
simulation of an instrument's directivity	○	○	○	○	○	○

Please comment your answers here:

28. Further to the previous question, is there anything you are missing or you wish you could do with your spatialization tools?

○ No

○ I don't know

○ Yes, (please specify):

CIRMMT Spatialization Questionnaire

14. Technical Section

29. In what consumer media format(s) do you publish your music?

- ☐ Audio CD
- ☐ DVD 7.1
- ☐ Blu-ray Disc
- ☐ Super Audio CD (SACD)
- ☐ DVD DTS encoded
- ☐ mp3
- ☐ DVD 5.1
- ☐ DVD dolby digital encoded
- ☐ mp3-surround™
- ☐ Other (please specify)

*** 30. Do you release your music as binaural/transaural preprocessed versions?**

- ○ I am not sure what that means
- ○ No
- ○ Yes, binaural
- ○ Yes, transaural
- ○ Yes, binaural and transaural

If yes, please specify the software and the set of HRTFs (Head Related Transfer Functions).

Page 14

CIRMMT Spatialization Questionnaire

15. Thank You!

31. If there are any aspects of your approach to the use of spatialization in your music that are not covered by this questionnaire, please outline here.

Page 15

Appendix B

Coding Examples for the Open-ended Questions in Chapter 2

Referring to Table 2.1: Why composers use spatial aspects in their music.
To enhance the listening experience
To affect the listeners' perception of space and time
To make it more interesting, to me and hopefully the audience.
To convey on another level and to intensify the sensory experience for the listener.
To keep things interesting.
It is more natural and also more interesting for the listener to hear sounds from more different directions rather than the traditional stage - audience setup.
Paradigm for artistic expression
Because space is one of the major aspects of sound, at least in my perspective, which focuses on space and matter....
It's an integral part of all sounds and my music.
There is no other way to express the ideas (geography) I am working with.
Music has to do with time and space, simple

For me, real-time signal processing is the way to continue to evolution of orchestration, and spatialization gives the composer the means to expand their gestural pallet into the spatial domain in a dynamic way not previously possible.

To organize and structure sounds

It adds one or more artistic layers to harmonize or create dissonance with other artistic layers in a piece.

Sometimes simply as another musical parameter.

When it comes to the mass of sound, and how the mass is organized, the way of presenting it in a space creates totally different perception in music.

For creating compositional structures.

Counterpoint

To experiment with technology and spatial effects

It offers more possibilities for playing with spatial forms.

To play with fix sounds.

There is still lots of room for innovation.

As a tool for development.

I experimented with grains distributed in an area. It was fun.

Perceptual motivation

It's a natural way of hearing in music, particularly in the electroacoustic field, but even the orchestra has taken that in care.

It's an integral part of all sounds and my music.

It's one of the essential dimensions of the sound.

Space is part of aural perception.

Because listening is a spatial experience and because humans have the capability to understand space through sound.

Segregation and blending of sounds

For enhanced clarity through separation,

Separating sources in order to make complex music more comprehensive,

To make my sources mix together better.

It allows me to present more sound material at the same time without losing sound quality. That is: clarity.

For allowing dense sound complexity yet clear definition.
To separate timbres in space.
To add motion and dynamic
Dramatic and dynamic use of directional cues; increased sense of speed and direction.
The ability to move sounds through space rhythmically and accurately.
To convey movement
To add motion
Creating density and movement.
To make sounds more natural and vivid
To give an auditory experience related to reality.
For giving sounds an identity.
Allows the music to become “alive” in concert.
There is no reason for sonic arts to be so artificially constrained.
For simulating reality.

Referring to Table 2.4: What are the main challenges of venues you have faced so far with respect to your compositional aspiration?
Acoustical conditions
Taking local acoustic into account
Greatly unequal dimensions
More reverberant halls will sometimes pose problems with feedback of the sound back into the computer processing
Unfavourable acoustic conditions in the given venue (too much or too little reverb)
Spaces that are far to live to work
Keeping a virtual space the same while playing it on different number of speakers and in concert rooms different in shape and size.
Venues add lots of reverb to the mix.
Unsuited, echoing spaces for electroacoustic music.
Within gallery/conference type presentations, sonic leakage from one exhibit to another is problematic.
Technical limitations of the venue
Dealing with the sound system that is installed in the concert hall.
Not enough speakers, not enough quality speakers, not enough cables, mixing consoles are often designed for stereo thus making “creative” routing necessary
In general hired systems never fit too well in existing venues, and physical placement of speakers is always tricky and a compromise at best.
If every venue had 8 Genelecs for me to plug in my output mixer I would have no challenges!
As I have moved more towards visual arts, I have discovered that even getting adequate stereo playback in a venue is problematic.
Poor speakers.
Having the same electrical phase for light and sound (hum and buzz...), having a poor quality system that is not eq-able, having a sound system power not in relationship with the size of the room, having to hide the speakers, ...
Time constraints
Set-up time before, during, after concerts

Lacking time
Having no time to make the sound.
Having a technical rehearsal time shared by light and sound
Non-ideal loudspeaker and audience location
Seating is always too close to the walls and flying speakers is almost always impossible in traditional concert spaces and opera theatres.
Raked and/or fixed seating.
Bad quality loudspeakers fixed in the wrong location.
Staff and audience
Deaf sound engineers, or at least unable to understand the aesthetic point... or unable to leave their habits (e.g. using compression)
Changes made by the managers of the venue who have, without warning repositioned the audience or placed extra furniture in the space.
Sweet spot
Some of the hall really doesn't create a "good" spatial aspect if the audience is not sitting on the center.
The focus point of a spatialization system is usually way too small...
Near impossibility of uniform audience placement versus sweet-spot(s).
To avoid "holes" in the listening space (i.e. between loudspeakers)
Cost of production
Scheduling and costs.
My absolute minimum surround setup uses 4 speakers effectively doubling rental costs).

Referring to Table 2.6: What is your motivation for working with your current spatialization equipment?
Usability, learning curve
Simplicity and reactivity close by an instrument (with hardware faders).
Simplicity
It's the more appropriate to my aesthetic needs or I didn't have the time to try out other systems.
IRCAM Spat is easy.
Ease of use.
Optimization and familiarity to control them
Ease of use, flexible, reliable
Quality of spatialization, fit to aesthetic goals
It sounds great.
Built to express my ideas. Tool building and music making happen together and depend upon each other.
Comfort in terms of confidence in using tools, also I feel I still have creative potential to reach with them.
I cannot forget the moment that I tried my sound sources with the WFS system. I felt like a dead sound becomes alive.
Quality of spatialization.
I have a system that does everything I am interested in doing, so no need to change.
Availability, accessibility and cost
I can have them.
Logistical constraints.
It's all I have access to.
Availability where I work.
Access at my job. The most important: free, active community.
Ease of access I guess.
Flexibility, versatility
They are adapted to my compositional style and they are easier to adapt to a variety of projects.
flexible
It is *very* versatile and I can adapt my work to any set up.
They provide different levels of use and results.

	The ZKM Zirkonium is kind of speakers independent. It means, that one can make a complex spatialisation work that is designed for a specific speakers configuration, that can be adapted quite easily to a different venue equipped with a different speakers configurations.
Integration into existing technical framework	
	I am using the RME Fireface, as far as specialization goes, it is neutral.
	It's intuitive, integrated in my production environment, and it works.
	It's integrates into Max/MSP.
Reliability	
	I find it to be a dependable interface.
	It works.
	reliable.
	It works easily.

Appendix C

Jamoma Implementations

During the course of this dissertation, a number of Jamoma modules were implemented to serve the needs of controlling and evaluating different spatialization algorithms. All implementations are either available through the github repository[1], or are included in the official Jamoma distribution[2].

Modules

Audio Type, classified through ∼

jmod.audiounit∼ A module to use stereo AudioUnit plug-ins in the Jamoma environment.

jmod.autotalent∼ Automatic pitch correction based on Thomas A. Barand's autotalent algorithm.

jmod.fffb∼ A flexible filterbank that splits the N frequency bands into a multichannel audio signal.

[1] http://github.com/Nilson/ViMiC-and-friends/tree/master, accessed Jun 2010

[2] To see the history of my commit to the Jamoma project, see https://www.ohloh.net/p/jamoma/contributors/117828132826515, accessed Jun 2010.

jmod.interpolate∼ Mixes N audio sources based on their spatial distribution into M audio signals.

jmod.samplePlayer∼ A stereophonic sample player reading audio from a buffer.

jmod.samplePlayer2∼ An extension of jmod.samplePlayer∼ with pitch-shifter and time-stretch option.

jmod.yin∼ A module for IRCAM's pitch-tracking algorithm *yin*.

Spatial Audio Type, classified through `.sur`

jmod.sur.ambipanning∼ Ambisonics equivalent panning, (Neukom and Schacher 2008).

jmod.sur.ambisonics∼ Ambisonics encoding and decoding up to 11th order using ICST Ambisonics externals V2.0, (Schacher 2010).

jmod.sur.ambi.decode∼ Ambisonics decoding module using ICST Ambisonics tools 2.0.

jmod.sur.ambi.encodeM∼ Ambisonics encoding using ICST Ambisonics tools 2.0.

jmod.sur.audiounit∼ A module to use multichannel AudioUnit plug-ins.

jmod.sur.aux∼ An auxiliary bus system for multichannel audio signals.

jmod.sur.input∼ An audio player that handles multichannel .aiff and .wav files.

jmod.sur.meters∼ A VU meter bridge visualization for multichannel audio signals.

jmod.sur.modulate∼ LFO Modulation of multichannel audio signals.

jmod.sur.panner5∼ An ITU 5-channel panning algorithm by Craven (2003).

jmod.sur.race∼ Implementation of the Ambiophonics algorithm by Glasgal (2007).

jmod.sur.vimic∼ Virtual Microphone Control spatialization for one audio source.

jmod.sur.vimic8∼ Virtual Microphone Control spatialization for up to eight audio sources.

jmod.sur.vimic8poly∼ Virtual Microphone Control spatialization for up to eight audio sources using multi-thread processing.

jmod.sur.zoomH2ambi~ Converts audio recordings from a *Zoom H2* field recorder into Ambisonics B-format.

Control Type

jmod.attributeInterface A graphical user interface to manipulate Jamoma attributes.

jmod.arduino An interface module for the popular i/o board Arduino.

jmod.boids3D Bird flight and animal flock simulator based on Eric Singer's *boids3D*.

jmod.logistic Computes the logistic map function $x_{[n]} = x_{[n-1]} * a(1 - x_{[n-1]})$

jmod.scene3D Manipulation and transformations of spatial audio scene descriptions.

jmod.sourceControl An authoring module for spatial audio scenes.

Video Type, classified through %

jmod.gl.waterfall% Visualization of a frequency spectrum over time (waterfall plot).

jmod.tap.ali% Idiosyncratic multiparametric interpolator by Momeni and Wessel (2003). Developed for *TrakHue* (Peters et al. 2007).

jmod.tap.colortrack% A fast colortracker from the Tap.Tools library.

jmod.television% An analog-TV distortion effect.

Components

jcom.idle Signalizes when input message stream is inactive.

jcom.instance2instance A tool to convert instances of an OSC namespace.

jcom.sur.ambi2UHJ~ Converts Ambisonics B-format into UHJ two-channel streams.

jcom.sur.UHJ2ambi~ Converts UHJ two-channel streams into Ambisonics B-format.

Appendix D

Coding Examples for the Analysis of the Exploratory Qualitative Study in Section 7.3

Sound Source Position
The piano is more to the left.
It's more to the front and a little to the left.
The voice is further over on the left.
In B the image seems slightly narrower and 30 degrees to the left.
B is located more front/centre.
Distance and Depth of Sound Sources
The room sounds larger and the piano further away.
The voice sounds further away.
Sounded dryer and closer.
In A, the vocal is "wider" (more channels?) and seems "closer".
The vocals are a coming from the left a bit and a bit bigger/deeper.

Reverberation, Perceived Room Sound (Ambient, Envelopment)
The room sounds larger.
Feels like some of the ambience is coming from behind.
Spatial spread and the feel of the room – it feels like there's a hole on the left where there's no reverb.
C had less reverb in the rear channels.
C has some reverberance in the rear channels (I only really hear left) that gives a bigger sound, but not a bigger room?.
Spread and Width of Sound Sources
much wider
B has the widest "image" in the front.
C has a more narrow width.
The vocals themselves seem 'wider.
Either the spatial spread is less, or something is flipped.
Timbre, Sound Coloration
C is more disorienting and thinner in timbre.
The high's are clearer.
A bit of pitch or timbre difference.
There seems to be more bass at the end and a bigger sound.
The low end on A feels mellow, warmer and rounder, but also muddier.
Clarity, Diffusion, and Phasing
C feels more diffuse.
C has a more stable, centred front image.
A seems to have some strange phase issues right at the beginning of the excerpt.
As it appears to come from outside of the speakers but cannot be easily located.
The image is less defined.
Spatial Coverage of the Scene
Seems more spread spatially, towards the back.
In B the image seems slightly narrower and 30 degrees to the left.
B is more localized to the front, with less movement and events in the rear channels.

	C makes greater use of rear channels.
	It is more narrow, tho vocal does jump forward nicely but the image seems pinched and not vewry enveloping, somewhat unnatural.
Loudness	
	A seems to be a little more shrill or maybe louder.
	The voice in A seems to be a bit closer, or louder and more crisp.
	The voice in C seems to be slightly closer, and louder.
	It seems like the voice in C is slightly further away, or perhaps quieter.
	The voice was farther away and quieter in A.

Bibliography

Allen, J. B. and D. A. Berkley (1979). Image method for efficiently simulating small-room acoustics. *J. Acoust. Soc. Am. 65*(4), 943–950. (cited on pages 6, 84, and 95)

Amatriain, X., T. Höllerer, J. Kuchera-Morin, and S. T. Pope (2007). Immersive Audio and Music in the AlloSphere. In *Proc. of the International Computer Music Conference*, Copenhagen, Denmark, pp. 276–283. (cited on page 7)

American National Standards Institute (1973). *Psychoacoustical terminology.* ANSI S3.20. New York, US: American National Standards Institute. (cited on page 9)

Ando, Y. (1998). *Architectural Acoustics: Blending Sound Sources, Sound Fields, and Listeners.* New York, US: Springer Verlag. (cited on pages 118, 150, and 154)

Baalman, M. (2007). *On Wave Field Synthesis and electro-acoustic music, with a particular focus on the reproduction of arbitrarily shaped sound sources.* Ph. D. thesis, Technical University Berlin, Germany. (cited on page 45)

Barron, M. (1971). The subjective effects of first reflections in concert halls and opera houses - the need for lateral reflections. *J. Sound Vib. 15*, 475–494. (cited on page 123)

Bates, E., G. Kearney, F. Boland, and D. Furlong (2007). Monophonic source localization for a distributed audience in a small concert hall. In *Proc. of the 10th Int. Conf. on Digital Audio Effects*, Bordeaux, France. (cited on page 17)

Batke, J.-M. and F. Keiler (2010). Investigation of robust panning functions for 3D loudspeaker setups. In *Proc. of the 128th AES Convention, Preprint 7979*, London, UK. (cited on page 7)

Bauer, B. (1960). Broadening the area of stereophonic perception. *J. Audio Eng. Soc. 8*(2), 91–94. (cited on pages 121 and 163)

Bayle, F. (2007). Space, and more. *Organised Sound 12*(3), 241–249. (cited on page 5)

Bech, S. (1998). Spatial aspects of reproduced sound in small rooms. *J. Acoust. Soc. Am. 103*, 434–445. (cited on pages 85 and 124)

Bech, S. and N. Zacharov (2006). *Perceptual Audio Evaluation—Theory, Method and Application*. Chichester, UK: John Wiley & Sons. (cited on page 119)

Beck, S. D., J. Patrick, B. Willkie, and K. Malveaux (2006). The Immersive Computer-controlled Audio Sound Theater: Experiments in multi-mode sound diffusion systems for electroacoustic music performance. In *Proc. of the International Computer Music Conference*, New Orleans, US, pp. 649–655. (cited on page 66)

Begault, D., E. Wenzel, and M. R. Anderson (2001). Direct comparison of the impact of head tracking, reverberation, and individualized head-related transfer functions on the spatial perception of a virtual speech source. *J. Audio Eng. Soc. 49*(10), 904 – 916. (cited on page 134)

Beranek, L. L. (2003). *Concert Halls and Opera Houses: Music, Acoustics, and Architecture*. New York, US: Springer Verlag. (cited on pages 1 and 155)

Berg, J. and F. Rumsey (2002). Validity of selected spatial attributes in the evaluation of 5-channel microphone techniques. In *Proc. of the 112th AES Convention, Preprint 5593*, Munich, Germany. (cited on page 125)

Berg, J. and F. Rumsey (2006). Identification of quality attributes of spatial audio by repertory grid technique. *J. Audio Eng. Soc. 54*(5), 365–379. (cited on page 167)

Berkhout, A. J., D. de Vries, and P. Vogel (1993). Acoustic control by wave field synthesis. *J. Acoust. Soc. Am. 93*, 2764–2778. (cited on page 88)

Berman, J. M. (1975). Behaviour of sound in a bounded space. *J. Acoust. Soc. Am. 57*, 1275–1291. (cited on page 84)

Bilsen, F. A. (1977). Pitch of noise signals: Evidence for a "central spectrum". *J. Acoust. Soc. Am. 61*, 150–161. (cited on page 124)

Bishop, C. (2005). *Installation Art: A Critical History*. Tate. (cited on page 1)

Blauert, J. (1997). *Spatial hearing: the psychophysics of human sound localization*. Cambridge, Mass.: MIT Press. (cited on pages 1, 9, 56, 119, 122, 124, and 216)

Blauert, J. (Ed.) (2005). *Communication acoustics*. Berlin, Germany: Springer. (cited on page 1)

Blesser, B. and L. Salter (2006). *Spaces speak, are you listening?: Experiencing aural architecture*. Cambridge, Mass.: MIT Press. (cited on pages 1, 3, 12, 15, and 26)

Blumlein, A. D. (1931). U.K. patent 394,325, Reprinted in Stereophonic Techniques, 1986, AES, New York. (cited on pages 3 and 127)

Boone, M. M., U. Horbach, and W. P. J. de Bruijn (1999). Virtual surround speakers with wave field synthesis. In *106th AES Convention, Preprint 4928*, Munich, Germany. (cited on page 211)

Borg, I. and P. J. F. Groenen (2005). *Modern Multidimensional Scaling - Theory and Applications* (2nd ed.). Springer Series in Statistics. Heidelberg, Germany: Springer-Verlag. (cited on page 195)

Borß, C. and R. Martin (2009). An improved parametric model for perception-based design of virtual acoustics. In *Proc. of the AES 35th Int. Conference*, London, UK. (cited on page 66)

Braasch, J. (2005). A loudspeaker-based 3D sound projection using Virtual Microphone Control (ViMiC). In *Proc. of the 118th AES Convention, Preprint 6430*, Barcelona, Spain. (cited on pages 15 and 78)

Braasch, J. (2006). *Global Reflections*. Kingston, New York, US: Deep Listening DL 34-2006. (cited on page 104)

Braasch, J. (2009). The telematic music system: Affordances for a new instrument to shape the music of tomorrow. *Contemporary Music Review 28*(4), 421–432. (cited on pages xvi and 103)

Braasch, J., C. Chafe, P. Oliveros, and D. V. Nort (2009). Mixing-console design considerations fortelematic music applications. In *Proc. of the 126th AES Convention, Preprint 7942*, New York, US. (cited on page 112)

Braasch, J., N. Peters, and D. L. Valente (2008). A loudspeaker-based projection technique for spatial music applications using virtual microphone control. *Computer Music Journal 32*(3), 55–71. (cited on pages 42, 66, 68, 73, 95, 156, and 165)

Bradley, J. and G. Soulodre (1995). Objective measures of listener envelopment. *The Journal of the Acoustical Society of America 98*, 2590–2597. (cited on page 123)

Bredies, K., N. Mann, J. Ahrens, M. Geier, S. Spors, and M. Nischt (2008). The multi-touch soundscape renderer. In *Proc. of the working conference on Advanced visual interfaces*, Napoli, Italy, pp. 466–469. ACM. (cited on page 6)

Bregman, A. (1990). *Auditory scene analysis: The perceptual organization of sound*. Cambridge, Mass.: MIT press. (cited on page 32)

Bregman, A. (1993). *Thinking in sound: the cognitive psychology of human audition*, Chapter Auditory Scene Analysis: Hearing in Complex Environments. New York, US: Oxford University Press. (cited on page 132)

Brüggen, M. (2001). Coloration and binaural decoloration in natural environments. *Acta Acoustica ACUSTICA 87*, 400–406. (cited on page 124)

Buchholz, J. M. and J. Mourjopoulos (2004). A Computational Auditory Masking Model Based on Signal-Dependent Compression. II. Model Simulations and Analytical Approximations. *Acta Acustica united with Acustica 90*(5), 887–900. (cited on page 150)

Caceres, J.-P. and C. Chafe (2009). Jacktrip: Under the hood of an engine for network audio. In *Proc. of the International Computer Music Conference*, Montreal, Canada, pp. 509–512. (cited on page 103)

Camerer, F. and C. Sodl (2001). Classical Music in Radio and TV - a multichannel challenge; The IRT/ORF Surround Listening Test. *http://www.hauptmikrofon.de, accessed Oct 2010*. (cited on pages 125, 127, 130, and 156)

Carroll, J. D. and J. J. Chang (1970). Analysis of individual differences in multidimensional scaling via an N-way generalization of "Eckart-Young" decomposition. *Psychometrika 35*(3), 283–319. (cited on page 187)

Casdorff, M., C. Kuhn-Rahloff, R. Moser, and M. Rosenthal (2008). A case study on sound reproduction and acoustic enhancement in concert halls using Wave Field Synthesis. In *Proc. of the 124th AES Convention, Preprint 7466.* (cited on page 7)

CBC News (2004). Archeologists discover ice age dwellers' flute. http://www.uni-tuebingen.de/uni/qvo/pm/pm2004/pm824.html, accessed Jun 2010. (cited on page 11)

Chadabe, J. (2001). Preserving performances of electronic music. *Journal of New Music Research 20*(4), 303 – 305. (cited on page 106)

Chang, J. and J. Carroll (1968). How to use MDPREF, a computer program for multidimensional analysis of preference data. *Unpublished report, Bell Telephone Laboratories*. (cited on page 195)

Chion, M. (2009). *Film, a Sound Art (Film and Culture Series).* West Sussex, UK: Columbia University Press. (cited on page 1)

Choi, I. Y., S. B. Chon, B. G. Shinn-Cunningham, and K.-M. Sung (2007). Prediction of perceived quality in multi-channel audio compression coding systems. In *Proc. of the AES 30th International Conference*, Saariselkä, Finland. Audio Eng. Soc. (cited on page 151)

Choisel, S. and F. Wickelmaier (2006, May). Relating auditory attributes of multichannel sound to preference and to physical parameters. In *Proc. of the 120th AES Convention, Preprint 6684.* (cited on page 149)

Chowning, J. M. (1971). The simulation of moving sound sources. *J. Audio Eng. Soc. 19*(1), 2 – 6. (cited on pages 5 and 79)

Christensen, K. B. and T. Lund (1999). Room simulation for multichannel film and music. In *Proc. of the 107th AES Convention, Preprint 4993*, New York, US. (cited on page 108)

Clarke, M. and P. Manning (2009). Valuing our heritage: exploring spatialization through software emulation of Stockhausen's *Oktophonie*. In *Proc. of the International Computer Music Conference*, Montreal, Canada, pp. 179–182. (cited on page 106)

Clozier, C. (2001). The gmebaphone concept and the cybernéphone instrument. *Computer Music Journal 25*(4), 81–90. (cited on page 5)

Cooperstock, J. R., J. Roston, and W. Woszczyk (2004). Broadband networked audio: Entering the era of multisensory data distribution. In *Proc. 18th International Congress on Acoustics*, Paris, France. (cited on page 103)

Cott, J. (1973). *Stockhausen: conversations with the composer*. New York, US: Simon and Schuster. (cited on page 51)

Craven, P. G. (2003). Continuous surround panning for 5-speaker reproduction. In *Proc. of the 24th International AES Conference: Multichannel Audio, The New Reality*, Banff, Canada. (cited on page 254)

Cremer, L. and H. A. Müller (1982). Principles and Applications of Room Acoustics, (translated by T. J. Schultz). *Applied Science Publishers 1*, 17–19. (cited on page 84)

Daniel, J. (2000). *Représentation de champs acoustiques, application à la transmission et à la reproduction de scènes sonores complexes dans un contexte multimédia*. Phd thesis, Universite Paris 6, France. (cited on pages 6 and 42)

Daniel, J., J. B. Rault, and J. D. Polack (1998). Ambisonics encoding of other audio formats for multiple listening conditions. In *Proc. of the 105th AES Convention*, San Francisco, US. (cited on page 128)

Daniel, P. and R. Weber (1997). Psychoacoustical roughness: Implementation of an optimized model. *Acta Acustica united with Acustica 83*(1), 113–123. (cited on page 150)

Davis, M. F. (2003). History of spatial coding. *J. Audio Eng. Soc. 51*(6), 554–569. (cited on page 3)

de Vries, D. (2008). *Wave Field Synthesis*. New York, US: Audio Engineering Society. (cited on page 44)

Delerue, O. (2006). Visualization of perceptual parameters in interactive user interfaces: Application to the control of sound spatialization. In *120th AES Convention, Preprint 6656*, Paris, France. (cited on page 6)

Evans, W., J. Dyreby, S. Bech, S. Zielinski, and F. Rumsey (2009). Effects of loudspeaker directivity on perceived sound quality - a review of existing studies. In *126th AES Convention, Preprint 7745*, Munich, Germany. (cited on page 210)

Färber, P. and P. Kocher (2010). The mobile ambisonics equipment of the ICST. In *Proc. of the International Computer Music Conference*, New York, US, pp. 207–210. (cited on page 45)

Fastl, H. and E. Zwicker (2007). *Psychoacoustics: Facts and Models* (3rd ed.). New York, US: Springer Verlag. (cited on page 150)

Fontana, S., A. Farina, and Y. Grenier (2007). Binaural for popular music: A case of study. In *Proc. of International Conference on Auditory Display*, Montreal, Canada, pp. 85–90. (cited on page 35)

Frank, M. (2009). Perceptual comparison of sound field reproduction methods with different spatial bandwidth. Master's thesis, Graz University of Technology, Graz, Austria. (cited on pages 44 and 171)

Frank, M., F. Zotter, and A. Sontacchi (2008). Localization experiments using different 2D Ambisonics decoders. In *25th VDT-Tonmeistertagung*, Leipzig, Germany. (cited on page 17)

Gabrielsson, A., U. Rosenberg, and H. Sjögren (1974). Judgments and dimension analysis of perceived sound quality of sound-reproducing systems. *J. Acoust. Soc. Am. 55*, 854 – 861. (cited on page 155)

Gamper, D. and P. Oliveros (1998). A performer-controlled live sound-processing system: New developments and implementations of the expanded instrument system. *Leonardo Music Journal 8*(1), 33–38. (cited on page 103)

Garavaglia, J. A. (2009). Raising awareness about complete automation of live-electronics: A historical perspective. In *Proc. of International Conference on Auditory Display*, Copenhagen, Denmark. (cited on page 5)

Gardner, W. and K. Martin (1994). HRTF measurements of a KEMAR Dummy-Head Microphone. http://sound.media.mit.edu/KEMAR.html, accessed Jun 2010. (cited on page 148)

Gasteiger, B. (2009). A tangible user interface for playing virtual acoustics. Master's thesis, Graz University of Technology, Graz, Austria. (cited on page 6)

Geier, M., J. Ahrens, and S. Spors (2008). The SoundScape Renderer: A Unified Spatial Audio Reproduction Framework for Arbitrary Rendering Methods. In *124th AES Convention*, Amsterdam, The Netherlands. (cited on pages 66, 69, 76, and 228)

George, S., S. Zieliński, and F. Rumsey (2006). Feature extraction for the prediction of multichannel spatial audio fidelity. *Audio, Speech, and Language Processing, IEEE Transactions on 14*(6), 1994–2005. (cited on page 148)

Gerzon, M. A. (1973). With-height sound reproduction. *J. Audio Eng. Soc. 21*(1), 2–10. (cited on page 42)

Gilford, C. (1959). The acoustic design of talks studios and listening rooms. *Proc. Inst. of Electrical Eng. 106*, 245–258. (cited on page 154)

Giordano, B. L., S. McAdams, and D. Rocchesso (2010). Integration of acoustical information in the perception of impacted sound sources: The role of information accuracy and exploitability. *Journal of Experimental Psychology: Human Perception and Performance 36*(2), 462–476. (cited on page 151)

Glaser, B. G. and A. L. Strauss (1967). *The discovery of grounded theory: Strategies for qualitative research*. Chicago, US: Aldine. (cited on pages 27 and 169)

Glasgal, R. (2007). 360° localization via 4.x RACE processing. In *Proc. of the 113th AES Convention, Preprint 7301*, New York, US. (cited on page 254)

Griesinger, D. (1997). The Psychoacoustics of Apparent Source Width, Spaciousness and Envelopment in Performance Spaces. *ACUSTICA 83*, 721–731. (cited on page 150)

Griesinger, D. (2001). The Psychoacoustics of Listening Area, Depth, and Envelopment in Surround Recordings, and their relationship to Microphone Technique. In *AES 19th Intern. Conference*, Schloss Elmau, Germany, pp. 182–200. (cited on pages 119 and 219)

Haller, H. P. (1995). *Das Experimental Studio der Heinrich-Strobel-Stiftung des Südwestfunks Freiburg 1971-1989. Die Erforschung der Elektronischen Klangumformung und ihre Geschichte*. Südwestfunk Schriftenreihe. Baden-Baden, Germany: Nomos Verlagsgesellschaft. (cited on page 5)

Halmrast, T. (2000, April). Orchestral Timbre: Comb-Filter Coloration from Reflections. *J. Sound Vib. 232*, 53–69. (cited on page 124)

Hamasaki, K. (2003). Multichannel recording techniques for reproducing adequate spatial impressionsion. In *Proc. of the Audio Engineering Society 24th Int. Conf. on Multichannel Audio: The New Reality*, Banff, Canada. (cited on page 129)

Hamasaki, K., S. Komiyama, H. Okubo, K. Hiyama, and W. Hatano (2004). 5.1 and 22.2 multichannel sound productions using an integrated surround sound panning system. In *Proc. of the 117th AES Convention, Preprint 6226*, San Francisco, US. (cited on page 112)

Hamasaki, K., T. Nishiguchi, R. Okumura, and Y. Nakayama (2007). Wide listening area with exceptional spatial sound quality of a 22.2 multichannel sound system. In *Proc. of the 122th AES Convention, Preprint 7037*, Vienna, Austria. (cited on page 224)

Hannemann, J. (2009). Broadband multipole–matched rendering: The current state of affairs. In *Proc. of the 1st International Symposium on Ambisonics*, Graz, Austria. (cited on page 7)

Harley, M. A. (1994). *Space and Spatialization in Contemporary Music: History and Analysis, Ideas and Implementations*. Ph. D. thesis, McGill University, Montreal, Canada. (cited on page 1)

Harley, M. A. (1998). Spatiality of sound and stream segregation in twentieth century instrumental music. *Organised Sound 3*(2), 147–166. (cited on pages 1 and 32)

Harrison, J. (1999). Sound, space, sculpture: some thoughts on the 'what', 'how' and 'why' of sound diffusion. *Organised Sound 3*(2), 117 – 127. (cited on page 45)

Heyer, P. (2008). Live from the Met: digital broadcast cinema, medium theory, and opera for the masses. *Canadian Journal of Communication 33*, 591–604. (cited on page 158)

Hiebert, G. (2005). *OpenAL 1.1 Specification and Reference*. (cited on pages 69 and 70)

Hiyama, K., S. Komiyama, and K. Hamasaki (2002). The minimum number of loudspeakers and its arrangement for reproducing the spatial impression of diffuse sound field. In *Proc. of the 113th AES Convention, Preprint 5674*, Los Angeles, US. (cited on page 123)

Holman, T. (2002). *Sound for Film and Television*. Focal Press, Boston. (cited on page 112)

Horbach, U., A. Karamustafaoglu, E. Corteel, and R. Pellegrini (2000). Implementation of an auralization scheme in a digital mixing console using perceptual parameters. In *Proc. of the 108th AES Convention, Preprint 5099*, Paris, France. (cited on pages 42 and 105)

Hosokawa, S. (2008). The walkman effect. *Popular Music 4*, 165–180. (cited on page 12)

Huopaniemi, J., L. Savioja, and T. Takala (1996). DIVA virtual audio reality system. In *Proc. Int. Conf. Auditory Display*, Palo Alto, US, pp. 111–116. (cited on page 6)

IEC (1997). *Draft 60268: Sound System Equipment - Part 13: Listening Tests on Loudspeakers*. Geneva, Switzerland: International Electrotechnical Commission. (cited on page 123)

IOSONO (URL 2009). http://www.iosono-sound.com, accessed sept. 2009. (cited on page 42)

Irimajiri, H., T. Kamekawa, and A. Marui (2007). Correspondence relationship between physical factors and psychological impressions of microphone arrays for orchestra recording. In *123th AES Convention, Preprint 7233*, New York, US. (cited on page 125)

ISO 389-8 (2004). *Reference zero for the calibration of audiometric equipment – Part 8: Reference equivalent threshold sound pressure levels for pure tones and circumaural earphones*. Geneva, Switzerland: International Organization for Standardization. (cited on pages 142, 168, and 177)

ITU (1992). *Recommendation BS.775-2, Multichannel stereophonic sound system with and without accompanying picture*, Volume 2006. Geneva, Switzerland: International Telecommunication Union. (cited on pages 4, 17, 65, 92, 112, 120, and 161)

ITU (1997). *Recommendation BS.1116-1, Methods for the Subjective Assessment of Small Impairments in Audio Systems Including Multichannel Sound Systems*. Geneva, Switzerland: International Telecommunication Union. (cited on pages 17, 120, 121, 135, 139, 141, and 165)

Jensenius, A. R. (2007). *ACTION - SOUND Developing Methods and Tools to Study Music-Related Body Movement*. Ph. D. thesis, University of Oslo, Oslo, Norway. (cited on page 91)

Jot, J. and A. Chaigne (1991). Digital delay networks for designing artificial reverberators. In *90th AES Convention, Preprint 3030*, Paris, France. (cited on page 86)

Jot, J.-M. (1992). *Etude et Réalisation d'un Spatialisateur de Sons par Modèles Physiques et Perceptifs*. Ph. D. thesis, France Telecom, Paris, France. (cited on pages 42, 43, 66, and 73)

Jot, J.-M. (1997). Efficient models for reverberation and distance rendering in computer music and virtual audio reality. *Proc. of the International Computer Music Conference*. (cited on page 5)

Jot, J.-M. (1999). Real-time spatial processing of sounds for music, multimedia and interactive human-computer interfaces. *Multimedia Systems 7*(1), 55–69. (cited on pages 43, 80, and 228)

Kamekawa, T. (2006). The effect on spatial impression of the configuration and directivity of three frontal microphones used in multi-channel stereophonic recording. In *Proc. of the AES 28th International Conference*, Piteå, Sweden. (cited on page 125)

Kang, J. (2006). *Urban sound environment*. London, UK: Taylor Francis. (cited on page 1)

Kassier, R., H.-K. Lee, T. Brookes, and F. Rumsey (2005). An informal comparison between surround-sound microphone techniques. In *Proc. of the 118th AES Convention, Preprint 6429*, Barcelona, Spain. (cited on page 125)

Keele, D. B. and D. J. Button (2005). Ground-plane constant beamwidth transducer (CBT) loudspeaker circular-arc line arrays. In *Proc. of the 119th AES Convention, Preprint 6594*, New York, US. (cited on page 211)

Kendall, G. S. and M. Ardila (2008). *The Artistic Play of Spatial Organization: Spatial Attributes, Scene Analysis and Auditory Spatial Schemata*, Volume 4969 of *Lecture Notes in Computer Science: Computer Music Modeling and Retrieval. Sense of Sounds*, pp. 125–138. Springer-Verlag Berlin Heidelberg. (cited on pages 9 and 41)

Kendall, G. S., N. Peters, and M. Geier (2008). Towards an interchange format for spatial audio scenes. In *Proc. of the International Computer Music Conference*, Belfast, UK, pp. 295–296. (cited on pages 14, 59, and 76)

Kim, S., M. de Francisco, K. Walker, A. Marui, and W. L. Martens (2006). An examination of the influence of musical selection on listener preferences for multichannel microphone technique. In *Proc. of the AES 28th International Conference*, Piteå, Sweden. (cited on pages 125, 127, 130, 156, and 164)

Kim, S., W. L. Martens, A. Marui, and K. Walker (2006). Predicting listener preferences for surround microphone technique through binaural signal analysis of loudspeaker-reproduced piano performances. In *Proc. of the 121th AES Convention, Preprint 6919*, San Francisco, US. (cited on page 149)

Klapholz, J. (1991). Fantasia: Innovations in sound. *J. Audio Eng. Soc. 39*(1/2), 66–70. (cited on page 3)

Koenig, W. (1950). Subjective Effects in Binaural Hearing. *J. Acoust. Soc. Am. 22*, 61–62. (cited on page 124)

Kruskal, B. and M. Wish (1978). *Multidimensional Scaling*. Beverly Hills: Sage Press. (cited on page 187)

Kuttruff, H. (2009). *Room Acoustics*. London, UK: Spon Press. (cited on pages 124, 135, and 177)

Landone, C. and M. Sandler (2001). Surround sound impact over large areas. In *Proc. of the AES 110th Convention, Preprint 5373*, Amsterdam, The Netherlands. Audio Eng. Soc. (cited on page 120)

Lee, H.-K. and F. Rumsey (2005). Investigation into the effect of interchannel crosstalk in multichannel microphone technique. In *Proc. of the 118th AES Convention, Preprint 6374*, Barcelona, Spain. Audio Eng. Soc. (cited on pages 123 and 154)

Letowski, T. (1989). Sound quality assessment: concepts and criteria. In *Proc. of the 87th AES Convention, preprint 2825*, New York, US. (cited on page 9)

Lipshitz, S. P. (1986). Stereo microphone techniques: Are the purists wrong? *J. Audio Eng. Soc. 34*(9), 716 – 744. (cited on page 4)

Litovsky, R. Y., H. S. Colburn, W. A. Yost, and S. J. Guzman (1999). The precedence effect. *J. Acoust. Soc. Am. 106*(4), 1633–1654. (cited on page 122)

Lombardo, V., A. Arghinenti, F. Nunnari, A. Valle, H. Vogel, J. Fitch, R. Dobson, J. Padget, K. Tazelaar, S. Weinzierl, et al. (2005). The virtual electronic poem (VEP) project. In *Proc. International Computer Music Conference*, Barcelona, Spain, pp. 451–454. (cited on page 6)

Lossius, T., P. Baltazar, and T. de la Hogue (2009). DBAP - Distance-Based Amplitude Panning. In *Proc. of the International Computer Music Conference*, Montreal, Canada. (cited on pages 42, 68, 73, 80, and 88)

Lund, T. (2000). Enhanced localization in 5.1 production. In *Proc. of the 109th AES Convention, Preprint 5243*, Los Angeles, US. (cited on page 123)

Lyon, E. (2008). Spatial orchestration. In *Proc. of the 5th Sound and Music Computing Conference*, Berlin, Germany. (cited on page 36)

Maconie, R. (2005). *Other Planets: The Music of Karlheinz Stockhausen*. Scarecrow Press. (cited on pages 4 and 107)

Malham, D. G. (1992). Experience with large area 3-D ambisonic sound systems. In *Institute of Acoustics Autumn Conference on Reproduced Sound 8*, Windermere. (cited on page 7)

Marentakis, G., E. Corteel, and S. McAdams (2008). Wave field synthesis evaluation using the minimum audible angle in a concert hall. In *Proc. of the 124th AES Convention*, Amsterdam, The Netherlands. (cited on page 17)

Marentakis, G., J. Malloch, N. Peters, M. T. Marshall, S. McAdams, and M. M. Wanderley (2008). The influence of the performer's gesture on the identification of spatial sound trajectories in a concert hall. In *Proc. of the International Conference on Auditory Display (ICAD)*, Paris, France. (cited on pages 7 and 19)

Marentakis, G., N. Peters, and S. McAdams (2007a). Auditory resolution in auditory virtual environments (A). *J. Acoust. Soc. Am. 122*(5), 3054. (cited on page 19)

Marentakis, G., N. Peters, and S. McAdams (2007b). DJ SPAT: Spatialized interactions for DJs. In *Proc. of the International Computer Music Conference*, Volume 2, Copenhagen, Denmark, pp. 360–363. (cited on page 16)

Marentakis, G., N. Peters, and S. McAdams (2008a). Auditory resolution in virtual environments: Effects of spatialization algorithm, off-center listener positioning and speaker configuration (A). *J. Acoust. Soc. Am. 123*(5), 3798. (cited on pages 19, 56, and 125)

Marentakis, G., N. Peters, and S. McAdams (2008b). Making spatial music accessible: Investigating elementary spatial movements. *Echoes - Newsletter of the Acoustical Society of America 18*(2), 3 – 4. (cited on page 19)

Marketo, C., T. Cooper, M. Petty, and F. Scriven (1994). The reliability of mdpref to show individual preference. *Journal of Sensory Studies 9*(3), 337–350. (cited on page 195)

Marshall, M., N. Peters, A. Jensenius, J. Boissinot, M. Wanderley, and J. Braasch (2006). On the Development of a System for Gesture Control of Spatialization. In *Proc. of the International Computer Music Conference*, New Orleans, US, pp. 360–366. (cited on pages 6, 16, and 48)

Martin, G. (online version Oct. 2006). *Introduction to Sound Recording.* available at http://www.tonmeister.ca/main/textbook. (cited on page 119)

Mason, R., T. Brookes, and F. Rumsey (2005). Frequency dependency of the relationship between perceived auditory source width and the interaural cross-correlation coefficient for time-invariant stimuli. *J. Acoust. Soc. Am. 117*(3), 1337–1350. (cited on page 150)

McAdams, S. and B. L. Giordano (2009). *The Oxford handbook of music psychology*, Chapter 7 - The perception of musical timbre, pp. 72–80. Oxford, UK: Oxford University Press. (cited on page 172)

McGinity, M., J. Shaw, V. Kuchelmeister, A. Hardjono, D. D. Favero, and A. Hardjono (2007). AVIE: a versatile multi-user stereo 360° interactive VR theatre. In *Proc. of the Workshop on Emerging Displays Technologies*, New York, US. (cited on page 98)

McLeran, A., C. Roads, B. L. Sturm, and J. J. Shynk (2008). Granular sound spatialization using dictionary-based methods. In *Proc. of the 5th Sound and Music Computing Conference*, Berlin, Germany. (cited on page 66)

Meltzer, S., L. Altmann, A. Gräfe, and J.-O. Fischer (2008). An object oriented mixing approach for the design of spatial audio scenes. In *Proc. of the 25th Tonmeistertagung*, Leipzig, Germany. (cited on page 76)

Merchel, S. and S. Groth (2009). Analysis and implementation of a stereophonic play back system for adjusting the "sweet spot" to the listener's position. In *Proc. of the 126th AES Convention, Preprint 7726*, Munich, Germany. (cited on page 163)

Meyer, J. (2009). *Acoustics and the Performance of Music* (5th ed.). Springer, originally published by PPVMedien GmbH, Edition Bochinsky, Bergkirchen, Germany. (cited on pages 32 and 111)

Mills, A. (1958). On the minimum audible angle. *J. Acoust. Soc. Am. 30*, 237–346. (cited on page 19)

Møller, H., D. Hammershøi, C. Jensen, and M. Sørensen (1999). Evaluation of artificial heads in listening tests. *J. Audio Eng. Soc. 47*(3), 83–100. (cited on page 132)

Momeni, A. and D. Wessel (2003). Characterizing and controlling musical material intuitively with geometric models. In *Proc. of the Conference on New Interfaces for Musical Expression*, Montreal, Canada, pp. 54–62. (cited on page 255)

Moore, F. R. (1983). A General Model for Spatial Processing of Sounds. *Computer Music Journal 7*(6), 6–15. (cited on pages 6, 42, 79, 80, and 88)

Moorer, J. (1979). About This Reverberation Business. *Computer Music Journal 3*(2), 13–28. (cited on page 6)

Morimoto, M. (1997). The Role of Rear Loudspeakers in Spatial Impression. In *Proc. of the 103th AES Convention, Preprint 4554*, New York, US. Audio Eng. Soc. (cited on page 123)

Muraoka, T. and T. Nakazato (2007). Examination of multichannel sound-field recomposition utilizing frequency-dependent interaural cross correlation (FIACC). *J. Audio Eng. Soc. 55*(4), 236–256. (cited on page 123)

Neher, T. (2004). *Towards A Spatial Ear Trainer*. Ph. D. thesis, School of Arts, University of Surrey, UK. (cited on pages 9 and 110)

Neukom, M. and J. Schacher (2008). Ambisonics equivalent panning. In *Proc. of the International Computer Music Conference*, Belfast, UK, pp. 592 – 595. (cited on pages 68, 72, and 254)

Normandeau, R. (2009). Timbre spatialisation: The medium is the space. *Organised Sound 14*(03), 277–285. (cited on page 50)

Okano, T., L. Beranek, and T. Hidaka (1998). Relations among interaural cross-correlation coefficient ($IACC_E$), lateral fraction (LF_E), and apparent source width (ASW) in concert halls. *J. Acoust. Soc. Am. 104*, 255–265. (cited on pages 123 and 150)

Olive, S. E. (2004). A Multiple Regression Model for Predicting Loudspeaker Preference Using Objective Measurements: Part II - Development of the Model. In *Proc. of the 117th AES Convention, Preprint 6190*, San Francisco, US. Audio Eng. Soc. (cited on page 150)

Olive, S. E. (2008). *Interaction Between Loudspeakers and Room Acoustics Influences Loudspeaker Preferences in Multichannel Audio Reproduction.* Ph. D. thesis, McGill University, Montreal, Canada. (cited on pages 10, 134, 154, 209, and 227)

Olive, S. E., P. L. Schuck, S. L. Sally, and M. Bonneville (1995). The variability of loudspeaker sound quality among four domestic-sized rooms. In *Proc. of the 99th AES Convention, Preprint 4092*, New York, US. Audio Eng. Soc. (cited on pages 133 and 209)

Olive, S. E., T. Welti, and W. L. Martens (2007). Listener loudspeaker preference ratings obtained in situ match those obtained via a binaural room scanning measurement and playback. In *122th AES Convention, Preprint 7034*, Vienna, Austria. (cited on pages 134 and 227)

Otondo, F. (2007). Creating spaces: an interview with Natasha Barrett. *Computer Music Journal 31*(2), 10 – 19. (cited on pages 13 and 26)

Otondo, F. (2008). Contemporary trends in the use of space in electroacoustic music. *Organised Sound 13*(01), 77–81. (cited on pages 26 and 64)

Pachet, F. and O. Delerue (1999). Musicspace: a constraint-based control system for music spatialization. In *Proc. of the International Computer Music Conference*, Beijing, China, pp. 272–275. (cited on page 93)

Peeters, G. (2004). A large set of audio features for sound description (similarity and classification) in the CUIDADO project. *CUIDADO Project Report*. (cited on page 150)

Pellegrini, R. (2002). Perception-based design of virtual rooms for sound reproduction. In *22nd AES International Conference, Preprint 000245*, Espoo, Finland. (cited on page 84)

Perrott, D. R. and A. D. Musicant (1977). Minimum auditory movement angle: Binaural localization of moving sound sources. *J. Acoust. Soc. Am. 62*, 1463. (cited on page 19)

Pestova, X., M. T. Marshall, and J. Sudol (2008). Analogue to digital: Authenticiy vs. sustainability in Stockhausen's MANTRA (1970). In *Proc. of the International Computer Music Conference*, Belfast, UK, pp. 201–204. (cited on page 106)

Peters, N. (2008). Proposing SpatDIF - The Spatial Sound Description Interchange Format. In *Proc. of the International Computer Music Conference*, Belfast, UK. (cited on pages 14, 69, 70, and 93)

Peters, N., M. Evans, and E. Britton (2007). Trakhue - intuitive gestural control of live electronics. In *Proc. of the International Computer Music Conference*, Volume 1, Copenhagen, Denmark, pp. 500–503. (cited on pages 16 and 255)

Peters, N., S. Ferguson, and S. McAdams (2007, September). Towards a Spatial Sound Description Interchange Format (SpatDIF). *Canadian Acoustics 35*(3), 64 – 65. (cited on pages 14 and 93)

Peters, N., T. Lossius, J. Schacher, P. Baltazar, C. Bascou, and T. Place (2009). A stratified approach for sound spatialization. In *Proc. of the 6th Sound and Music Computing Conference*, Porto, PT, pp. 219–224. (cited on page 91)

Peters, N., G. Marentakis, and S. McAdams (2011). Current technologies and compositional practices for spatialization: A qualitative and quantitative analysis. *to appear in Computer Music Journal 35*(1). (cited on pages 64, 90, and 94)

Peters, N., T. Matthews, J. Braasch, and S. McAdams (2008). Spatial Sound Rendering in Max/MSP with ViMiC. In *Proc. of the International Computer Music Conference*, Belfast, UK, pp. 755–758. (cited on pages 91, 95, and 165)

Peutz, V. M. A. (1978). The Variable Acoustics of the Espace de Projection of Ircam (Paris). In *Proc. of the 59th AES Convention, Preprint 1310*, Hamburg, Germany. (cited on page 7)

Pfanzagl-Cardone, E. and R. Höldrich (2008). Frequency dependent signal correlation in surround- and stereo-microphone systems and the blumlein-pfanzagl-triple (BPT). In *Proc. of the 124th AES Convention, Preprint 7476*, Amsterdam, The Netherlands. (cited on page 133)

Place, T. and T. Lossius (2006). Jamoma: A modular standard for structuring patches in Max. In *Proc. of the International Computer Music Conference 2006*, New Orleans, US, pp. 143–146. (cited on pages 15, 72, 91, and 96)

Place, T., T. Lossius, A. R. Jensenius, and N. Peters (2008). Flexible control of composite parameters in Max/MSP. In *Proc. of the International Computer Music Conference*, Belfast, UK, pp. 233–236. (cited on page 16)

Place, T., T. Lossius, A. R. Jensenius, N. Peters, and P. Baltazar (2008). Addressing Classes by Differentiating Values and Properties in OSC. In *Proc. of the Conference on New Interfaces for Musical Expression*, Geneva, Switzerland, pp. 181–184. (cited on pages 16, 93, and 225)

Place, T., T. Lossius, and N. Peters (2010a). A flexible and dynamic C++ framework and library for digital audio signal processing. In *Proc. of the International Computer Music Conference*, New York, US, pp. 157–164. (cited on pages 16, 91, and 225)

Place, T., T. Lossius, and N. Peters (2010b). The Jamoma Audio Graph Layer. In *Proc. of the 13th Int. Conference on Digital Audio Effects*, Graz, Austria. (cited on pages 16, 225, and 228)

Poletti, M. (2007). Robust two-dimensional surround sound reproduction for nonuniform loudspeaker layouts. *J. Audio Eng. Soc. 55*(7), 598–610. (cited on page 7)

Pottier, L. (1998). Dynamical spatialisation of sound. HOLOPHON: a graphical and algorithmical editor for Sigma 1. In *Proc. of the International Conference on Digital Audio Effects*, Barcelona, Spain. (cited on pages 6, 42, and 74)

Pottier, L. and O. Stalla (2000). *Trends In Gestural Control Of Music*, Chapter Interpretation and Space. Paris, France: IRCAM–Centre Pompidou. (cited on page 6)

Pulkki, V. (1997). Virtual sound source positioning using vector base amplitude panning. *J. Audio Eng. Soc. 45*(6), 456 – 466. (cited on page 68)

Pulkki, V. (2000). Generic panning tools for MAX/MSP. In *Proc. of the International Computer Music Conference*, pp. 304–307. (cited on pages 73 and 79)

Pulkki, V. (2001a). Localization of amplitude-panned virtual sources, part 2: Two and three dimensional panning. *J. Audio Eng. Soc. 49*, 753 – 767. (cited on page 6)

Pulkki, V. (2001b). *Spatial Sound Generation and Perception by Amplitude Panning Techniques*. Ph. D. thesis, Helsinki University of Technology, Helsinki, Finland. (cited on pages 42 and 43)

Pulkki, V. (2002). Microphone techniques and directional quality of sound reproduction. In *112th AES Convention, Preprint 5500*, Munich, Germany. (cited on page 125)

Pulkki, V. (2007). Spatial sound reproduction with directional spatial sound reproduction with directional audio coding. *J. Audio Eng. Soc. 55*(6), 503–516. (cited on page 68)

Pulkki, V. and T. Hirvonen (2005). Localization of virtual sources in multichannel audio reproduction. *IEEE Transaction on Speech and Audio Processing 13*(1), 105 – 119. (cited on pages 56 and 83)

Rabenstein, R., S. Spors, and P. Steffen (2004). *Wave Field Synthesis Techniques for Spatial Sound Reproduction*, Chapter 13, pp. 517 – 545. Topics in Acoustic Echo and Noise Control. Heidelberg, Germany: Springer. (cited on pages 6, 42, and 44)

Rakerd, B., W. Hartmann, and J. Hsu (2000). Echo suppression in the horizontal and median sagittal planes. *J. Acoust. Soc. Am. 107*, 1061–1064. (cited on page 124)

Ramakrishnan, C. (2009). Zirkonium: Non-invasive software for sound spatialisation. *Organised Sound 14*(3), 268 – 276. (cited on page 42)

Ramakrishnan, C., J. Goßmann, and L. Brümmer (2006). The ZKM Klangdom. In *Proc. of the Conference on New Interfaces for Musical Expression*, Paris, France, pp. 140–143. (cited on pages 7 and 66)

Reiter, U. (2007). Tanga - an interactive object-based real time audio engine. In *Conference Proceedings - Audio Mostly 2007*. (cited on page 75)

Rose, J., P. Nelson, T. Takeuchi, and B. Rafaely (2001). Variance of sweet spot size with head location for virtual audio. In *Proc. of the 110th AES Convention, Preprint 5391*, Amsterdam, The Netherlands. (cited on page 120)

Ruby, S., D. Thomas, and D. H. Hansson (2009). *Agile Web Development with Rails* (3rd ed.). The Pragmatic Programmers LLC. (cited on page 92)

Rumsey, F. (2001). *Spatial Audio*. Oxford, UK: Focal Press. (cited on pages 104, 110, 123, and 127)

Rumsey, F. (2002). Spatial quality evaluation for reproduced sound: Terminology, meaning, and a scene-based paradigm. *J. Audio Eng. Soc. 50*(9), 651 – 666. (cited on pages 9 and 41)

Rumsey, F., S. Zieliński, R. Kassier, and S. Bech (2005a). On the relative importance of spatial and timbral fidelities in judgments of degraded multichannel audio quality. *J. Audio Eng. Soc. 118*(2), 968–976. (cited on pages 124 and 155)

Rumsey, F., S. Zieliński, R. Kassier, and S. Bech (2005b). Relationships between experienced listener ratings of multichannel audio quality and naïve listener preferences. *J. Acoust. Soc. Am. 117*(6), 3832–3840. (cited on page 124)

Sacks, O. W. (2007). *Musicophilia - Tales of Music and the Brain*. New York, US: Random House. (cited on page 1)

Schacher, J. C. (2007). Gesture control of sounds in 3D space. In *Proc. of the Conference on New Interfaces for Musical Expression*, New York, US, pp. 358—362. (cited on page 6)

Schacher, J. C. (2010). Seven years ISCT ambisonics tools for MaxMSP - a brief report. In *Proc. of the 2nd International Symposium on Ambisonics and Spherical Acoustics*, Paris, France. (cited on pages 44, 80, and 254)

Schacher, J. C. and P. Kocher (2006). Ambisonics Spatialization Tools for Max/MSP. In *Proc. of the International Computer Music Conference*, New Orleans, US, pp. 274 – 277. (cited on pages 71, 73, and 80)

Schmeder, A., A. Freed, and D. Wessel (2010). Best practices for Open Sound Control. In *Proc. of the Linux Audio Conference*, Utrecht, The Netherlands, pp. 131–139. (cited on page 93)

Schroeder, M. R., D. Gottlob, and K. Siebrasse (1974). Comparative study of european concert halls: Correlation of subjective preference with geometric and acoustical parameters. *J. Acoust. Soc. Am. 56*, 1195–1201. (cited on page 133)

Schuck, P. L., S. E. Olive, J. G. Ryan, F. E. Toole, S. L. Sally, M. Bonneville, K. Momtahan, and E. Verreault (1993). Perception of reproduced sound in rooms: Some results from the athena project. In *Proc. of the 12th Int. Conference of the Audio Eng. Soc.*, Copenhagen, Denmark. (cited on pages 133 and 209)

Schwarz, D. and M. Wright (2000). Extensions and Applications of the SDIF Sound Description Interchange Format. In *Proc. of the International Computer Music Conference*, Berlin, Germany, pp. 481–484. (cited on page 71)

Smalley, D. (2007). Space-form and the acousmatic image. *Organised Sound 12*(01), 35–58. (cited on page 1)

Snedecor, G. W. and W. G. Cochran (1980). *Statistical Methods* (7th ed.). Ames, Iowa: Iowa State Press. (cited on page 136)

Sonic Emotion (URL 2009). M3S, http://www.sonicemotion.com, accessed Sept. 2009. (cited on page 42)

Soulodre, G. A., M. C. Lavoie, and S. G. Norcross (2003). Objective measures of listener envelopment in multichannel surround system. *J. Audio Eng. Soc. 51*(9), 826–840. (cited on page 123)

Spors, S. and J. Ahrens (2008). A comparison of wave field synthesis and higher-order ambisonics with respect to physical properties and spatial sampling. In *Proc. of the 125th AES Convention, Preprint 7556*, San Francisco, US. (cited on page 44)

Stallmann, K. (2007, October). Songs Inside My Head: ICAD 2007. *Newsletter of the Society for Electro-Acoustic Music in the United States 4*, 14–15. (cited on page 104)

Steinberg, J. and W. Snow (1934). Auditory Perspective-Physical Factors. *Electrical Engineering 53*(1), 3–7. (cited on pages 3 and 79)

TC Electronics (URL 2009). System 6000, http://www.tcelectronic.com/System6000.asp, accessed Sept. 2009. (cited on page 42)

Theile, G. (1980, April). *On the localisation in the superimposed soundfield.* Ph. D. thesis, Technical University Berlin, Germany. (cited on page 124)

Theile, G. (2001). Multichannel Natural Music Recording Based on Psychoacoustic Principles. In *Proc. of the AES 19th International Conference on Surround Sound: Techniques, Technology, and Perception*, Schloss Elmau, Germany, pp. 201–229. (cited on pages 95, 128, and 219)

Thompson, E. A. (2004). *The soundscape of modernity: architectural acoustics and the culture of listening in America, 1900-1933*. Cambridge, Mass.: The MIT Press. (cited on page 3)

TiMax (URL 2009). Audio Imagine System, http://outboard.co.uk/pages/timax.htm, accessed Sept. 2009. (cited on page 42)

Tintinnabulate & Soundwire, J. Braasch, C. Chafe, P. Oliveros, and B.Woodstrup (2009). *Tele-Colonization*. Kingston, New York, US: Deep Listening DL-TMS/DD-1. (cited on page 104)

Toole, F. E. (1991). Binaural record/reproduction systems and their use in psychoacoustic investigations. In *Proc. of the 91st AES Convention, Preprint 3179*, New York, US. Audio Eng. Soc. (cited on page 132)

Toole, F. E. (2008). *Sound reproduction: loudspeakers and rooms*. Oxford, UK: Focal Press. (cited on pages 16, 17, 121, 123, and 125)

Tyler, R. S., A. E. Perreau, and H. Ji (2009). Validation of the spatial hearing questionnaire. *Ear and Hearing 30*(4), 466–475. (cited on page 223)

Vaananen, R. and J. Huopaniemi (2004). Advanced AudioBIFS: virtual acoustics modeling in MPEG-4 scene description. *Multimedia, IEEE Transactions on 6*(5), 661–675. (cited on page 70)

Valente, D., D. Lewis, J. Franco, and E. Heinrichs (2010). Comparing head rotation angle, visual localization, and recall-proficiency of school-aged children 8–12 while listening to a story by multiple discreet talkers in a virtual classroom. (A). *J. Acoust. Soc. Am. 127*, 1822. (cited on page 111)

Valente, D. L. and J. Braasch (2006). A Hybrid Approach to Rendering Multi-Channel Audio, Using Wave Field Synthesis and Stereophonic techniques, with Virtual Microphone Control (A). *J. Acoust. Soc. Am. 119*, 3281. (cited on page 88)

Varèse, E. (1936). *Audio Culture: Readings in Modern Music (2004)*, Chapter 4: The Liberation of Sound, pp. 17–22. New York, US: Continuum. (cited on page 4)

Vinet, H. (2007). Science and technology of music and sound: the IRCAM roadmap. *Journal of New Music Research 36*(3), 207–226. (cited on page 7)

Vortex (URL 2009). surround tools, http://im-research.com/products, accessed Sept. 2009. (cited on page 42)

Wakefield, G. (2006). Third-Order Ambisonic Extension for Max/MSP with Musical Applications. In *Proc. of the International Computer Music Conference*, New Orleans, US, pp. 123–126. (cited on page 80)

Waves (URL 2009). 360° surround tools, http://www.waves.com/content.aspx?id=265, accessed Sept. 2009. (cited on page 42)

Weber, R. (1990). The continuous loudness judgement of temporally variable sounds with an "analog" category procedure. In *Results of the third Oldenburg Symposium on Psychological Acoustics. A. Schick.*, Oldenburg, Germany, pp. 267–294. (cited on page 131)

Weinzierl, S., H. Rosenheinrich, J. Blickensdorff, M. Horn, and A. Lindau (2010). Die Akustik der Konzertsäle im Leipziger Gewandhaus. Geschichte, Rekonstruktion und Auralisation (Acoustics of the concert halls in the Leipzig Gewandhaus. history, reconstruction and auralization). In *Fortschritte der Akustik: DAGA Tagungsband*, Berlin, Germany. (cited on page 6)

Weinzierl, S. and K. Tazelaar (2006). *sonambiente berlin 2006. klang kunst sound art*, Chapter Raumsimulation und Klangkunst. Vom künstlichen Nachhall zur virtuellen Akustik (Room simulation and sound art. From artificial reverb to virtual acoustics). Heidelberg, Germany: Kehrer. (cited on page 3)

Wiggins, B. (2004). *An Investigation into the Real-Time Manipulation and Control of Three-Dimensional Sound Fields.* Ph. D. thesis, University of Derby, Derby, UK. (cited on page 68)

Williams, M. and G. Le Dû (2004). The Quick Reference Guide to Multichannel Microphone Arrays, Part 2: using Supercardioid and Hypercardioid Microphones. In *116th AES Convention, Preprint 6059*, Berlin, Germany. (cited on page 104)

Wright, M. and A. Freed (1997). Open Sound Control: A New Protocol for Communicating with Sound Synthesizers. In *Proc. of the International Computer Music Conference*, Thessaloniki, Greece, pp. 101–104. (cited on pages 66, 78, and 93)

Yadegari, S., F. R. Moore, H. Castle, A. Burr, and T. Apel (2002). Real-time implementation of a general model for spatial processing of sounds. In *Proc. of the International Computer Music Conference*, Goteborg, Sweden, pp. 244–247. (cited on pages 66 and 80)

Zatorre, R., M. Bouffard, P. Ahad, and P. Belin (2002). Where is 'where' in the human auditory cortex? *Nature Neuroscience 5*(9), 905–909. (cited on page 9)

Zicarelli, D. (1998). An extensible real-time signal processing environment for Max. In *Proc. of the International Computer Music Conference*, Ann Arbor, Michigan, US, pp. 463–466. (cited on pages 15, 77, and 96)

Zielinski, S. K., F. Rumsey, R. Kassier, and S. Bech (2005). Comparison of basic audio quality and timbral and spatial fidelity changes caused by limitation of bandwidth and by down-mix algorithms in 5.1 surround audio systems. *J. Audio Eng. Soc. 53*(3), 174 –192. (cited on page 172)

Zmöling, J., A. Sontacchi, and W. Ritsch (2003). The IEM-Cube, A Periphonic Re-/Production System. In *Proc. of the 24th AES International Conference: Multichannel Audio, The New Reality*, Banff, Canada. (cited on page 7)

Zouhar, V., R. Lorenzy, T. Musil, J. Zmölnig, and R. Höldrich (2005). Hearing Varèse's Poème électronique inside a virtual Philips Pavilion. In *Proc. of the 11th International Conference on Auditory Display*, Limerick, Ireland, pp. 247–252. (cited on page 6)

Zurek, P. M. (1979). Measurements of binaural echo suppression. *J. Acoust. Soc. Am. 66*, 1750–1757. (cited on page 124)

Zvonar, R. (1999). A history of spatial music. `http://cec.concordia.ca/econtact/Multichannel/spatial_music.html`, accessed Sep. 2010. (cited on page 2)