

Sound - Space - Body
Reflections on Artistic Practice

Trond Lossius
Bergen National Academy of the Arts
Submitted in partial fulfillment for
Research Fellowship in the Arts

February 8, 2007

Contents

1	Aknowledgements	3
2	Introduction	5
2.1	Background	5
2.2	Methods	6
2.2.1	Contextualization	6
2.2.2	Artistic Practice	6
2.2.3	Art as code	8
2.3	Results	9
2.4	Discussion	9
3	Sound and Space	11
3.1	Introduction	11
3.1.1	Coordinate System Conventions	12
3.1.2	Technical Equipment	12
3.2	Acoustics and Psychoacoustics	13
3.2.1	Psychoacoustics and Spatial Sound Perception	13
3.2.2	Acoustics	20
3.3	Spatialisation Techniques	22
3.3.1	Stereo	22
3.3.2	ITU 5:1 Surround Sound	26
3.3.3	Vector-Based Amplitude Panning	28
3.3.4	Ambisonics	33
3.3.5	Wave field Synthesis	40
3.3.6	Distance-Based Amplitude Panning	43
3.3.7	The Ircam Spat Library and Binaural Sound	49
3.3.8	Loudspeakers as Direct Sources of Sound	52
3.3.9	Distant Voices	53
3.4	A Space for Sound	54
3.5	Data Space	61
4	Issues of the Performative	64
4.1	Introduction	64
4.2	Performative Issues in Installation Contexts	65
4.2.1	Music as Process, Music as Weather	65
4.2.2	The Memory of a Gold Fish	71
4.2.3	Installations with Sound of Their Own Making	79
4.2.4	A Shared Space	98

4.3	Performative Issues in Live Art	102
4.3.1	Works for Stage	102
4.3.2	Working with Real-Time Technology in Live Projects	104
5	Concluding Remarks	109
6	Artistic outcome	111
6.1	Installations	111
6.2	Stage works	112
6.3	Music commission	112
6.4	Various projects and events	113
6.5	Prices	113

Chapter 1

Acknowledgements

I would like to thank my supervisors Jerney Welsh and Morten Eide Pedersen for the many valuable discussions over the last years and the general trust and interest in my work. Jeremy Welsh have not only filled the role of supervisor, for the last three years *LMW* has been a framework for a series of collaborations between Jeremy Welsh, Jon Arne Mogstad and myself. Unorthodox as it might be to mix roles this way in academia, it has enabled exchange of thoughts, ideas and hands on experience between the three of us that I do not believe would be possible otherwise. Jon Arne Mogstad, although not formally a supervisor, have been an invaluable additional resource to me, not only because of the projects we have been doing together, but also for the continuous ongoing discussions we have had at our studios over numerous cups of coffee. The same holds true for Karen Kipphoff, also a collaborating partner and colleague at the Department of Fine Art.

I would like to thank colleagues and students at Bergen National Academy of the Arts for making it such an inspiring place to work. In particular I would like to thank the headmaster, Nina Malterud, director Johan Haaberg and dean Paula Crabtree for their support and interest in my work. I would also like to express my thanks to the participants at the gatherings of the fellowship program.

I am fortunate to be surrounded by a network of artists, curators, theoreticians and friends that have contributed to my work one way or another:

Thanks to Mari Aare and the rest of the staff at Hordaland kunstsenter for hosting *Cubic Second*. I am particularly grateful to Hilde Hauan Johnsen for valuable discussions and suggestions concerning the design of the room, Are Hauffen and Kristian Skjold for their help in the construction, Hanc Christian Gilje for lending me technical equipment, and Matthias Arvasstson, Marieke Verbiessen and Elin Solvang for additional assistance.

Q-art 2006 was organized and curated by Kari Skippervold, Petter Kleven, Johan Butenschøn Skre and Anne Vesaas. *Staged Spaces* was a collaboration with Karen Kipphoff. *Quiet and Relaxed, but Alert* was a collaboration with Jana Winderen. *White-out* was a collaboration with Kurt Ralske. My contribution to *Generator.x* was curated by Marius Watz, and produced by Anne Britt Strømnes and Jana Windren. Data used for the project were provided by Frode Flatøy at Bjerknes Centre for Climate Research. *Trickster* was a collaboration with Andrea Sünder-Plassman. *Ex te lux oritur* was a collaboration with Agnethe Christensen. *Elektropoesia* was a collaboration with Kurt Ralske for Electrohype, curated by Anna Kindvall and Lars Midtbø. *Lydspor* was a collaboration with Asbjørn Flø, with additional assistance from Henrik Sundt,

Ståle Stenslie, Astrid Midtbøe, Odd Magnus Nilsen, Andreas Mork and Cato Langnes. Prøverommet Beta was curated by Ellen Røed and Torunn Skjelland. *Kunsten å gå på vannet* was a collaboration with Jana Winderen and Jørgen Træen. *Hacker* and *Tracker* were collaborations with Frode Thorsen and Gitte Bastiansen, produced by Helge Øye for Rikskonsertene and Oppland fylkeskommune. *Cracker* was a collaboration with Frode Thorsen, Gitte Bastiansen and Gisle Frøysland.

Verdensteatret has been of particular importance to my development as an artist, and the methods for cross-disciplinary work and thoughts on integration of art and technology of Verdensteatret have formed the basis for working methods in other projects: Lisbeth J. Bodd, Asle Nilsen, Hkon Lindbäck, Piotr Pajchel, Ali Djabbar, Per Flink Basse, Corinne Campos, Øyvind B. Lyse, Bergmund W. Skaslien, Erik Balke, Lars Øyno, Morten Pettersen, Petter Steen, Kenneth Korstad Langaås, Reinert Mithassel, Kjeld Kemp, Christian Blom and Jannicke Lie.

This text discuss some works done prior to the period of the research fellowship: *Ekkofisk*, 2000-2001 was a collaboration with Reinert Mithassel, with additional input from Ole Henrik Moe and Karle Lange. *Norway Remixed* was a collaboration with Asbjørn Flø and Risto Holoapinen, curated by Bjarne Kvinnsland and Tilman Hartenstein. *Erotogod* was a collaboration with Ståle Stenslie, Asbjørn Flø and Knut Mork Skagen. *Living Room* was produced by Janne Stang Dahl of Produksjonsnettverk for elektronisk kunst in collaboration with Trondheim Elektrokisk kunstsenter, with the following participants: Andrea Sünder-Plassmann, Jeremy Welsh, Reinert Mithassel, Trine E. Eidsmo, Tinna Ludviksdottir, Thorolf Thuestad, Moritz Tredup, Nils Claesson, Espen Gangvik, Robin Stöckert, Kristin Bergaust, Kenneth Langås and Per Jonas Lindstöm and myself.

The *Lydgalleriet* collective has been a valuable resource: Jørgen Larsson, Maia Urstad, Roar Sletteland, Bjørnar Habbestad and Thorolf Thuestad, Steinar Sekkingstad and Erlend Hammer. Special thanks to BEK - Bergen Center for Electronic Arts for all assistance provided. Atelier Nord, NoTAM and PNEK have all supported projects in one way or another.

The initial ports of the Max library *tl.objects* to Mac OSX and Windows were done with the assistance of Jan Schacher and François-Eudes Chanfrault respectively. *Jamoma* has been a joint development with Tim Place, Alexander Refsum Jensenius, Pascal Baltazar and Dave Watson with additional input and suggestions from Alexander Fogar, Mathew Aidekman, John Hudak, Thomas Grill and Jeremy Bernstein. André Sier and Mathieu Charmange have provided valuable feedback on the distance based amplitude panning method. Several mailing lists have been continuous valuable sources of information and help, in particular the maxmsp, sursound and ftm mailing lists. I would also like to express my thanks to the elves at Cycling'74.

If anyone has been accidentally left out, please accept my apologies and let me know, so that I can do the necessary corrections.

I am deeply grateful to my wife, Laila Linde Lossius, for her endless support, understanding, encouragement and patience. I would also like to express my gratitude to our three sons, Thomas, Martin and Alexander, as well as the rest of the family.

Chapter 2

Introduction

2.1 Background

Coming from a musical background, having studied composition at the Grieg Academy, I became more and more involved with sound installations in the late 1990s, influenced by among others the work of Brian Eno, John Cage and Erik Satie. In a somewhat similar manner to Risset (2003) I wanted to sculpt and organize directly the sound material, focusing on the sonic qualities.

In 2000 I started working at BEK, Bergen Center for Electronic Arts¹, and quickly became involved in a number of cross-disciplinary projects, collaborating with other artists most often coming from other backgrounds than my own. In the next couple of years I increasingly realized that the kind of art projects I wanted to work on did not only depend on a musical understanding, but also to a high degree had to be informed by other art practices, most notably contemporary fine arts. I felt a growing need to widen my own horizon and deepen the understanding of the different fields I was getting involved in. For this reason I applied for, and was accepted, a research fellowship in the arts in 2003, at Bergen National Academy of the Arts, Department of Fine Art.

During my research fellowship I have aimed, through work on installations and other interdisciplinary projects, mainly in collaboration with other artists, to explore installation and other interdisciplinary forms of expressions as points of tension and intersection between contemporary music and contemporary fine art.

My goal has been to get closer in my own artistic practice to achieving a synthesis of these fields. I therefore have aimed to develop a greater degree of interdisciplinary competence, increasing my knowledge and understanding of several different artistic fields, and integrate this competence into my ongoing artistic work. An important aspect of this contextualization has been a deepening of my knowledge of contemporary art, with a particular emphasis on sound art, installations and issues of site.

Albeit sound art, encompassing a range of interdisciplinary and multimedia working methods, has a long tradition (Kahn 1999 and LaBelle 2006), it currently seems to get a rapidly growing significance with more and more use of and attention to sound in the visual arts. In the last years my understanding of my own work has changed. At the start of the fellowship I defined my work as primarily existing in a field of intersection and tension between music and arts (Lossius 2005). I no longer consider my work as being merely positioned somewhere between different established art practices. In the

¹<http://www.bek.no>

last two years it has also manifested itself as a platform of its own.

2.2 Methods

2.2.1 Contextualization

Borgdorff (2006) discuss various types of arts research in terms of research *on*, *for* or *in* the arts. Research *on* the arts implies a theoretical distance as practiced by the academic humanities disciplines, with the object of research remaining “untouched under the inquiring gaze of the researcher”. This only partly holds true for the kind of contextualization I have to do as a practicing artist. Theoretical reading and understanding is not only aimed at acquiring knowledge for its own sake. Just as important is the question of how this can become a resource pool informing and influencing my own artistic practice. For this reason my investigation of theory at times takes unorthodox turns and twists, colored by my own artistic and aesthetic needs, interests and preferences.

Likewise the contextualization can not only be theoretical, but also to a large degree have to be practical and intuitive. It is not only a question of acquiring fact-based knowledge and understanding. It is equally important to develop a heightened aesthetic sensitivity towards the various forms and expressions I get in touch with, for example a trained eye for visual artistic expressions, as a prerequisite for meaningful cross-disciplinary communication in collaborative projects.

2.2.2 Artistic Practice

My main modus operandi has been what Borgdorff would name research *in* the arts, the artistic practice itself becoming an essential component of both the research process and the research results. The actual making of art has been the most fundamental method for investigating the various problems I have been dealing with. The application and addressing of new insights, ideas and problems in projects become the benchmark for how artistically fertile they turn out to be.

The research has been a matter of working at two parallel levels. At the same time as I work on specific art projects, they also become experimental playgrounds for investigations of various issues at a meta level. Issues of particular importance have been crisscrossing as a number of red threads through the maze of projects I have been involved in, often with radically different approaches from one project to the next in order to building a broader base of understanding and experience.

The development of each project has been process-oriented. Rather than a strongly predefined concept, the starting point has tended to be a set of issues to deal with and questions to investigate, negotiated between the involved artists. The making of the works have typically progressed through a series of phases involving preliminary research, technical investigations and development, a gradual refining of concept, material and content, finalizing of the work, and public presentations.

For a number of reasons most of the projects I have been involved in have been collaborations with other artists. Many of the projects tend to have a scale that makes them difficult to realize alone. Due to their cross-disciplinary qualities they also require skills and knowledge on more levels than one person can be expected to master. It is also very much a preferred way of working. I find collaborations artistically stimulating, generally finding myself working and developing much faster. I am continually

confronted with new problems and ways of thinking, pushing me past current limitations and insights. I also find the very situation of collaborating, sitting next to and working with someone else, creatively stimulating, letting ideas and thoughts bounce back and forth. The performance group Verdensteatret has clearly articulated methods of collaboration in order to achieve true cross-disciplinary results, where all layers of the final product bear equal importance. Pakes (2004) provides an accurate description of the social complexity of this way of working:

The performing arts necessarily involve collective production and collective action, a number of agents working together to produce performance events. So these events take place within and are the result of an intersubjective context in which it is crucial to have a creative sensitivity to others participating in the process, to the materials at hand and to the evolving situation.

I have found that this way of working can be equally successfully applied to other forms of collaborative cross-disciplinary projects.

In most collaborations the involved artists have had different backgrounds and knowledge. I believe it is important to break down the barriers between the contributions and responsibilities of the involved artists. For the end result to be truly cross-disciplinary all media and layers of the art work have to be equally important and aware of each other, functioning as a multimedia orchestration or counterpoint with a giving and taking between the various layers. I find this difficult to achieve unless everyone involved are permitted and forced to take full responsibility of all parts of the work. An alternative model would be a clear hierarchy with one person having the final word on all matters, but for democratic reasons I find that model less appealing. Schafer (1977), quoting Mumford (1934), draws parallels in the structuring of roles between the symphonic orchestra emerging in the 19th century and the factory of the industrial revolution:

...with the increase in the numbers of instruments, the division of labor within the orchestra corresponded to that of the factory: the division of the process itself became noticeable in the newer symphonies. The leader was the superintendent and production manager, in charge of the manufacture and assembly of the product, namely the piece of music, while the composer corresponded to the inventor, engineer and designer, who had to calculate on paper, with the aid of such minor instruments as the piano, the nature of the ultimate product - working out its last detail before a single step was made in the factory.

Although there is a strong tradition within fine arts and classical and contemporary music for the creative artist to be working individually, I have ended up questioning this convention in my own work. To me cross-disciplinary collaborations do not imply a repression of my own artistic integrity. Instead it represents a different way of realizing myself artistically within a social context. Looking to other parts of society, I find that important progress nowadays seldom is the achievement of a single individual, but rather the result of large-scale long-term distributed collaborations. Drawing on my musical background I believe this should be recognized as a different but equally valid way of working, more resembling the collaborative model of rock and pop bands.

A good dialogue is a prerequisite for good collaborations. This becomes a question of establishing a continuous dialogue, including finding a common cross-disciplinary

language as required. It is also a question of discussing how one is working along the way, not only the actual work. Finally it is important to build an atmosphere of mutual trust and honesty, where it is possible to provide ideas and suggestions, encouragement and critique as required. It is a potential pitfall that the dialogue gets too polite, problems of the projects not being properly dealt with. To avoid this it is important to build personal relationships that can stand confrontations and disagreements, and find constructive ways of solving them.

2.2.3 Art as code

Technical development in my work amounts to more than what Borgdorff (2006) labels research *for* the arts. It is not only a question of developing tools. In my projects sound is generated in real time by one or more computers running custom made software programs developed in the multimedia graphical programming environment Max/MSP/Jitter environment (Puckette and Zicarelli 1990, Zicarelli 1998 and Zicarelli 2002). Often other elements, such as video, sensors or light, will be controlled by Max as well. As such, the medium of my installations is not only sound or multimedia understood as a combination of e.g. sound and video. Programming code becomes a meta-medium, and creating the program is creating the art work.

Some solutions and functions might be reused and further developed from one project to the next. I do not consider this only as a development of tools, it also becomes a matter of developing a compositional language and new compositional strategies. The Austrian composer Karlheinz Essl seems to work in a similar way, and the creation of the *Lexicon-Sonate* also resulted in the Real-time composition library for Max (RTC-lib) (Essl 1995). The Generator.x touring exhibition, conference, blog and concert tour is subtitled “art from code”. This is further elaborated by Marius Watz in the curatorial statement:

A new generation of artists and designers are turning to code as a means of new expression and a way to better control their medium. They have realized that software is not the transparent interface it has conventionally been thought to be. Instead, software is a material that both limits and permits personal expression. True literacy means being able to both read and write. If to use pre-existing software is to ‘read’ digital media, then programming is the equivalent to writing.²

At the time being it seems to me that the emphasis on the importance of craftsmanship is much stronger in contemporary music than contemporary fine arts, at least in Norway. Lately I have come to think that the term “new media art” might have to be understood not only as the merging of media and information technology (Manovich 2001) in an art context. Technological development offers at times radically new means for artistic expression. Within society at large there is a general tendency towards digitizing information wherever possible in order to access new possibilities for further processing of the information, increasingly influencing our understanding and perception of the world at fundamental levels. If art is to be able to deal with issues stirred up by this development, I believe thorough technical and artistic understanding and mastering of new technology and its application in the arts to be a prerequisite. Somehow “new media art” also will have to implicitly reinstate the importance of the artistic medium in order to be able to deal with issues relating to the information age.

²<http://www.generatorx.no/generatorx-introduction>

2.3 Results

The primary product of the research has been a series of art projects, carried out over the time span of the fellowship. Most of the projects have been carried out in collaboration with other artists, and are multi-medial. Sound invariably forms an integral and important part of the projects. The projects fall in different categories. Installations make up the greater bulk of the projects. Sound is not merely functioning as an aural background or soundtrack for the visuals in these installations, instead the projects should be understood as intermedial with all parts being equally important and depending on each other. Work on cross-disciplinary projects have also been extended to live art projects. The final art project carried out within the framework of the fellowship, the installation *Cubic Second* was an installation exhibited at Hordaland kunstsenter November 3 - 12 2006.

Often artistic development have implied the need for new technological development. As a consequence of this I have developed libraries adding functionalities to Max/MSP/Jitter on a continuously basis. Development has been split between two libraries. `tl.objects`³ is a set of functions mainly aimed at audio synthesis, math operations, etc. During the period of the fellowship `tl.objects` has been gradually extended, and it has also been ported to Mac OSX and Windows XP with the assistance of Jan Schacher and François-Eudes Chanfaut respectively.

Jamoma⁴ (Place and Lossius 2006) is a framework for developing high-level modules in the Max/MSP/Jitter environment, encouraging a structured approach, interchangeability and flexibility. Jamoma started as an off-spring from the Jade software developed by Tim Place of of Electrotap Inc.⁵, and was released according to a GNU Lesser General Public License⁶ in the spring of 2005. Since then further development has been a joint effort between Tim Place, Trond Lossius, Alexander Refsum Jensenius, Pascal Baltazar and Dave Watson with additional input and suggestions from Alexander Fogar, Mathew Aidekman, John Hudak, Thomas Grill and Jeremy Bernstein. Jamoma has also been used for research in musical gestures at Department of Musicology, University of Oslo and Schulich School of Music, McGill University (Jensenius 2006) and (Marshall, Peters, Jensenius, Boissinot, Wanderley, and Braasch 2006).

2.4 Discussion

The research has been a matter of working at two parallel levels. At the same time as I worked on specific art projects, they also became experimental playgrounds for investigations of various issues at a meta level. Issues of particular importance have been crisscrossing as a number of red threads through the maze of projects I have been involved in, often with radically different approaches from one project to the next in order to build a broader base of understanding and experience.

Gradually I have seen that these threads are themselves interwoven, forming broader bands. Two topics have been of particular importance; relationships between sound and space, and between sound and body. These will be the main concerns of this reflection on my artistic practice over the last three years. In addition some thoughts will be provided on artistic collaboration and art and technology.

³<http://www.bek.no/~lossius/download>

⁴<http://www.jamoma.org>

⁵<http://www.electrotap.com>

⁶<http://www.gnu.org/licenses/lgpl.html>

Although this text represents the primary reflection on my work over the last years, I have also been maintaining a blog⁷ throughout the duration of the fellowship, in order to log and document progress and process on an ongoing basis. The blog has proven particularly useful during the early phase of projects, while doing background research, reading, raising new questions, testing new approaches, etc. In periods of intense development, typically as the opening/premiere draw close, I tend to get short on time. In this phase I also find myself working at the edge or past current understanding and knowledge, also depending on intuition and a sensitivity to the needs and inherent aesthetic qualities of the project. Apart from the limited time available for deadline reasons, I have to work at a certain speed in order to stay in the flow of things, being vulnerable to becoming to self-aware and self-critical. For a while the inner critical voice has to be thrown a bit off guard, and only when the project has been completed can this phase of the process be reexamined in order to better understand and decode what happened in the final stages. The blog has been a valuable tool for logging and tracking progress for my own sake, encouraging a continuous reflection and questioning of my own work, a systematic examination of issues over extended time spans, and forcing me to articulate myself more thoroughly.

Several projects will be discussed in this text. I do not attempt at any exhaustive discussion of each project, but emphasis aspects that has been important to the long-term exploration of various topics. Almost all projects have been collaborations with other artists. They bring their own agendas to the shared work, and often their emphasis, focus and understanding of the process, work and gained experience will differ from mine. There is not *one* story about each project, but multiple. The following discussion reflects what I have come away with, or rather what I think that I have come away with at the time of this writing. The artists I have been collaborating with will all be able to supplement the discussions, sometimes elaborating, sometimes expanding in other directions and sometimes telling a completely different story. That is what makes collaborations so stimulating and challenging.

⁷<http://www.bek.no/Members/lossius/lostblog>

Chapter 3

Sound and Space

3.1 Introduction

Almost all of the art projects I have realized during the last 6 years have depended on sound being played back using multiple loudspeakers. The initial motivation was an interest in sound objects, working and listening *inside* of sound rather than emphasizing relationships between sound events. Early on this lead to a slowing down and sound starting to suggest places rather than narratives. The use of multiple loudspeakers have been important in my efforts to create sonic environments, with the spectator/listener immersed or embedded in sound. It has been my hope that by inviting the audience to physically move inside the sound, it would also encourage encourage them to listen inside the sound.

Surround sound, derived from technology for cinema, has become widespread in the consumer marked, in part due to the advent of DVD, with Dolby Surround radio and television broadcasts and computer gaming following suite. Consumer surround sound as defined in ITU-R BS.775-1 standard is first and foremost a specification of how a specific number of loudspeakers are to be positioned relative to a spectator/listener sitting in a fixed position, heading in a specific direction. For a number of reasons that will be further elaborated when discussing the ITU 5.1 standard in section 3.3.2 I have generally considered it inadequate for my artistic needs in installation contexts. Instead I have been doing a more or less systematic practical, theoretical and aesthetic survey of other approaches to sound spatialisation in my projects over the last 6 years. The aim has not necessarily been to search for a specific optimal solution that could serve any project. Rather it has been an investigation of what solutions do exist, how and to what degree they work and in what situations they might be artistically useable. Methods for spatialisation are invariably based on a number of assumptions that might make them suitable for one project but inappropriate for the next.

While some techniques are well documented, others have been more difficult to penetrate. Most of the more advanced approaches discussed are documented mainly in scientific papers. With the exception of Wiggins (2004) I have seen few attempts at systematic reviews of advanced algorithms and techniques for spatialisation, and none that have been aimed primarily at artists and musicians. For this reason I have chosen to include a fair bit of technical details in some of the following sections.

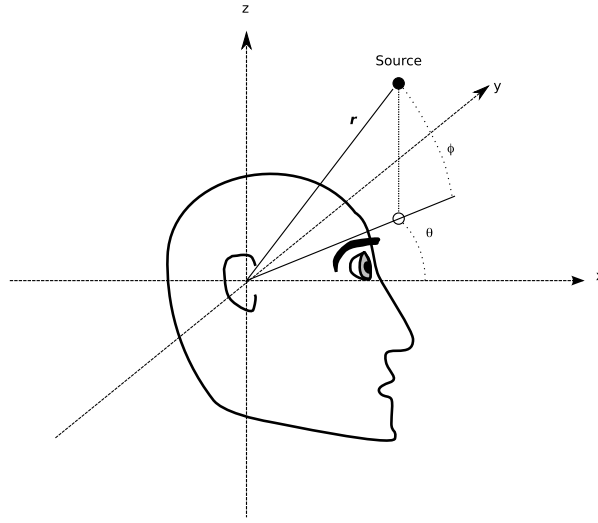


Figure 3.1: Cartesian and polar coordinate systems.

3.1.1 Coordinate System Conventions

Most spatialisation techniques are optimized for a listener positioned at the *sweet spot*, a specific position, often also with a specific frontal direction. In these cases the cartesian coordinate system will be defined with x - and y -axes in the *horizontal* plane at the height of the ears so that x -axis is frontal direction and the y -axis pass through the two ears. The coordinate system is right-handed, with z -axis pointing upwards as illustrated in figure 3.1. The x - and z -axes defines the *median plane*, symmetrically dividing the space related to a listener into left and right parts.

Spherical coordinates are defined with radius r being the distance from *origo* to the point. Azimuth θ is the angle between x -axis and the projection of the point onto the horizontal plane. Elevation ϕ is the angle between the point and its projection onto the horizontal plane. Conversion between cartesian and spherical coordinates is as follows:

$$x = r * \cos \theta * \cos \phi \quad (3.1)$$

$$y = r * \sin \theta * \cos \phi \quad (3.2)$$

$$z = r * \sin \phi \quad (3.3)$$

3.1.2 Technical Equipment

As all projects carried out so far have been temporary I have been able to reuse much of the technical equipment from one project to the next. Starting with a grant from the Municipality of Bergen in 2001, I have gradually been collecting Bowers & Wilkins 602 S2 and S3 speakers and Denon AVR-280n multichannel receivers, currently being able to work with 18 channels. This has been a reasonable compromise between loudspeaker size and look, sound quality and pricing. Although sound colorization without doubt differs between the speakers, it is still far more homogeneous than using a combination of speakers of different brands and types. I have also grown

accustomed to these speakers, and believe that I have a feeling for how the sound will play back using the system, more or less unconsciously adjusting for any spectral colorization along the way.

I have the impression that other artists tend to prefer active to passive speakers. Personally I have found passive speakers to have an advantage in gallery contexts in terms of requiring far less cabling, substituting power and XLR cables for a simple loudspeaker cable only. Often using several hundred meters of cabling in one installation this quickly becomes both a visual and financial issue.

In spite of being a regular customer at a local HiFi shop, I have never been attempted to have technical equipment sponsored. I have seen to many works disturbed by the combined effect of a sponsorship sign being the first thing to meet your eyes as you approach the work and oversized and flashy technical equipment distracting from the real content of the work. Sponsorship is of course possible, but the artist and curator have to be cautious about how it might influence the perception of the work.

3.2 Acoustics and Psychoacoustics

3.2.1 Psychoacoustics and Spatial Sound Perception

It is beyond the scope of this text to go into details of perception and cognition of spatial sound. The text will be limited to a brief presentation of terms and phenomenas of relevance to the further discussion. A thorough review of the topic has been done by Blauert (1996) in his book *Spatial Hearing* and by Moore (2003) in *An Introduction to the Psychology of Hearing*. A more compact and pragmatic presentation of relevance to musicians can be found in chapter 2 of the book on *Spatial Audio* by Rumsey (2001).

In the following discussion a *sound event* will describe a physical event while *auditory event* describes a perceived event. The discussion will start with the special case of a single sound source, and then discuss spatial hearing with two and more sound sources.

Spatial Hearing with one Sound Source

Perceptual localization of a point source to a large degree depends on the difference between the sound signals reaching the two ears. If the source is positioned to the left of the listener as in figure 3.2, the path of the sound to the two ears differs. There is a difference in the distance to the two ears, and the right ear is shadowed by the head, causing time and amplitude differences.

The maximum time delay between the ears is on the order of 0.65 ms and called the *binaural delay* or the *Interaural Time Difference* (I.T.D.). This is one of the cues used by the ear/brain system to calculate the position of sound sources. The ear/brain system analyses interaural time differences according to subclasses; time shifts in the envelope of the sound as well as phase offsets of the carrier between the two ears.

Time shifts in the envelope is particular important to determine spatial localization at the onsets and offsets of sounds, as well as if the sound contains transient information. For more static sound, e.g. sinusoidal signals, the timing difference is experienced as phase offsets between the signals reaching the two ears. The phase difference depend on the localization and frequency of the source. For low frequencies the ear/brain system is sensitive to phase differences, but the sensitivity starts deteriorating for higher frequencies, as half a wavelength of the sound becomes equal to or smaller than the

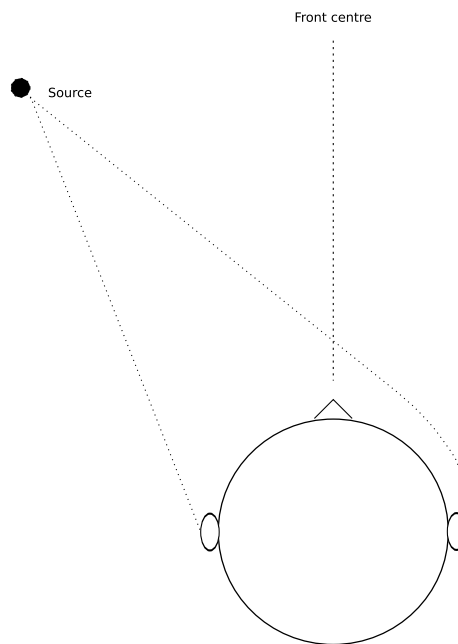


Figure 3.2: The two paths that sound from a source to the left of the listener must follow to the ears.

distance between the ears. Phase information starts becoming ambiguous for source localization above 800 Hz, and at 1600 Hz phase differences no longer provide any localization cues.

In reverberant environments the direct signal will interfere with reflections to create more complex phase patterns, further confusing the value of phase differences as cues for spatial localization. An effective demonstration of this is the complex interference patterns experienced when visiting the *Dream House* installation by La Monte Young and Marian Zazeela as discussed by Hammer (2005).

The difference in distance that sound has to travel to the two ears, is too small to cause any noticeable difference in level by itself. However, if the source is positioned off centre the shadowing effect of the head will cause level differences known as *Interaural Level Difference* (I.L.D.). This is depending on frequency due to *diffraction*, the bending of the sound wave around the obstacle of the head. As a simple rule of thumb, any sound that has a wavelength larger than the diameter of the head will tend to be diffracted around and any sound with a wavelength shorter than the diameter of the head will tend to be attenuated, causing a lowpass filtering effect.

Figure 3.3 shows the frequency ranges in which carrier phase offset, envelope time shifts, and level differences can have an effect. According to Blauert (1996) “carrier time shifts have an effect only below 1.6 kHz. Envelope time shifts have less of an effect as the frequency is decreased; the lowest frequency at which they have an effect depends on the shape of the envelope. Interaural sound pressure level differences are effective throughout the audible frequency range. The relative importance of interaural time and level differences depends on the type of sound signal. (...) Interaural level differences have their greatest importance when the signal includes components above

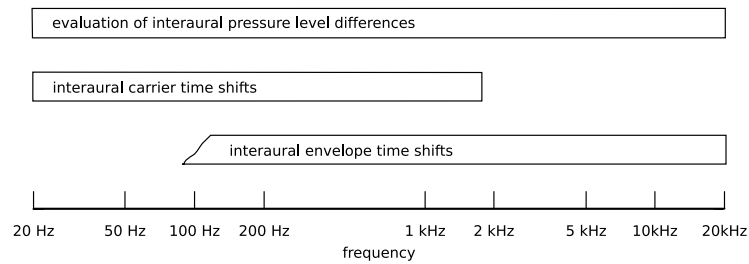


Figure 3.3: Frequency ranges in which different attributes of interaural signal differences are evaluated (after Blauert (1996)).

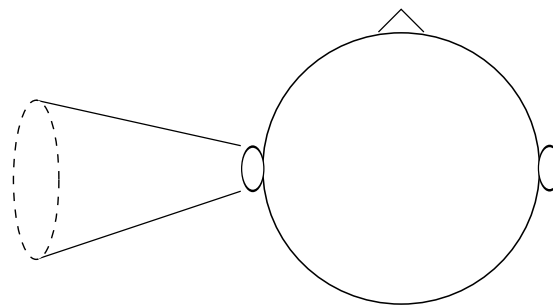


Figure 3.4: The cone of confusion.

1.6 kHz and the level is low.”

The diffraction of the sound due to the obstacle of the torso, head and in particular the pinna is depending on both the direction and frequency of the sound. Acoustically the pinna functions as a linear filter whose transfer function, the *head-related transfer function* (HRTF), depends on the direction and distance of the source. “By distorting incident sound signals linearly, and differently depending on their direction and distance, the pinna codes spatial attributes of the sound field into temporal and spectral attributes” (Blauert 1996). The HRTF differs from one individual to another.

If virtual localization of a sound source is simulated over headphones using only amplitude and time differences, it is insufficient to ensure a precise localization of the source due to a phenomenon known as the *cone of confusion*. Any sound that is coming from a cone of directions as indicated in figure 3.4 will have the same level, phase and time differences associated with it making the actual position of the source potentially ambiguous. When listening to an external source HRTF will provide additional spatial clues. Movement of the head also contributes towards locating the position of the source.

Auditory localization is generally less precise than visual localization, and depends on the position of the source as well as the type of signal. “For directional hearing in the horizontal plane the minimum localization blur occurs in the forward direction. With increasing displacement from the forward direction towards the right or left, the localization blur increases. At right angles to the direction in which the subject is facing, the localization blur attains between three and ten times its value for the forward direction. (...) Behind the subject the localization blur decreases once more, to ap-

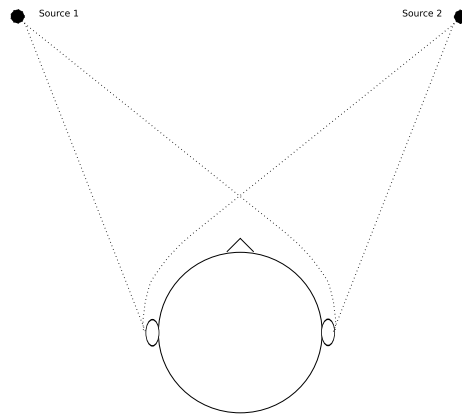


Figure 3.5: With two sources there are four paths to the ears. This example shows two loudspeakers in the standard stereophonic setup.

proximately twice its value for the forward direction.” (Blauert 1996). In the forward direction the localization blur typically varies from 1° for impulses to to approx. 4° for sinusoids.

Spatial Hearing with Two Sound Sources

The following discussion will be limited to deal with two sources radiating *coherent* signals only. Two signals are coherent if they differ only in one or more of the following three ways: (I) differing amplitude, (II) one signal being delayed relative to the other and (III) one signal inverse respectively to the other (Blauert 1996).

The idealized study of sound fields originating from two sources has an immediate application in stereophonic reproduction, further discussed in section 3.3.1.

As illustrated in figure 3.5, with two sound sources four paths to the ears have to be considered. Depending on the sources, three different classes of auditory events can result:

1. One auditory event appears in a position depending on both sources.
2. One auditory event appears, but its position is determined by one of the sources only.
3. Two auditory events appear, the position of one event depending more or less exclusively on one of the sources only, the position of the other event depending on the position of the other source only.

The first case occurs if the levels of the two signals do not differ much, and the time of arrival differs by less than 1 ms. The two signals will then be perceived as coming from a single *phantom sound source*. This phenomena is coined *summing location*, and is at the core of stereophonic reproduction. The differences in levels or shift in time will determine the position of the auditory event. As the delay increase towards 0.63 ms the auditory event moves towards the direction of the lead speaker, being located near or at the speaker for time differences between 0.63 ms and 1 ms.

The relationship between auditory localization and differences in level is more complex, depending on frequency and envelope of the sound signal. Due to the phenomena of diffraction it can be shown that “a difference in the levels of the loudspeaker signals at low frequencies leads only to a time difference between the ear input signals, and conversely, a time difference between the loudspeaker signals leads only to a level difference between the ear input signals”, (Blauert 1996). For sinusoids with frequencies above 800 Hz the relationship between levels of loudspeaker signals and perceived localization gradually break down, but in a similar way as for single sources, differences in levels of signal envelopes seem to influence perceived localization.

If the signals differ by more than 1 ms and less than approximately 5 ms, they will still be perceived as a single auditory event, but the position is felt to be constant, more or less determined by the position of the lead source only. This is known as the *law of the first wavefront* or the *precedence effect*, (Blauert 1996) and (Litovsky, Colburn, Yost, and Guzman 1999). The second signal arriving tend to be perceptually masked by the first signal, but will alter the tone color of the auditory event, as well as increase the spatial extent of the event.

As the delay time is further increased, the fusion breaks down, and the auditory event separates into two events in different locations, the position of each to a high degree being determined by the position of the two sources respectively. The second auditory event, the *lag*, is perceived as an *echo* of the lead signal. The time delay threshold for the echo to appear depends on the signal as well as the spatial separation of lead and lag sources. Quantitative estimates of thresholds varies tremendously (2-50 ms), being short for brief stimuli such as clicks and longer for complex ongoing sound such as noise or speech. The steeper the slopes of the signal, the shorter the delay time at which the auditory event disassociates.

More than Two Sound Sources

The results for spatial hearing with two sources can be generalized to spatial hearing in sound fields generated by more than two sources (Blauert 1996). This is relevant in a discussion of sound reproduction using more than two speakers, as well as when discussing sound fields of enclosed spaces, where sound reflections of surfaces can be thought of as being generated by mirror-image sound sources. Blauert (1996) states that “generally all spatial hearing effects observed in connection with two sources can also occur in connection with multiple sources, though these effects are modified in some cases”. Accordingly, summing location, the law of the first wavefront and echo might all occur.

In the two-sources case the coherent signals had to be delayed by less than 1 ms for summing location to happen. When more sources are present, the auditory system takes into consideration coherent components of the ear input signals that arrive within up to a couple of milliseconds after the first component. The criteria seems to be that each component arrives within less than 1 ms compared to the previous component.

The law of the first wavefront can be generalized in a similar way. A particular component is less likely to be audible if additional components are present between the primary sound and this component. This also applies to the echo threshold. Blauert (1996) refers to tests using impulse signals. With the primary sound and a test reflection being the same level, the echo threshold was 10 ms. Inserting an additional reflection at the same level the echo threshold could be prolonged to 20-30 ms. Adding more reflections the echo threshold could become as great as 200 ms.

Distance Hearing

In a free field the sound from a point source will propagate outwards as a sphere. As the surface of the sphere is proportional to the square of the radius, the intensity I of the sound will be inverse proportional to the square of the distance r in a free field, *the inverse square law for intensity*:

$$I \sim \frac{1}{r^2} \quad (3.4)$$

As intensity is proportional to the square of pressure v , this leads to the *inverse distance law* for sound pressure:

$$v \sim \frac{1}{r} \quad (3.5)$$

Logarithmically this is expressed as a *6 dB decrease per doubling of the distance* (Everest 2000). As will be discussed further in section 3.2.2 free field propagation of sound is a highly idealized case with no reflections. This seldom occur in real in- or outdoor spaces.

The mechanisms for perceiving distance differs depending on the distance, and can be divided into three or four regions:

At intermediate distances from the sound source, approximately 3-15 meters, the sound wave can be considered planar, and sound pressure level at the ears is the only factor varying with distance.

At greater distances it is also necessary to take into consideration the absorption of sound in air. This is frequency-dependent, with high frequencies attenuated more than lower. Air absorption also varies with moisture content and wind speed, and as well as fluctuating in time due to turbulence.

Close to the sound source the sound wave can no longer be considered planar. Due to the curvature of the wave HRTF is no longer independent of the distance of the source. (Wiggins 2004) points to the fact that a source close to the listener will have a greater binaural distance associated with it. As a consequence movements of the head will not only help to determine the direction of the source, but also the distance.

Headphone listening eliminates the filtering effect of the pinnae, thus removing important distance cues. This can lead to *inside-the-head-locatedness*. The phenomena can be compensated in binaural audio by artificially introducing HRTF filtering, further discussed in section 3.3.7, or by introducing artificial reverberation cues to create the illusion of a sound source outside of the head (McKeeg and McGrath 1997).

Changes in distance of a source leads to perceived changes in the spectral balance of the signal, and the tone color is perceived to become darker (Blauert 1996). Diagrams of equal loudness contours illustrates how perceived loudness of a source depends both on frequency and intensity, the auditory system being most sensitive at about 2000 Hz, and not very sensitive to frequencies below 100 Hz or above 10.000 Hz. The difference is more pronounced for quieter sources (Mathews 1999). When a broadband source moves closer, loudness is increased, and the low-frequency components gain more weight.

As the sound source moves sufficiently close to the listener, it can no longer be considered a point source, a source of sound emitted from an infinitesimal small point in space. The source becomes a bodily volume increasing in size as it draws closer, and the direction that sound is perceived to originate from is no longer a single beam but a wider field of beams.

Studies of how sound pressure relates to perception of distance, concludes that if the sound level pressure of the ear input signals is the only attribute of the sound event available to the auditory system to form the distance of the auditory event, then a trend exists for this distance to increase less rapidly than the distance of the sound source. On this basis von Békésy suggested that auditory space is of limited total extent, in other words that there is an outer limit to the distance of auditory events, an *auditory horizon* (Blauert 1996).

Familiarity with the source signal improves the accuracy of distance perception, but this also depends on the expected context of the sound: Sound of whisper, ordinary speech or somebody calling out loud played back at the same sound level will lead to an impression of the sound of whisper being located nearest and the sound of calling furthest away (Blauert 1996).

If the source is moving, the distance from source to listener will change over time, unless the source is moving in circles around the listener. Changes in distance will lead to change in signal delay due to the speed of sound propagation in air, causing the well-known *Doppler shift*, a change in perceived frequency and phase as the source moves towards and then away from the listener.

Interaction Between Hearing and Other Senses

Sound localization do not depend on the ears only, but also on other senses. Motion of the head would not be able to contribute to sound localization unless the auditive signals are interpreted in conjunction with other sensuous information such as sight and information from the vestibular organ that provide the subject with an idea of his or her position and direction in space.

Auditive information concerning source localization is also matched with visual information. At least to some degree the sound localization is a trained ability, and can be retrained in the case that hearing capacity change. This must be assumed to be a process of calibrating auditive and visual cues concerning localization.

Blauert (1996) provides several illustrations of the tendency for visual and auditory events to merge spatially. Klemm in 1909 suggested a rule of spacial complication, claiming that there is a general tendency for perceptions of different senses to merge spatially when stimuli are presented simultaneously to the different senses.

The composer Natasha Barrett points to yet another factor that might be of importance. The associative and contextual quality of the sound source is likely to influence how the localization of the source is perceived:

The only 'lab tests' looking at how the perceived localization of the auditory event correlates to the intended position of the event that I am aware of involve simple tones and filtered noise bursts. The problem is that real-world sound involves at least two complications:

1. Complex spectra and temporal morphologies don't correlate very well with lab tests, and even though one can imagine some nice math, looking at for example centroid weightings to find spectral approximations, there appears to be no clear answer.

2. Sound identity recognition by the listener is extremely important. If the listener recognises the sound, then this understanding interacts with physiological perception, creating a clearer concept of the perceived location. So lab tests normally use abstract sound to avoid the problem of sound identity. Sound identity recognition can be different for all listeners and is a biologically conditioned listening state. This goes for distance and angle.¹

3.2.2 Acoustics

In section 3.2.1 it was pointed out how intensity of sound propagating in a free field is proportional to the inverse of the square of the distance. This equals a 6 dB decrease in intensity with each doubling of distance.

Even in outdoor conditions a free field is rare, with the reflections from ground generally being inescapable except for a field in the vicinity of the source. The amount of reflection depends on the surface, but will generally reduce the decrease in loudness with distance. Instead of a 6 dB fall in intensity with each doubling, the fall will be more in the range of 4 to 5 dB (Everest 2000).

In enclosed spaces the surfaces of walls, floor, ceiling and objects in the room functions as acoustic mirrors, reflecting the sound beams in similar ways to how a mirror reflects light beams. The reflections propagate out into the room again eventually reaching new surfaces to bounce off from. With each reflection some energy is lost, and gradually the energy of the sound signals dissipates, transformed to heat.

In an enclosed space the inverse law still applies close to a point source. If sound pressure is measured at increasing distances from the source, it will be found to first fall off with 6 dB per doubling, but gradually this evens out as the effects of reflected sound begins to be effective. At the *critical distance* the direct and reflected sound are equal. The higher the level of the primary sound in comparison to that of the diffuse field, the more precisely located is the auditory event. If the level of the diffuse field is overwhelmingly higher than that of the primary sound, there is no primary auditory event but only a diffusely located auditory event (Blauert 1996).

If random noise is suddenly emitted from a loudspeaker in a room, sound quickly builds up to a certain level where the amount of energy radiated from the source equals the amount of energy lost through absorption. If the source of the random noise is then muted, earlier emitted sound will continue to bounce around the room, but eventually the energy will dissipate. The *reverberation time* of the room is defined as the time required for the sound in the room to decay by 60 dB (Everest 2000). Roughly speaking this is the time required for the sound to die away. Reverberation time depends on the physical shape of the room including all objects present, and the absorbing properties of the surfaces.

Sabine's equation can be used to calculate the reverberation time:

$$RT_{60} = \frac{0.16V}{Sa} \quad (3.6)$$

where RT_{60} is reverberation time in seconds, V is the volume of the room in cubic meters, S is the total surface of the room in square meters and a is the average absorption coefficient of room surfaces. Sa is the total absorption, measured in sabins.

¹Private communication.

From Sabine's equation it can be seen that a rectangular room with fixed floor area will be more reverberant the higher the ceiling is. Likewise a rectangular room of fixed height and floor area will be maximally reverberant if the shape of floor is a perfect square. The more rectangular the room is the less reverberant it will be.

Different surface materials have different absorbing properties, and the absorbing properties of most materials are frequency-dependent. For this reason room reverberation is generally calculated and measured at several frequencies (125 Hz, 250 Hz, 500 Hz, 1 kHz, 2 kHz and 4 kHz) in order to describe the properties of the room for low, mid and high frequency sound. A sabine coefficient of 0 indicated that all energy is reflected, while a sabine coefficient of 1 indicates that all energy is absorbed. Tables of absorption coefficients can be found for common building materials, e.g. in the appendix of Everest (2000). Bricks and concrete surfaces absorb little or no energy at any frequencies, although coarse concrete blocks absorb energy at low frequencies. Glass and wooden floors absorb low frequency energy to a certain amount. Acoustic tiles, drapes and carpets absorb mainly at high frequencies, but heavy carpets and drapes also absorb at mid frequencies.

If a sound source produce a brief click, and the resulting sound is recorded at a distance, a complex pattern will be seen. First comes the direct click, and then several discrete reflections, caused by the first reflections of the surfaces near the source, the *early reflections*. The number of reflections gets denser, forming a *cluster* eventually becoming the dense and complex *reverb tail*. The delay between the direct signal and the first reflection is the *arrival time gap*.

Such recordings, *impulse responses* or echograms, are finger marks of the transitory behavior of a source at a specific location in relation to a specific listening position. The impulse response can be convolved with a dry signal to produce the sound that would result from playing back the dry signal at the source position and listen to it at the destination position. This technique form the basis for several commercial hardware and software reverberation units and plug-ins, offering a variety of impulse responses captured at real spaces such as cathedrals, churches, orchestral halls, rock stadium venues and domestic spaces. One such yet to be released plug-in, Speakerphone by AudioEase, seems to combine impulse responses of various speakers such as cell phones and radios with field recordings sampling various environments, so that one can get the artist to rap through "one of Speakerphones actual GSM connections surrounded by the acoustics and sounds of an idling car at a gas station"². From teaser demonstrations on their web pages this seems to produce very convincing displacements or aural scenography, at least when used to process spoken language.

The density D of reflections per time unit increase as the square of the elapsed time after the sound source radiates the pulse:

$$D = \frac{4\pi c^2}{V} t^2 \quad (3.7)$$

where c is the speed of sound and V is the volume of the room (Blauert 1996).

²<http://www.audioease.com/Pages/Speakerphone/speakerphone.html>

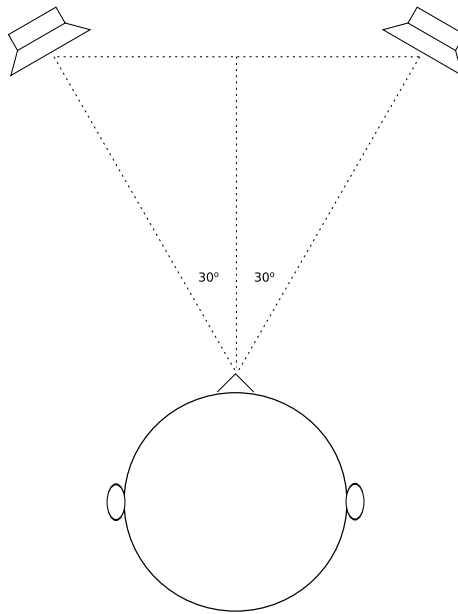


Figure 3.6: Optimum arrangement of two loudspeakers and listener for stereo listening (after Rumsey 2001).

3.3 Spatialisation Techniques

3.3.1 Stereo

Principles of Loudspeaker Stereo

Standard two-loudspeaker stereo is aimed at providing a frontal listening experience, with the optimum configuration generally accepted to be an equilateral triangle with the listener positioned just to the rear of the point of the triangle as illustrated in figure 3.6 (Rumsey 2001).

If the outputs of the two speakers differ only in amplitude and not in phase (time) then it can be shown (at least for low frequencies up to around (d)700 Hz) that the vector summation of the signals from the two speakers at each ear results in two signals which, for a given frequency, differ in phase angle proportional to the relative amplitudes of the two signals (Rumsey 2001 p. 54). Recalling the discussion in section 3.2.1 this will create the illusion of an auditory event positioned somewhere between the two speakers.

For higher frequencies the ability of a system based on level differences only to create a stereo image is harder to analyze. Still Rumsey (2001) p. 57 points out that “if a system based only on level differences did not cope accurately with transients then one would expect transients to be poorly localised in subjective tests, and yet this is not the case, with transients being very clearly located in Blumlein-style recordings”.

Equal intensity panning, also known as the *-3 dB panpot law*, is a well-established technique for panning mono sources in a stereo mix. Only amplitude is used to position the source in the stereo image, and the intensity I remains independent of the stereo positioning. As intensity is related to perceived loudness, this ensures that the loudness

or distance of the source appears constant to the listener. The stereo position is defined in terms of the angle θ . For a mono source with amplitude v_s the left and right channel amplitudes v_L and v_R are given as:

$$v_L = v_s * \cos \theta, \quad v_R = v_s * \sin \theta, \quad 0 \leq \theta \leq \frac{\pi}{2} \quad (3.8)$$

$\theta = 0$ is panned hard left and $\theta = \frac{\pi}{2}$ hard right. If the mono source have unit amplitude, the intensity I , defined as the sum of the squares of the amplitudes for each speaker, will remain constant and equal to 1:

$$I = v_L^2 + v_R^2 = 1 \quad (3.9)$$

Limitations of Two-Channel Loudspeaker Stereo

For a listener positioned at the ideal position the spatial experience of a stereophonic reproduction remains frontal. Even if stereo is originally developed for reproduction of concert music and movies, this limits the spatial impression of the reproduction. The reverberation and ambiance of a real concert hall will reach the listener from all sides, but is coming from the front only in a stereo reproduction. The impression created by this reduction can be described as “it is there” instead of “you are there”. As I have been aiming at creating a sonic environment surrounding the audience, I have generally found stereo reproduction to be insufficient, not only for the actual works, but also a recurring problem when documenting my own installation works as video with stereo sound.

Documentation of installations is generally challenging. Installations emphasize first-hand experience, have no privileged view, and depend on the audience being situated inside the work. As discussed by Bishop (2005) they are not easily reduced or framed:

Works of installation art are directed at and demand the presence of the viewer. This point is further reinforced by the problem of how to illustrate installations photographically. Visualisation of a work as a three-dimensional space is difficult via a two-dimensional image, and the need to be physically *inside* an installation renders photographic documentation even less satisfactory than when it is used to reproduce paintings and sculpture.

This is equally much the case for the sound of the installation. My use of multiple speakers and spatialisation resist the idea of the *sweet spot*, the privileged listening position. The audience is situated within the soundscape, and the sound is explored by moving around the space. Foreground becomes background and vice versa, and layers of the sound might open up to reveal spatial and sonic details as they are approached. Contrary to vision auditory perception is not restricted to what is in front of us. Thus one of the important functions of the sound in the installations is to function as a perceptual glue, creating an awareness of all of the space, also those parts that your attention is not currently directed at. You can't leave the sound behind. All of this is lost in a sound recording reduced to stereo.

Phantom images are subject to some tonal colorization as they are panned across the sound stage, owing to the way that the signals from two loudspeakers sum at the ears of the listener. A phantom central image will have a certain amount of midrange colorization compared to that of an actual speaker in that position (Rumsey 2001).

In an installation and gallery setting the audience will generally be allowed and expected to move around the space. The stereo reproduction will quickly start to fall apart as the audience leaves the *sweet spot*. Using a Cartesian coordinate system with *origo* at the sweet spot and *x*-axis pointing towards the speakers, the relative delay in milliseconds between the direct signals from the two speakers to any point (x, y) in the space can be expressed as:

$$t(x, y) = \frac{1000}{c} * (\sqrt{d^2 + x^2 + y^2 - \sqrt{3} * d * x - d * y} - \sqrt{d^2 + x^2 + y^2 - \sqrt{3} * d * x + d * y}) \quad (3.10)$$

In this equation c is the speed of sound, and d is the distance between the two speakers.

As time difference increase past 1 ms summing localization gives way to the precedence effect, causing the auditory event to collapse into the nearest loudspeaker. As time difference continue to increase past 5 to 10 ms, the two signals from the speakers will gradually be perceived as separate auditory events, one being an echo of the other. In addition, assuming that loudspeakers are positioned in reasonably distance to walls, roof and ceiling, one or more reflections of the signal from the nearest loudspeaker might reach the listener before the direct signal from the far speaker.

Figure 3.7 plots the difference in time for the direct signal from the two speakers depending on the position of the listener. Both plots assume square spaces with the speakers positioned at two of the corners. One of the spaces is relatively small with a distance between speakers of 3.5 m. The other is larger, with a distance between speakers of 10 m. As can be seen, the audience do not need to walk far off the middle axis of the room before localization is determined by the precedence effect, hence collapsing towards the nearest loudspeaker. In the larger room there are also regions at the left and right sides of the rooms where echo between the speakers is likely to occur.

Also in the case that the listener only moves along the *x*-axis marking the middle of the room, the stereo image will change as he moves closer to or further away from the speakers. Moving closer the loudspeaker angle will increase, introducing a hole in the middle. Further away the stereo image will grow narrower as the loudspeaker angle decrease.

In gallery contexts the interaction between the stereo signal reproduced and the acoustics of the space should be taken into consideration. Gallery spaces tend to be reverberant, as they are generally large and fairly empty with mainly hard and non-absorbing surfaces, with the possible exception for acoustic tiling. The critical distance defined in section 3.2.2 is the distance from the source where the reverberant signal is as strong as the direct. Unless the direct signal is clearly louder than the reverberant signal there will be no primary auditory event but only a diffusely located auditory event. This might be the reason why sound reproduced over stereo in gallery spaces often sound muddy, degraded by the interference of the reverb of the space. If more than two speakers are used, distributed over the space, there is a higher likelihood that the audience will be in the vicinity of least one of the speakers regardless of where they move in the space. This will be particularly beneficial if a spatialisation technique is utilized that always employ all of the loudspeakers, such as ambisonics, to be discussed in section 3.3.4. Using more loudspeaker might thus be a means to improve the perceived directness and preciseness of the sound.

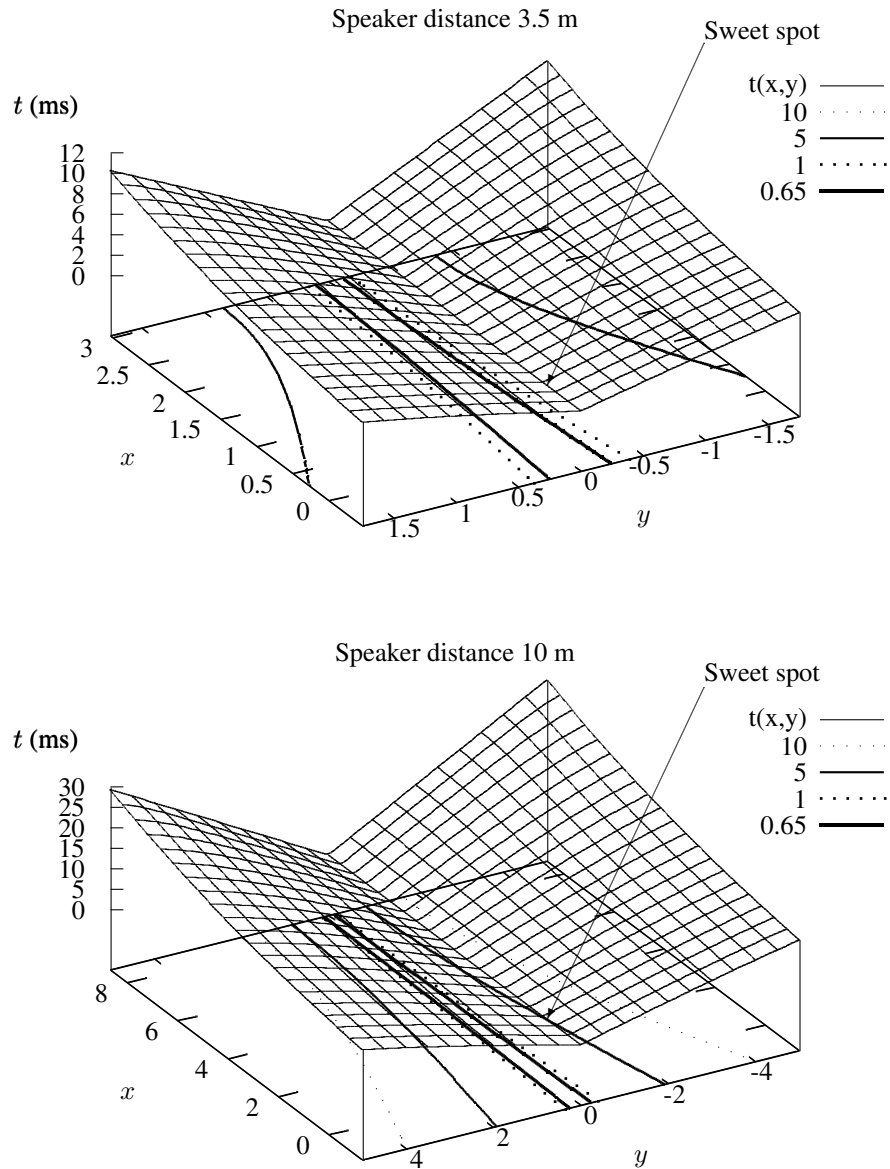


Figure 3.7: Difference in time that sound takes to propagate from two speakers positioned at the far left and right corners in a small and larger room. Contours indicate the transitional zones between summing localization and precedence effect (roughly 0.65 to 1.0 ms) and between precedence effect and echo (5 ms and upwards depending on the sound source).

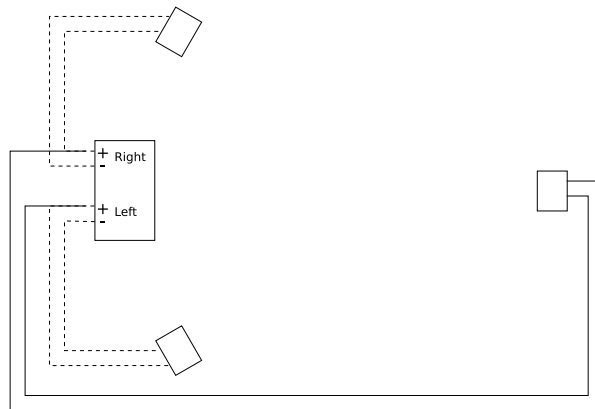


Figure 3.8: Ambient three-way speaker system as suggested by Brian Eno in 1982.

Eno's Ambient Speaker System and Dolby Stereo

In an attempt to circumvent the frontal flatness of traditional stereo Brian Eno suggested an ambient speaker system on the cover notes for the record *Ambient 4: On Land* (1982), connecting a third loudspeaker as illustrated in figure 3.8

This system resembles a simplified layman version of Dolby Surround decoder, introduced in the 1970s and 1980s, a matrix system decoding a stereo signal to four channels (Left, Center, Right, Surround), with a 20 ms delay for the surround channels to reduce matrix side-effects that could otherwise result in front signals appearing to come from behind. The Dolby system works well for movie sound but is not particularly suited to music reproduction, because the stereo image tends to shift around, and the front image tends to be sucked towards the centre unless the original recording was mixed extremely wide (Rumsey 2001).

3.3.2 ITU 5:1 Surround Sound

5 channel surround sound is a standardized setup for channel assignment and loudspeaker layout for cinema, television and consumer applications as specified in ITU-R BS.755 (ITU 1993). The layout is shown in figure 3.9. This configuration should be considered an extension of ordinary stereo rather than a system for true reproduction of sound from all sides. The front left and right channels are positioned at the same angles as for stereo, to ensure downward compatibility. A center channel is added to improve localization of sources positioned at the center. In film reproduction this help anchoring dialogues more clearly in the middle of the screen. The addition of a center channel also widens the listening area. Preferably similar loudspeakers should be used for all channels. The low frequency effects channel (LFE) is considered optional in consumer audio, and is intended to be used for conveying special low frequency content, not for conveying the low frequency component of main channel signals. For this reason recordings should be made to sound satisfactory even if the LFE channel is absent.

The angle between front left and surround left channel is too large to create stable phantom images, and the same goes for the angle between the two surround channels, causing a “hole at the back”. Essentially “the three front channels are intended to be

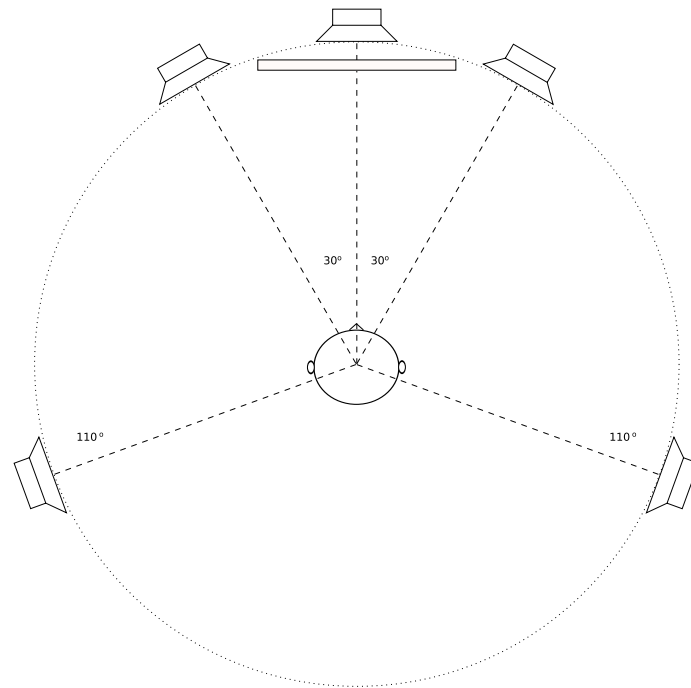


Figure 3.9: 3-2 format according to the ITU-R BS.775 standard.

used for a conventional three-channel stereo sound image, while the rear/side channels are only intended to be used for generating supporting ambience, effects or ‘room impressions’ ” (Rumsey 2001). ITU 5.1. enables a more enveloping experience than stereo, and is more tolerant towards audience positioned off-centre, but it still assumes a frontal presentation not always suitable for installations. It has a rigid definition for loudspeaker positions that in installation might be difficult to achieve, not fitting in with the layout of the space and installation. It has to be added that 5:1 is fairly tolerant to deviations from this specification, stated thus by Martin Leese³:

The point about 5.1 is that while there are preferred speaker positions, nothing particularly bad happens if you ignore this advice. That is to say, there is a lot of “slop” in the system. Ambisonics⁴ is more particular, but then it is attempting to achieve a lot more. I could have rephrased the above to say that with 5.1 nothing particularly *good* happens when you do place speakers in their preferred positions.

Personally I often find myself wanting to use more than five speakers, as well as higher degree of flexibility in speaker layout. Still ITU 5:1 has some obvious advantages due to the standardization, offering a relatively cheap and technically very simple and stable setup. In addition it is supported by most digital audio workstation software (DAW), and an abundance of plug-ins exist for surround sound effect processing. If

³E-mail to the sursound mailing list on January 29 2007: <https://mail.music.vt.edu/mailman/listinfo/sur-sound>.

⁴Ambisonics is discussed in section 3.3.4

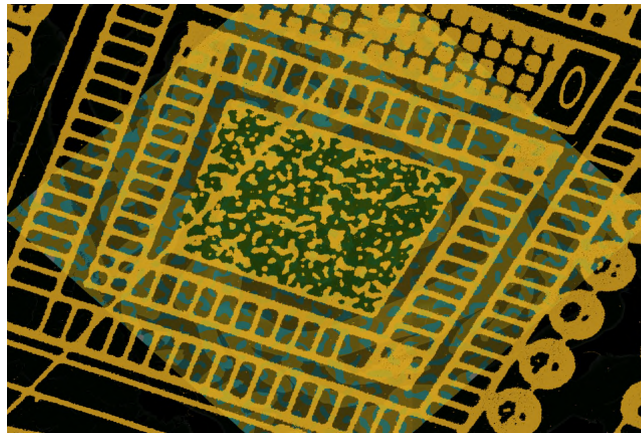


Figure 3.10: Screen capture from *77 Million Paintings* by Brian Eno, 2006. The visual component of *77 Million Paintings* is a dynamically altered onscreen image composited by dynamic layering of still images. The work relates to screensavers as ambient music relates to muzak. Used by permission.

5:1 sound is combined with video on a DVD, sound will be exposed to a lossy Dolby Digital (AC-3) data compression, degrading the quality of the audio signals. In particular for hi-fidelity sound and experimental music and sound art this might be critical, and some artists consider the use of digital compression a destruction of their work (McKenzie 2006).

I have not used ITU 5:1 myself in any artistic work so far, but I can see two uses for it that could potentially be very attractive: In documentation of installations sound reproduced using 5:1 instead of stereo might overcome or reduce the loss of spatial feel, bringing it closer to the real experience. For documentation on DVD it would be able to provide two alternative audio tracks, uncompressed stereo and compressed 5:1. If sound documentation of installations is encoded as second order ambisonic audio, according to section 3.3.4, Wiggins (2004) propose convincing ways of decoding to ITU 5:1 as well as stereo played back over headphones.

As wide screen television and surround sound gets common in private houses (and Norwegians are generally early adaptors of new consumer technology) it is also possible to imagine audiovisual art for private spaces. The recent *77 million paintings* DVD/Art Software Package by Brian Eno is one such example (figure 3.10). There are also several examples of electroacoustic composers adopting 5:1 for releases of their work. The combined DVD-a (ambisonics 5.1 decoding) and CD *Kraftfelt* by Natasha Barrett was the first release of this kind by a Norwegian composer. She was awarded the Nordic Council Music Prize 2006 for one of the works on this release, "...fettters...".

3.3.3 Vector-Based Amplitude Panning

Vector based amplitude panning (VBAP) is an extension of the panning principle to an array of loudspeakers (Pulkki 1997). In the two-dimensional horizontal case loudspeakers will be encircling the listener, the position of each of the speakers defined using angular coordinates. The position of the virtual source is likewise defined according to angle. The two loudspeakers closest to and surrounding the virtual source

are activated, and amplitude panning used to adjust balance between the two speakers. If the position of the source is rotating, one of the two speakers will fade out as the source draws near to the other. Passing it, the next speaker in the row will gradually fade in.

In the three-dimensional case, speakers are assumed to be positioned on a sphere surrounding the listener. Positions of speakers and source are given in spheric coordinates, the three speakers closest to and surrounding the source are found, and amplitude panning is applied adjusting gain levels for each of these speakers.

The basic conception described above has been extended by adding the possibility to define directional spread or diffuseness, by describing position of the virtual source by means of a cluster of vectors instead of one vector only, and then apply panning to each of these vectors. If all vectors converge to the same direction, minimum locational spread possible for the system will be achieved. As the vectors spread, the perceived position of the virtual source will be more diffuse (Pulkki 1999). At maximum spread the signal is omnidirectional.

VBAP is implemented for Max/MSP (Pulkki 2000) as well as other audio processing programming environments, and is computationally efficient. Compared to ITU 5:1 it has several obvious benefits. The number and positions of loudspeakers can be custom defined, and the system is not frontal, but enables sound originating from all surrounding angle as well as three-dimensional speaker setups.

VBAP used for Ekkofisk

I first used VBAP for *Ekkofisk* (2000-2001), an installation in collaboration with Reinert Mithassel. A large aquarium containing two gold fishes were continuously monitored by two surveillance cameras. The images were analyzed and information combined to track the 3-dimensional positions of the two fishes. This information was used to generate sound in real time, a synthesized singing voice for each fish when they were in the vicinity of each other, and bell-like sounds if they moved further away. The decision to translate video tracking data to parameters controlling singing voices caused a slower interactive response, as discussed in section 4.2.2. Most parameters were updated only at discrete intervals as the decision about what note to sing next were taken. The spatialisation of the sound was the only parameter that could be updated continuously as the fishes moved, and as a result this parameter, once we got it right, became the most immediate and obvious link between the visual sight of the fishes swimming and the sound generated.

Four speakers were used for the installation. For three of the public displays they were resting horizontally at the top of the aquarium, one at each side, as illustrated in figure 3.11. For the presentation at World Wide Video Festival in Amsterdam, 2001, four tall and slim Bang & Olufsen speakers were used, positioned at the corners below the aquarium. This represented a 45° rotation of the speaker setup relative to the other presentations. One of the advantages of VBAP was that this kind of changes were very easy and fast to implement, simply a question of changing the description of the speaker coordinates. No further changes to the MaxMSP patches were required.

The speaker setup for *Ekkofisk* differed from standard VBAP setups. Instead of speakers surrounding a listener positioned at the sweet spot, the speakers now formed an inner circle with the audience positioned outside. VBAP still served this project well. Sound would rotate around the aquarium as the fishes moved. Due to the directional properties of the speakers, the fish would be felt to be further away when at the opposite side of the aquarium as sound would not only be panned to the speaker

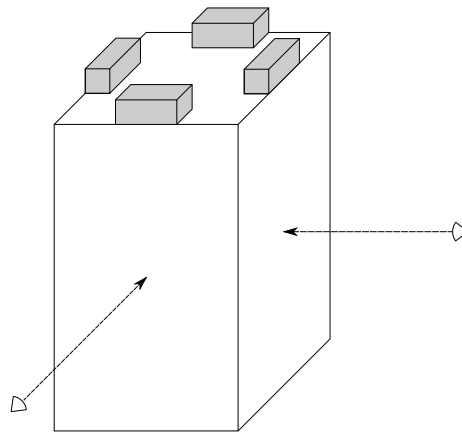


Figure 3.11: Illustration of camera and loudspeaker setup for the Ekkofisk installation. Two cameras were monitoring the aquarium from perpendicular angles. Four speakers at the top of the aquarium were used for sound playback.

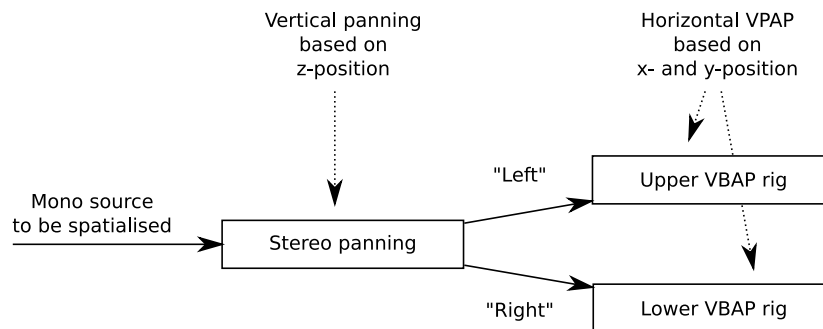


Figure 3.12: Possible approach to three-dimensional panning in Ekkofisk combining stereo panning and VBAP.

furthest away, but that speaker would also be projecting the sound away from the spectator/listener. One possible pitfall of the system would be if the fish moved from one side to the other through the center of the aquarium, causing a sudden 180° jump in panning position. In order to avoid that, the VBAP spread parameter was linked to the distance of the fish from the z -axis running vertically through the centre of the aquarium. The closer to the vertical axis, the higher the spread would be, with a 100% spread at the z -axis.

The most important limitation of the system was its inability to provide vertical positional clues. In hindsight I believe that we would have benefited from adding another four speakers positioned below the aquarium in a similar way to the ring at the top. If so two possible approaches to the vertical panning could be imagined. One would be to define a three-dimensional VBAP system with eight speakers. Without having tested it, I fear that this solution might not work satisfactory. 3-dimensional VBAP use only 3 speakers at a time, at least for a 0% spread. As the fish moves this would often lead to rapid series of skips back and forth between sound playback through 2 speakers at the



Figure 3.13: The listening room of *Norge - et lydrige* with the centrally placed controller.

top and one below, or 2 below and one at the top. With the audience often being very close to the aquarium and hence the speakers, we could risk that the skips would cause audible discontinuities. An alternative solution would be to use regular stereo panning to balance levels between the lower and upper sets of speakers, and then use a separate horizontal VBAP setup for each of the rigs, as illustrated in figure 3.12. This method would ensure that four speakers always would be active.

Decoupling Description of Position for Sources versus Speakers

Following Wiggins (2004) systems for spatialisation can be divided into two categories. The systems discussed in the previous sections, stereo and 5-channel surround, define a speaker layout and/or carrier medium but with no reference to how signals are captured and/recorded for the system.

The alternative is systems that capture or encode information concerning the position(s) of the source material, and then decode it for a specified speaker layout. In these systems information regarding the positioning of the sources is decoupled from information concerning the positioning of loudspeakers in the reproduction system. This introduces a flexibility lacking in the first category as the description of source positioning is independent of and can be conceived independently from the speaker layout that eventually will be used for reproduction. Examples of such systems include VBAP, ambisonics, and wave field synthesis.

There are several benefits to this category of systems in the creative process, as the process of developing *Norge - et Lydrige*, *Norway Remixed* illustrates:

Norge - et Lydrige, *Norway Remixed* was a public sound art project in three parts, realized in 2002 and curated by Bjarne Kvinnsland of NoTAM and Tilman Hartenstein

of the Norwegian Broadcasting Corporation (NRK) (Rudi 2003). One part, *Norway Remixed*, was a listening space custom built inside a vacant retail space at Oslo S, the railway station in Oslo, with sound developed in collaboration between Asbjørn Flø, myself and Risto Holopainen. Walking through a passageway providing shielding from the background noise of the station, the inner part of the room was acoustically controlled, with acoustic design by Lars Strand of NRK. The circular space had 24 loudspeakers hidden in roof, walls and floor. A vertical rod in the middle of the room contained 24 on/off push buttons provided an interface for interaction. 16 of the buttons would trigger sound recorded or streamed by the regional offices of the Norwegian Broadcasting Corporation one for each button. The sounds were processed to balance at the borderline between being recognizable representation and abstract sound objects, utilizing granulation with parameters dynamically changed so that each button produce dynamic physical gestures, with different identities for each buttons. The remaining 8 buttons functioned as modifiers, applying tremolo, ring modulation, delays and reverb to the 16 sound sources. VBAP was used for spatialisation of the interactive layers of sound. The movement of sound in the space was a critical important part of the shaping of the gestures, and different patterns were scripted, e.g. sound spiraling from the roof downward to the floor, sound running in circles around the spectators or sound starting at a point at one side of the spectator and dissolving into grains coming from all sides.

For the installation me and Asbjørn Flø developed a generic system for scripting and controlling of gestures, and used it to compose complex progressive gestures.⁵ For development and prototyping in Bergen two to four speakers were used. At the studio of Asbjørn we had access to 16 speakers. Only a few days before the opening we were able to start testing with 24 speakers at Oslo S. The flexibility in describing the loudspeaker setup was crucial to this project. With 24 or 16 loudspeakers we were able to fully hear the three-dimensional spatial movements. With four loudspeakers only the horizontal component could be simulated, and two loudspeakers would reduce spatial gestures to linear movements between left and right channel. Still, the flexibility in definition of speaker setups enabled prototyping and verification of the technical solutions, and provided setups that were sufficient to be able to imagine what the effects would be like on the full-fledged system.

Spatial Impulse Response Rendering and Directional Audio Coding

VBAP as discussed so far is a technique for synthesized spatialisation; mono sources are virtually positioned according to desired direction and diffuseness. VBAP has been integrated in recent research at Laboratory of Acoustics and Audio Signal Processing, Helsinki, expanding the technique to work on captured signals as well as incorporating simulation of room reverberation.

Spatial Impulse Response Rendering (SIRR) can be used to reproduce room acoustics with any multichannel loudspeaker system. In SIRR impulse responses of a room are captured as B-format signals (see section 3.3.4, and analyzed within frequency bands to extract information concerning direction, diffuseness and sound intensity. A multichannel response suitable for reproduction with any chosen surround loudspeaker setup is synthesized using the analysis data. The resulting loudspeaker responses are used in a multichannel convolving reverberator, and the synthesized responses create a natural perception of space corresponding to the measured room. Resulting impulse

⁵This system has later been licensed to Electrotap LCC and Cycling'74 and incorporated as part of Hipno audio processing plug-ins (Place, Wolek, and Allison 2005).

responses are commercially available for the Waves IR convolution reverbs⁶ (Merimaa and Pulkki 2004).

Directional audio coding (DirAC) applies a similar method applied to spatial sound reproduction. The sound signals are divided into frequency bands using filterbanks. The diffuseness and direction of arrival of sound at each frequency band is analyzed depending on time. A mono channel is transmitted with additional directional information, or, in applications targeting for best quality, all recorded channels are transmitted. In the decoding/synthesis process the sound at each frequency channel is first divided into diffuse and non-diffuse streams. The diffuse stream is then produced using a method which produces maximally diffuse perception of sound, and the non-diffuse stream is produced with a technique which produces as point-like perception of sound source as possible (VBAP). DirAC was demonstrated at AES 120th convention in Paris 2006 as applied in teleconferencing (Pulkki and Faller 2006) and high-fidelity reproduction of B-format signals (Pulkki 2006). Compared to higher-order ambisonic recordings demonstrated by Trinnov⁷ and others at the same event, DirAC had a remarkable sense of depth in the encoded hi-fidelity sound field. The teleconferencing application can probably be useful for the intended purpose, but the lower data rates introduced audible artifacts that would make it unsuitable for the purposes discussed in this text.

The encoded DirAC signal contains both a ordinary mono signal and a control rate signal providing information on direction and diffuseness of the encoded signal. If the control rate signal is itself sampled at full audio sample rate, the signal might not be possible to integrate with most music software, else this might limit the applicability of the system.

DirAC is still in development, and not publicly available. It also remains to see if it will be as easily accessible as VBAP. While the algorithm for VBAP has been open-sourced, SIRR was patented and licensed to Waves Audio Ltd. One consequence of this is that SIRR impulse responses are not freely available. Although the technique itself is flexible in terms of speaker layouts, the commercial distribution restricts flexibility, only offering implementations for stereo and ITU surround. As such this is a archetypical example of how access to the source code and the right to alter it and redistribute would provide additional freedom, as advocated by the Free Software Foundation⁸.

3.3.4 Ambisonics

Background

Ambisonics was a system pioneered mainly by Michael Gerzon in the 1970s (Wiggins 2004), and is based on spherical harmonic decomposition of a sound field. First order ambisonics represents the sound field using four signals or audio channels, collectively known as *B-format*. The *W* signal is an omni-directional pressure signal representing the zeroth order component of the sound field. *X*, *Y* and *Z* make up the first order components and record particle velocity in three dimensions. These three signals can be considered equivalent to recorded signals using figure of eight microphones. A mono source will be encoded as B-format by multiplying the signal with a coefficient depending on the direction of the source, using spherical coordinates, as defined in section 3.1.1:

⁶<http://www.acoustics.net>

⁷<http://www.trinnov.com>

⁸<http://www.fsf.org>

$$W = \frac{1}{\sqrt{2}} \quad (3.11)$$

$$X = \cos \theta \cos \phi \quad (3.12)$$

$$Y = \sin \theta \cos \phi \quad (3.13)$$

$$Z = \sin \phi \quad (3.14)$$

The spherical decomposition can be extended to higher orders, providing additional more detailed spatial information on the sound field. The second order encoded signal adds five more channels to B-format, giving a total of 9 channels:

$$R = \frac{3}{2} \sin^2 \phi - \frac{1}{2} \quad (3.15)$$

$$S = \cos \theta \sin(2\phi) \quad (3.16)$$

$$T = \sin \theta \sin(2\phi) \quad (3.17)$$

$$U = \cos(2\theta) \cos^2 \phi \quad (3.18)$$

$$V = \sin(2\theta) \cos^2 \phi \quad (3.19)$$

Third order encoded signals introduce seven additional channels, giving a total of 16 channels:

$$K = \sin \phi \left(\frac{5}{2} \sin^2 \phi - \frac{3}{2} \right) \quad (3.20)$$

$$L = \cos \theta \cos \phi (5 \sin^2 \phi - 1) \quad (3.21)$$

$$M = \sin \theta \cos \phi (5 \sin^2 \phi - 1) \quad (3.22)$$

$$N = \cos(2\theta) \sin \phi \cos^2 \phi \quad (3.23)$$

$$O = \sin(2\theta) \sin \phi \cos^2 \phi \quad (3.24)$$

$$P = \cos(3\theta) \cos^3 \phi \quad (3.25)$$

$$Q = \sin(3\theta) \cos^3 \phi \quad (3.26)$$

Generally a full three-dimensional ambisonic encoded signal of n th order will have $(n+1)^2$ channels.

Ambisonic encoding is linear, and thus implies that several encoded signals can be combined by simple channel by channel signal additions. Ambisonics is a hierarchical format, so in limited reproduction settings redundant channels can be ignored. For example in a horizontal only decoding the Z signal can be ignored in first order ambisonics, while Z , U and V can be ignored for second order ambisonics. In addition the K signal will become independent of azimuth and hence can be substituted for a scaled version of W .

Spatial manipulations can be performed on the encoded signal such as rotation around the z -axis, tumble around the y -axis or tilt around the x -axis by means of linear matrix operations on the signals. Some formulas for rotation of first, second and third order ambisonics can be found in Malham (2003). Wiggins (2004) also provides equations for how to zoom in on a part of a first-order sound field.

Recording

First order ambisonic microphones combine four capsules, mounted in a tetrahedral array. The signal can be processed to B-format (Craven and Gerzon 1977) as well as many conventional mono and stereo microphone patterns (Wiggins 2004). B-format microphones are produced by SoundField⁹. CoreSound¹⁰ is shipping TetraMic early 2007, introducing a welcome alternative at a lower cost. Combined with the increasing availability of portable two and four track recorders, this enables light-weight and affordable portable systems for B-format field recordings.

Research into higher-order ambisonic microphones is ongoing (Laborie, Bruno, and Montoya 2003) and (Moreau, Daniel, and Bertet 2006), and applied in the Trinnov Surround Recording Platform¹¹, a high spatial resolution solution for ITU 5:1 recordings.

In an effort to make ambisonics more accessible, the surround sound community loosely organized via the sursound mailing list¹² has made available a wide range of B-format recordings for free download¹³. A file format for ambisonics encoded sound has been defined¹⁴ as a custom subtype of Microsoft WAVE-EX (Microsoft 2001).

Decoding

The discussion so far of methods for capturing and encoding of signals has been independent of any specification of the layout of the speakers used for reproduction. For the encoded signals to be played back they have to be *decoded*, converted to signals appropriate for the speaker layout that will be used, so that the spatial information of the encoded signal can be restored as far as possible. The separation of encoding and decoding is a major advantage of the system.

If we had wanted to record the spatialized sound of the *Norway remixed* installation discussed in section 3.3.3, we would have to record one track per output channel, a total of 24 tracks. If the same installation were to be mounted again, but with a slightly different speaker layout, we would be unable to use the previous recording for playback as it would not correspond to the new speaker layout. If ambisonics was used instead, we would be able to record the encoded signal, and then adjust the decoding process to the speaker layout. This would also greatly reduce the number of tracks required for recording. For first order ambisonics 4 tracks would be sufficient, second order would require 9 tracks, and 3rd order ambisonics 16 tracks.

The spatial encoding is a prerequisite for ambisonics, but the term *ambisonics* is defined as a set of requirements for the decoder (Gerzon and Barton 1992), restated by Wiggins (2004) as:

A decoder or reproduction system for 360 surround sound is defined to be Ambisonic if, for a central listening position, it is designed such that:

1. velocity and energy vector directions are the same at least up to around 4 kHz, such that the reproduced azimuth $qV = qE$ is substantially unchanged with frequency,

⁹<http://www.soundfield.com>

¹⁰<http://www.core-sound.com>

¹¹<http://www.trinnov.com>

¹²<https://mail.music.vt.edu/mailman/listinfo/sursound>

¹³<http://www.ambisonicbootlegs.net>

¹⁴<http://dream.cs.bath.ac.uk/researchdev/wave-ex/bformat.html>

2. at low frequencies, say below around 400 Hz, the magnitude of the velocity vector is near unity for all reproduced azimuths,
3. at mid/high frequencies, say between around 700 Hz and 4 kHz, the energy vector magnitude, rE , is substantially maximised across as large a part of the 360 sound stage as possible.

The definition is motivated by psychoacoustic theory. As can be recalled from the discussion of figure 3.3 in section 3.2.1 localization of frequencies below 800 Hz predominantly depends on interaural carrier time shifts while localization depends on interaural pressure level differences for frequencies above 1600 Hz. The definition of an ambisonic decoder responds to this by emphasizing the summed velocity vector from all reproduced signals at low frequencies and summed pressure vector for high frequencies.

If \bar{w} , \bar{x} , \bar{y} and \bar{z} are the four channels of a B-format signal, decoding the signal s_i for the i th loudspeaker of the setup is done by adding the encoded signals:

$$s_i = k_{wi} W \bar{w} + k_{xi} X \bar{x} + k_{yi} Y \bar{y} + k_{zi} Z \bar{z} \quad (3.27)$$

where W , X , Y and Z are calculated according to equations 3.11 to 3.14 using the spherical coordinate of the speaker, and k_{wi} , k_{xi} , k_{yi} and k_{zi} are decoding coefficients depending on the speaker layout. Determining the optimal decoder is thus a question of finding optimal values for these coefficients. For a setup with n speakers $4n$ coefficients have to be determined. In three-dimensional systems of second and third order $9n$ and $16n$ coefficients have to be determined. It goes without saying that this in general is a daunting task, most likely not possible to be solved analytically.

In symmetrical horizontal systems the Z term can be ignored, the coefficient k_{wi} will be the same for all channels, and the coefficients k_{xi} and k_{yi} are the same and equal for all channels. In this case the decoding equation can be solved as:

$$s_i = \frac{\sqrt{2}}{N} (\bar{w} + \frac{\sqrt{2}k}{2} (\cos \theta \bar{x} + \sin \theta \bar{y})) \quad (3.28)$$

with $1 \leq k \leq 2$ being the *fudge factor*, a coefficient controlling the quality of the encoding. $k = 2$ results in the idealized matching mathematical form of the problem. This produce the best localization at the sweet spot, but due to the fact that some of the speakers will have inverse phase the sweet spot will be small, and artifacts due to phase cancellations occur outside the sweet spot. $k = 1$ avoids reversed phases, known as *controlled opposite*, resulting in a larger listening area at the expense of some directional information. Matched and controlled opposite coefficients are readily available¹⁵ for a number of symmetric two- and three-dimensional rigs for decoding of B-format and second order signals, known as the *Furse-Malham higher-order format*.

N denotes the number of speakers, ensuring consistent over-all level independently of the number of speakers used. In digital audio it is advisable to ignore N as increasing values for n will cause decreasing levels for each channel, degrading effective sample rate and signal to noise ratio in the digitally sampled signal. Instead the equation should be tuned for maximum level while avoiding digital clipping, and the overall level should instead be adjusted by the amplifying units.

The definition of ambisonic decoding require separate sets of decoding coefficients for low frequencies as compared to higher frequencies. This is done by applying a shelf

¹⁵<http://www.muse.demon.co.uk/ref/speakers.html>

filter to the first order encoded component, boosting levels of first order components relatively to the omni component at low frequencies. The shelf filter typically has a cross-over frequency of 500 Hz, and in digital applications it is generally implemented as a finite-impulse-response filter using convolution. Shelf filters boosting first order components for low frequencies easily introduce reversed phases at low frequencies. For this reason when decoding for large rigs where a large listening area is desired, it is recommended to use controlled opposite decoding without shelf filter¹⁶. Malham (1992) discuss work carried out at the Music Department at York University on large ambisonics rigs for concert reproduction in the early 1990s.

No general approach exist for decoding to asymmetric loudspeaker rigs. Gerzon and Barton (1992) attempted to find first order decoders for a number of asymmetric loudspeaker setups of relevance to cinema and high definition television, the so-called *Vienna decoder*. None of the speaker arrays were as irregular as the ITU 5:1 surround sound ended up being specified. The Vienna decoder have been shown by Wiggins (2004) to be suboptimal. Instead he solves the problem by means of numerical methods, using a Tabu search algorithm, providing superior results to those shown by Gerzon and Barton (1992).

Farina and Ugolotti (1998) propose a different strategy for decoding using a small dedicated program (Farina and Righini 1997) that can be set to emulate a standard decoder. The decoder includes convolution with inverse filters which compensate for the irregularities of the loudspeakers. The system is first calibrated by recording B-format impulse responses of all loudspeakers. These are analyzed by the program to determine the positions and characteristics of the speakers, and from this create an inverse filter designed so that the frequency response and the absolute delay of each loudspeaker are perfectly equalized. This means that the array can be built with speakers of different kinds and makers, placed at varying distances from the center and even at varying angular positions: the self-adjustment procedure takes care of all these mismatches, and compensates as far as possible.

Concluding the discussion it should be mentioned that there is a minimum number of speakers needed to successfully reproduce each Ambisonic order, which is always greater than the number of transmission channels available for the decoder. This problem can be compared with the aliasing problem in digital audio, that is, enough samples must be used in the reproduction array. Higher order polar patterns, when decoded, do not imply that fewer speakers are working at the same time; they are just working in a different way to reconstruct the original sound field (Wiggins 2004).

Technical Implementation of Ambisonics in MaxMSP

Several matrix-based implementations of ambisonics encoding and decoding has become available for Max/MSP in the later years. The implementation of ambisonics in Jamoma (Place and Lossius 2006) is based on the externals developed at Institute for Computer Music and Sound Technology (ICST) in Zurich by Schacher and Kocher (2006), offering first, second and third order encoding and decoding. Positions of sources are given in or converted to spherical coordinates including distance, and the encoder incorporates a gain stage depending on distance. By default this is set to a 3 dB decrease in amplitude when doubling the distance, but the amount of decrease can be customized. If sources moves close to origo, inside the unit circle or sphere, the omni component continue to increase, but all other components are lowered towards

¹⁶D. Malham, personal communication.

zero as the source approach origo. At origo this results in a omni-directional signal, thus ensuing continuity in the signal if a virtual source is made to move from one side of the circle to the other passing through origo.

Upstream of the encoding module in Jamoma modules for Doppler-effect and air filtering can be inserted, the latter based on Ircam Spat (Jot 1996).

Centre de Recherche Informatique et Création Musicale (CICM) in Paris has developed externals for two- and three-dimensional first order encoding and decoding¹⁷. Encoding and decoding is done in the same external, and it only accommodates for one source. Thus from a practical point of view some of the benefits of ambisonics are not taken advantage of. One do not get access to the encoded signal so that it can be recorded, encoded signals can not be mixed with other B-format signals originating elsewhere, if the patch is to be used for several different speaker layouts it has to be changed again and again instead of just substituting one decoding module for another. Finally, from a computational point of view the implementation of the encoding and decoding matrixes are suboptimal. The strength of the CICM implementation is that source and loudspeaker positions can be provided in Cartesian coordinates, and the encoding and decoding takes into account the relative distances of both source and speakers. The CICM externals have generally produced very convincing results, in particular when used with non-standard and non-symmetric speaker layouts, maybe more so than the ICST externals. As both libraries are released under a GNU LGPL open source license, it is possible to combine the best of both worlds by integrating code from both sources. This however remains to be done.

Experience using ambisonics

The general purpose of my investigations into different techniques for spatialisation has been a general search for ways to work with sound in room that would make the listener immersed in and surrounded by sound. It has not necessarily been important that the experience should be the same regardless of position in the room, quite the contrary, but I have wanted to avoid the experience of the immersive sound field from collapsing in parts of the room as far as possible. From one project to the next I have also been playing with speaker layouts, considering the placement of sound and sound sources in the space a sort of audio sculpting. Interestingly work on sonic qualities is also often discussed as audio sculpting, *Audiosculpt* indeed being the name of a software program used for this purpose; the experience of space and the vertical color of sound (Tamm 1995) seem related. The work with sound in space is also very much a response both to the specific site at hand and to the contributions of my collaborating partners.

Most of my experiments with ambisonics have been violating the fundamental assumptions of the system in one way or another. Ambisonics have seemed to be more tolerant in this respect than most other systems, and at times I have been able to use the artifacts and errors introduced due to my abuse of the system in ways that have been artistically interesting.

I first used ambisonics for *Elektropoesia*, a sound and video installation at Malmö konsthall for the Electrotype biennale in 2004-2005 in collaboration with Kurt Ralske, figure 3.14 . The video format was extreme wide screen, approx. 8x1 meters or 2048x240 pixels. In the installation I wanted to indirectly comment on the discussion of noise and the line by Kahn (1999), with a literary take on his claim that the line is

¹⁷<http://cicm.mshparisnord.net/dl/ambipan.htm>



Figure 3.14: Elektropoesia. Installation of sound and video.

“a point where the meeting of *audio* (‘I hear’) and *video* (‘I see’) has been particularly conspicuous”. Just underneath the video 16 loudspeakers were laying sideways on a shelf. The initial idea was to have the line of speakers working as a virtual string, and I imagined the sensation of sound moving sideways on the string in a similar way to the horizontal movements that were a predominant part of the custom-coded slitscan technique used by Kurt Ralske.

Different approaches were attempted with very poor results, up until I started testing ambisonics using the CICM externals. This was probably as far away as one could possibly get from the recommended standard Ambisonic speaker setups where loudspeakers are supposed to be evenly spaced around the listener. I ended up virtually positioning speakers a few meters behind the row of speakers, and had the various layers of sound move sideways. The details concerning how movements were controlled are further discussed in section 4.2.1. I was impressed by the results, in particular as I seemed able to move the sound way beyond the borders of the line segment defined by the speakers, creating a sensation of the sound effortlessly transcending the walls of the space. Although the movements did take place in a sideways direction, a peculiar phenomena manifested itself. The sound made a physical sensation and impact that I have never experienced before or after. Standing in front of the video projection and loudspeaker row, sound was felt to arrive from the front and pass past and through me at tremendous speed, as if I was standing in a wind of sound. Thus the direction that totally dominated my experience of the work was not the line tangential to the projection surface, but the perpendicular line pointing from the work to and through me.

To this day I am not sure what caused this effect. The most likely explanation I

have is that the loudspeakers set up next to each other must have started acting together in a way resembling wave field synthesis (see section 3.3.5). In ambisonics all speakers are active all of the time, but the gain levels will be more pronounced for the speakers closest to the virtual source position. I suspect that the signals from each of the speakers have started to combine according to Huygen's principle, resulting in a wave with a much wider front.

All I know for sure is that the effect was not caused by high sound levels. This was a group exhibition and our installation was positioned in a semi-open room, so to high levels would have interfered with all other sound-based works in the exhibition. In fact I assisted in setting overall sound levels for all works of the exhibition the day before the opening, mainly turning down volumes for most of the works, including my own, to reduce the sphere of sound leaking from each work, thus protecting the sound integrity of each work not by raising levels of it but by reducing how much sound from surrounding works were permitted to interrupt.

Two months later I had the opportunity to test the same material in another setting. At a workshop with the choreographer Per Roar Thorsnes I played the same sound material, controlled in the same way through a 8 speaker setup, but this time the speakers were not lined up but evenly distributed at all sides of a dance floor. I made sound traverse a line passing through the dance floor continuing out on both sides, ignoring the walls. This fully illustrated one of the strengths of ambisonics, the ability to compose movements in sound independently of the setup that will be used for reproduction. The spatialisation was just as pronounced in this situation, but the sensation of sound blowing past you were gone. Interestingly I experienced another phenomena of great potential: Standing in the audience area, well outside the dance floor, and well outside the speaker rig, the movements of the sound remained pronounced, but now I had a feeling of the movement happening 'over there', at the dance floor. I was sitting outside, hearing sound happening at the stage in the same way that I was outside of the stage were the dancers were moving. I have seen the same observation mentioned by others in mailing list discussions.

Ambisonics have been used in several later works, such as *Staged Bodies*, a media installation in collaboration with Karen Kipphoff. For the first installment at Bergen kunsthall in the autumn 2005 first order ambisonics was used. I was not entirely happy with the results, and for the installation at Kunstneres Hus in 2006 second order controlled opposites ambisonics was used instead, giving more articulation and spaciousness in the sound.

In some projects it has been one of several spatialisation techniques used, such as *Cubic Second*.

3.3.5 Wave field Synthesis

The systems for reproduction of spatial audio discussed so far are all able to reproduce sound so that it can be perceived correctly only in a small part of the listening area, the so-called sweet spot. Wave field synthesis (WFS) is a technique for reproducing signals resulting in a large listening areas which is not restricted to a particular sweet spot. The spatial properties of the acoustical scene can be perceived correctly by an arbitrary large number of listeners regardless of their position inside this area. An introduction to WFS can be found in Spors, Teutsch, Kuntz, and Rabenstein (2004).

Huygen's principle states that any point of a propagating wave at any instant conforms to the envelope of spherical waves emanating from every point on the wavefront at the prior instant. This principle can be used to synthesize acoustic wave fronts of ar-

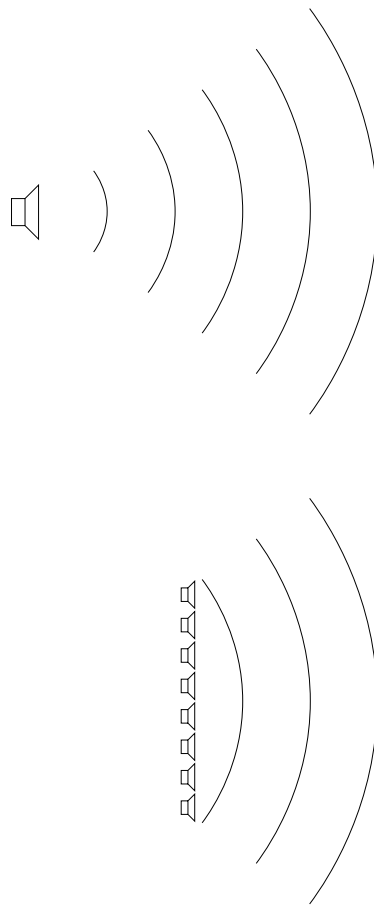


Figure 3.15: A row of speakers are used to synthesize the propagating wavefront according to Huygens' principle.

bitrary shape, using an infinite number of small sound sources, whose waves together would form the next wavefront. WFS approximates this by a large quantity of small loudspeakers positioned next to each other. In practical implementations several simplifications have to be done (Caulkins, Corteel, and Warusfel 2003). The infinite plane is reduced to a line to focus the reproduction domain on the horizontal plane. Often the line is further restricted to a segment therefore limiting the area of correct reproduction. If the line is extended “around the corners” to envelope the room, the sound field of sources originating in the room can be synthesized as well.

The fact that infinite small speakers do not exist leads to two sources for errors or deviations from true sound pressure distribution; spatial aliasing and truncation effects. Spatial aliasing is caused by errors due to the distance from one speaker to the next. Spatial aliasing will occur for all frequencies above f_{Nyq} :

$$f_{Nyq} = \frac{c}{2\Delta\lambda \sin \alpha} \quad (3.29)$$

where c is the speed of sound, $\Delta\lambda$ is the distance from the center of one speaker to the center of the next and α denotes the angle of incidence on the speaker array of the synthesized plane wave relative to the loudspeaker array. Spatial aliasing might lead to a wave field that is not correctly synthesized anymore, resulting in a bad localizable sound source. As can be recalled from section 3.2.1 directional hearing becomes gradually more ambitious for frequencies from 800 Hz to 1600 Hz. If the distance $\Delta\lambda$ between loudspeakers is 10 cm, the minimum spatial aliasing is $f_{Nyq} = 1700$ Hz. If the distance is 25 cm, the minimum spatial aliasing occurs at 680 Hz.

Truncation effects appear when the line is represented by a loudspeaker array of finite extension. They can be understood as diffraction waves propagating from the ends of the loudspeaker arrays. Truncation effects can be minimized, e.g. by filtering in the spatial domain.

I have not so far experimented with wave field synthesis myself. Only recently I became aware of available software, the program WONDER (Wave field synthesis Of New Dimensions of Electronic music in Realtime) for Linux (Baalman 2004)¹⁸. As part of The AES 120th Convention in Paris 2006 an excursion was made to Ircam for a demonstration of their experimental setup consisting of 76 flat panel speakers. The results were very impressive, in particular when the system was put to real artistic use, and not just demonstrated using a three track recording of a soprano and two guitar players, virtually positioned and moved.

WFS seems able to provide a solution for creating a sound field that is consistent throughout the space in a way that none of the previous technologies might be. In works for stage one might imagine a row of speakers positioned between stage and audience offering the possibility of using reproduced sound not only for amplification and and sonic background layers, but virtually position sound on-stage with a much higher degree of precision.

On the downside there are limitations to WFS when used in installation contexts. The system demonstrated at Ircam was designed for a frontal experience. It was combined with Spat (see section 3.3.7 for details) played back over 8 speakers surrounding the space. In my opinion the direct signal reproduced using WFS via flat panel speakers and the reverberation did not blend well. The two classes of signals were clearly coming from different directions, as well as differing in the spectral qualities, probably due to very different types of loudspeakers being used. In addition I did not find the

¹⁸<http://www.kgw.tu-berlin.de/baalman/index.html>



Figure 3.16: Wave field synthesis test setup at the Fraunhofer Institute for Digital Media Technology IDMT in Ilmenau for the Carouso research project 2001-2003 (<http://www.emt.iis.fraunhofer.de/projects/carouso>).

reverb created in Spat sonically convincing. In a horizontal only setup loudspeakers have to be positioned at ear height, thus competing for wall space with visual context desired to be at eye height. Visually the flat panel setup at Ircam appeared surprisingly discrete, as a simplistic white line along the wall. With other kinds of loudspeakers it might appear a lot more intrusive.

In gallery spaces with limited time and assistance for preparing the installation, the amount of work required to set up a system for WFS might be a challenge. WFS seems to require a great deal of technical expertise to use, and is still in a process of scientific development. It is also expensive technology, and as far as I remember the costs of the setup at Ircam was estimated to more than 50.000 €.

In spite of the above discussion, I have to admit that I was so impressed with the impact of the system when used for real artistic context, that it was almost depressing. If I could get my hands on this kind of toy I know exactly what I would use it for, and experimenting with wave field synthesis is high up on the agenda for future work.

3.3.6 Distance-Based Amplitude Panning

The spatialisation techniques discussed so far all assume the listener to be situated at an ideal listening position, surrounded by speakers distributed on a circle (2D) or sphere (3D) with equal distance from the listener and constant angles between the speakers. For sound installations in gallery spaces this might be inappropriate for both aesthetic and practical reasons.

The privileged listening position can be considered an auditive equivalent of the centered viewer facing a Renaissance perspective painting within fine arts. This was understood as a hierarchical relationship between the viewer and the world of painting in front of him, and the position was increasingly criticized from the late 1960s onwards, simultaneously to the rise of installation art (Bishop 2005):

Instead of a rational, centred, coherent humanist subject, poststructuralist theory argues that each person is intrinsically dislocated and divided, at

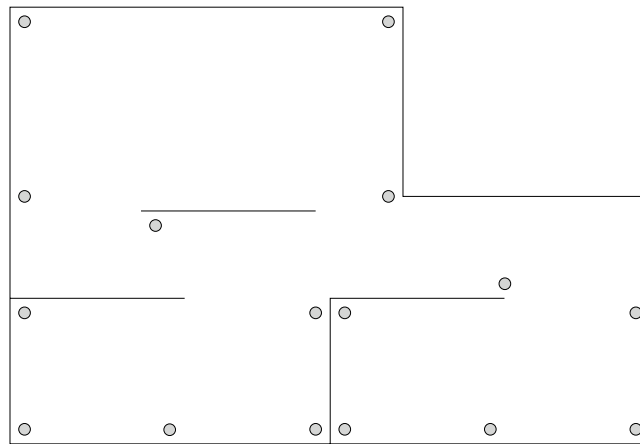


Figure 3.17: 16 channel speaker layout for the *Living room* workshop and installation at Galleri KiT, Trondheim, 2003.

odds with him or herself. In short, it states that the correct way in which to view our condition as human subjects is a s fragmented, multiple and *decentred* - by unconscious desires and anxieties, by an interdependent and differential relationship to the world. (...) There are no one right way of looking at the world, nor any privileged place from which such judgments can be made. As a consequence, installation art's multiple perspectives are seen to subvert the Renaissance perspective model because they deny the viewer any one ideal place from which to survey the work.

Even if an uniform distribution of speakers is desired, the architecture of the space and the visual aspects of the installation might make it impossible to achieve. And in the case that the installation extend to more than one room, the idea of a central listening position becomes utterly meaningless.

The *Living Room* Installation at Galleri KiT in 2003

In March 2003 The Production Network for Electronic Arts (PNEK)¹⁹ arranged a workshop at Galleri KiT at Trondheim Academy of Fine Art, ending with a work in progress installation. The project, *Living room*, involved 12 artists investigating cross-disciplinary and collaborative art practice within new media. Sound for the installation was a collaboration between myself and Thorolf Thuestad. Galleri KiT consists of a number of semi-connected rooms (figure 3.17). We wanted to make a consistent sonic texture for all of the space, made up from several layers. Each layer would be given a virtual position slowly drifting around the space. This way the local sound at any point in the gallery would gradually change as one source moved away and others approached. As the spectators moved around the gallery they would create their own spatial mix of the layers. What layers of sound seemed close by or distant would depend on the movement of the virtual sound sources as well as the movement of the audience, and the audience might very well meet one of the layers at several different

¹⁹<http://www.pnek.no>

locations within the gallery during the course of their visit. 16 speakers were used for the installation, spread around the space as illustrated in figure 3.17.

Initial Tests using Extended VBAP Systems

A number of different approaches were tested during the workshop. E.g. the gallery space was subdivided into a number of square zones defined by a subset of 4 speakers according to the layout of the rooms in such a way that two adjacent zones always would share one or two speakers. For each of these zones a VBAP system distributed sound between the four speakers. As the virtual source moved, it would jump from one zone to the next. Listening tests concluded that this solution was unsatisfactory. The jump from one zone to the next was audible as discontinuities in spatial position, and within each zone the source tended to gravitate toward the closest speaker. There might be at least two explanations for this: Two-dimensional VBAP position the source along the edge of a circle but is not able to create the illusion of sound moving inside the circle. If a virtual source is meant to pass through one region and enter the next, this will instead cause it to move along the boarder of one region and then move onto the border of the next region. In addition the transition from one region to the next will happen abruptly, and is likely to be perceived as an audible change in much the same way as described by Pulkki (2000) when discussing audible loudspeaker directions:

In pair-wise and triplet-wise amplitude panning it can be found disturbing that spread and timbre of a virtual source varies when it is moved. This happens because the spread of a virtual source is dependent on the amount of loudspeakers producing it. When a virtual source is in same direction with a loudspeaker, the virtual source is localized very consistently to that direction. When it is panned between loudspeakers, it will appear spread. This is sometimes described as “loudspeaker directions are audible”.

This artefact is not present with Ambisonic system, when it is used as a panning tool (Malham and Myatt 1995). In Ambisonics the directional spread is not dependent on virtual source positioning directions, because the amount of loudspeakers producing a sound is not dependent on it. However, in Ambisonics system the sound is applied to almost every loudspeakers at one time. The produced virtual sources are spatially spread, and the perceived virtual source direction is dependent on listening position.

Maybe a similar strategy to multiple-direction amplitude panning (Pulkki 1999) could be used to reduce the problem: Instead of describing source position as a single point in space, it could be described in terms of a field of points distributed within a limited region. This would blur the transition from one zone to the next, but would also make spatial positioning somewhat less articulated. While multiple-direction amplitude panning is computationally cheap and already implemented in the Max externals for vbap (Pulkki 2000), a similar approach for the extended setup of several vbap zones would have to be custom implemented. The limited amount of speakers used at any time, and the problem concerning positioning virtual sources inside the circle of each zone were expected to remain problematic, and hence further investigations were not prioritized within the limited time available for the workshop. Instead a new method, *distance-based amplitude panning (DBAP)*, was developed, extending the well-established technique of equal intensity panning of mono sources from stereo to multiple speaker setups.

Distance-Based Amplitude Panning (DBAP)

Distance-based amplitude panning (DBAP), first developed by the author in 2003 for the *Living room* workshop, can be seen as an extension of the principle of equal intensity panning to loudspeaker setups of any size, with no *a priori* assumptions to their position in space or relative to each other.

For simplicity the method will be discussed in terms of a two-dimensional setup with speakers and source all positioned at the same horizontal plane. This can easily be extended to three dimensions.

The position of the virtual source will be described in terms of cartesian coordinates (x_s, y_s) . For a loudspeaker setup with N speakers, the position of the i th speaker is given as (x_i, y_i) . The distance from source to each of the speakers can then be calculated as

$$d_i = \sqrt{(x_i - x_s)^2 + (y_i - y_s)^2} \quad \text{for } 1 \leq i \leq N. \quad (3.30)$$

For simplicity the source is assumed to have unity amplitude. DBAP makes two assumptions. The first is that intensity is to be constant regardless of the position of the virtual source. If the amplitude of the i th speaker is v_i , this implies

$$I = \sum_{i=1}^n v_i^2 = 1 \quad (3.31)$$

The second is based on the inverse distance law for sound pressure for sound propagating in free field as discussed in section 3.2.1. DBAP applies an adaption of this by assuming that the relative amplitude for each speaker is inversely proportional to the distance from source to speaker:

$$v_i = \frac{k}{d_i} \quad (3.32)$$

where k is a constant depending on the position of the source. Combining equations (3.31) and (3.32) k can be found as:

$$k = \frac{1}{\sqrt{\sum_{i=1}^N \frac{1}{d_i^2}}} \quad (3.33)$$

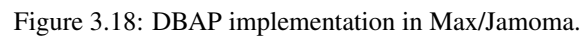
If the virtual source is located at the exact position of one of the loudspeakers, (3.33) will cause a division by zero. Combining equations (3.32) and (3.33) we get

$$v_j = \frac{1}{\sum_{i=1}^N \frac{d_j^2}{d_i^2}} \quad (3.34)$$

From this it can be shown that

$$\lim_{d_j \rightarrow 0} v_i = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases} \quad (3.35)$$

When the virtual source is located at the exact position of one of the loudspeakers, only that speaker will be emitting sound. This might cause the same kind of problems with loudspeaker directions being audible as discussed earlier for vector-based amplitude panning. One way of ensuring a certain degree of spread would be by introducing a *diffusion constant* $r_s \geq 0$ in equation (3.30):



The diffusion constant r_s can be thought of as a displacement in vertical position introduced between the source and the speakers, as if the speakers are positioned at $z_i = 0$ while the vertical position of the source is $z_s = r_s$. The effect of this is that a certain amount of spatial blurriness is introduced. The larger r gets, the less the source will be able to gravitate towards one speaker only. There should be a potential for using this artistically in terms of contrasts between highly articulated or diffuse spatial positioning of sources.

The diffusion constant r now can be understood as a displacement in the fourth dimension of source as compared to speakers. The practical implications remains the same as for the two-dimensional speaker setup, but conceptually the idea of a source offset in the fourth dimension is intriguing.

DBAP is included in Jamoma (Place and Lossius 2006), figure 3.18. The Portuguese artist André Sier have implemented DBAP as C and compiled as external for Max, adding other modes that are variations on the formula presented here. They do not necessarily extend from any acoustic principles anymore, but still are able to produce interesting results. E.g. one of the modes have a very strong gravitation towards the closest speaker. Moving sources while using this mode make the sound jump softly

from one speaker to the next. Conventional spatialisation techniques tries to overcome the discrete speaker layout so that sound positioning can be treated as a continuum of possibilities. From an artistic point of view, that is just one of several possibilities. A very different but interesting approach would be to use widely differing loudspeakers, and move a sound from one speaker to the next to expose their characteristics, so to speak play the speakers, exposing them as non-transparent medium.

Galleri KiT was a highly reverberant space, and it was difficult to get an exact impression of how well DBAP really worked. This was also due to a phenomena when preparing installations that I by now know far to well. The process of preparing an exhibition is generally noisy, and it is challenging to get the time and silence to be able to work on sound material in the space to get an impression of how it interacts with acoustics, and adjust sound to the space. I ended staying up all night prior to the opening in order to have the space to myself. This is becoming something of a bad but sometimes necessary habit. I had some interesting moments when sound seemed to pass through the walls from one room to the next.

Possible extensions of DBAP

DBAP works well as long as the virtual source is positioned within the field demarcated by the speakers. If it moves outside and away from the field, the relative difference in distance between the speakers decrease, and instead of the signal disappearing, it becomes omnipresent. Mathematically speaking it is much the same as increasing the diffusion constant.

A possible solution to this would be to let the loudspeaker positions define a two- or three-dimensional convex hull. In two dimensions a convex hull can be visualized by imagining an elastic band stretched open to encompass all speakers; when released, it will assume the shape of the required convex hull. Tests could then be performed to determine if the virtual source is positioned inside or outside the convex hull. If positioned outside amplitude could start falling, e.g. according to the inverse distance law for sound pressure. One could also imagine the possibility of adding reverb depending on distance from the convex hull. The relative distribution between speakers could be determined by projecting the point onto the border of the convex hull. This would avoid the signal from turning omnipresent as it moves away and fades out. Standard numerical methods are readily available for these tasks.

DBAP could also be used for other means, as it is basically a panning algorithm distributing energy to several destinations scattered over a n -dimensional space. In digital audio workstation software auxiliary effects are sound processing effects that any channel can be routed to, to a smaller or larger degree. One could imagine several DBAP loudspeakers substituted for audio effect units distributed over a two or three-dimensional space. The dry source could be routed in parallel to all effects, and the amount of energy fed to each of them (or the sound level of the return signal from each of them) controlled dynamically by the position of the dry source in this *space of sound effects*. This kind of weighting resemble how Momeni and Wessel (2003) distribute parameter states on a surface to be traversed, leading to dynamic interpolations between presets. That method later formed the basis for the Hipnoscope, the controlling interface particular to the Hipno plug-ins for audio processing (Place, Wolek, and Allison 2005).

The sound artist Justin Bennet has been preoccupied by maps and borders in a num-

ber of works. *Europe*²⁰ is an ongoing project to map European borders with sound. For the installation version, sounds from borders are distributed on a virtual map, and the computer program traverse the borders of the map, tuning into the sounds positioned nearby²¹. His way of traversing the sound recordings resembles DBAP in reverse; providing a system for weighting sources instead of destinations or speakers.

3.3.7 The Ircam Spat Library and Binaural Sound

The Ircam Spat library

Several of the techniques discussed so far have mainly been concerned with virtual positioning of a source in space by means of controlling only amplitude. While able to describe the direction of the auditory event, these methods are not able to create any illusion of a virtual space that the event unfolds in. The two possible exceptions are Directional Audio Coding (DirAC) and wave field synthesis (WFS). Spatial Impulse Response Rendering (SIRR) and DirAC extends vector based amplitude panning (VBAP) by taking into account information concerning the acoustics of the space. Similarly WFS “is a technique for reproducing the acoustics of large recording rooms in smaller sized listening rooms”.

The Spatialisateur project started in 1991 as a collaboration between Espaces Nouveaux and Ircam, with the goal of designing a virtual acoustics processor allowing composers, performers or sound engineers to control the diffusion of sounds in a real or virtual space. Spat, available as part of the Ircam forum²², is an effort to organize and optimize the experimental patches developed in the Spatialisateur project, in order to make them accessible to musicians and researchers who work with Max. It supports multi-channel systems of 2 to 8 loudspeakers in a studio or relatively small concert hall as well as 3D stereo reproduction modes for headphones or two loudspeakers.

The room acoustical quality is not controlled through a model of the virtual room’s geometry and wall materials, but through a formalism directly related to the perception of the virtual sound source by the listener, described by a small number of mutually independent perceptual factors:

- Source perception: Source presence, brilliance and warmth (energy and spectrum of direct sound and early reflections).
- Source - room interaction: Envelopment and room presence (relative energies of direct sound, early and late room effect), running reverberance (early decay time).
- Room perception: Late reverberance (late decay time), heaviness and liveness (variation of late decay time with frequency).

The synthesis of a virtual sound scene relies on a description of the positions and orientations of the sound sources and the acoustical characteristics of the space. The signal processing in Spat is formed by cascade association of four configurable sub-modules controlling source, room, panning and output.

²⁰<http://www.soundscape.com/andere/docs/europa.htm>

²¹Private communication

²²<http://forum.ircam.fr>

The input signal (assumed devoid of reverberation) can be pre-processed by the source module, which can include a low-pass filter and a variable delay line to reproduce the air absorption and the Doppler effect, as well as spectral equalizers allowing additional corrections according to the nature of the input signal.

The room module is a multi-channel reverberator, and can be further broken down into four time sectors according to the discussion on impulse responses in section 3.2.2: The direct sound, early reflections, a cluster containing a denser pattern of diffuse later reflections, and the late diffuse reverberation.

The directional distribution module converts the multi-channel output to various reproduction formats or setups, while allowing to control the perceived direction of the sound event.

Generally speaking, the output module can be used as a decoder for adapting the multi-channel output of the panning module to the geometry or acoustical response of the loudspeaker system or headphones, including spectral and time delay correction of each output channel (Jot 1996).

Binaural Sound

In section 3.2.1 it was discussed how diffraction of sound due to the obstacle of the torso, head and in particular the pinna depends on direction and frequency of the sound, causing the sound reaching the inner ears to be spectrally colored in a manner that encodes information on the direction and distance of the source. The spectral filtering of the pinna can be described as a linear filter, the head-related transfer function (HRTF). Binaural sound is intended to be listened to over headphones, thus bypassing the natural spectral filtering of the pinna. Instead spectral filtering is applied to the sources to be reproduced, so that an illusion of sound coming from specific positions can be achieved.

Binaural recordings can be achieved either using torsos with built in microphones or by placing closely-matched microphones in the ears. At research trips with Verdensteatret this has been a handy and non-intrusive way of doing field recordings, as the microphone used, Soundman OKM head-microphone²³, looks the same as earphones.

Binaural recordings, intended for headphone listening, normally do not translate well in direct stereo playback, with limited channel separation at low frequencies. A crosstalk cancelation method can be used to translate the recording (Eargle 2004).

HFTF differs from one individual to another in the same way finger prints does, and this can cause translation problems: Location of auditory events in binaural recording might differ widely from one listener to another. Head movements and visual cues, important for perceptual localization, are missing and might have impact on the perceived effect. In immersive 3D environments this might be compensated using 3D visual reality systems and head tracking (Rumsey 2001).

Trickster

Trickster, a collaboration with Andrea Sünder-Plassmann combined anaglyphic stereo-photography as digital print and sound reproduced over headphones. The work captured girls at the threshold between childhood and womanhood and their mental spaces, dreams and idols. In addition to be dressed up and photographed in environments different from scenes of their every-day life, they were also interviewed, asked for their thoughts on the past and dreams for the future. The recordings formed the

²³<http://www.soundman.de>



Figure 3.19: *Trickster* Anaglyphic stereo-photography, digital print, sound, 90 x 110 cm. 2005

basis for the sound of the work, presented using CD-players and headphones, with one recording for each portrait. Binaural spatialisation was used, based on Ircam Spat, in the hope that it would create a three-dimensional soundscape similar to the illusion when wearing anaglyph glasses.

Technically this worked well, but conceptually I was not entirely happy with the result. The use of headphones resulted in a situation where the photography and sound would not be approached simultaneously. One would first encounter the photograph, and would then have to put on the headphones to hear the sound. This decoupled sound and image in a way that is contrary to the integration that I am generally aiming for in my work. In addition it was difficult to wear headphones and anaglyphic glasses simultaneously, they got in the way of each other. Taking on headphones implies sheltering or isolating oneself from the surrounding context, reducing or rejecting the gallery space as a social context. Not only was this problematic in relationship to the other spectators, but also when facing the photos. If a portrait represents a kind of *meeting* with the person portrayed, the headphones suggested a contradicting act of cutting oneself off, closing in on oneself. Schafer (1977) describes this eloquently:

The ultimate private acoustic space is produced with headphone listening, for messages received on earphones are always private property. (...) In the head-space of earphone listening, the sounds not only circulate around the listener, they literally seem to emanate from points in the cranium itself, as if the archetype of the subconscious were in conversation. (...) When sound is conducted directly through the skull of the headphone listener, he is no longer regarding events on the acoustic horizon; no longer is he surrounded by a sphere of moving elements. He *is* the sphere. He is the universe.

Headphone listening directs the listener towards a new integrity with himself. But only when he releases the experience by pronouncing the sacred

Om or singing the Hallelujah Chorus or even the “Star Sprangled banner” does he take his place again with humanity.

The collective aspect of the exhibition space has become increasingly important to me, as discussed in section 4.2.4. There is no doubt a strong potential for creating powerful artistic expressions dealing with issues of displacement, isolation and estrangement by means of virtual reality in sound and vision, but at the moment I doubt that I myself will be moving in this direction in the near future.

From a practical point of view, in development of sound for installations binaural sound can be very useful. Sound that eventually will be reproduced using multiple speakers can be ‘virtualized’ by decoding to a number of speakers, describing the position of them relative to the listener, and then apply HRTF. I have done this myself in development of several works, e.g. the sound for *Staged Bodies*. Although several software solutions are available for this, the Ircam Spat library offers a wide range of HRTF filters, enabling experiments to find transfer functions that works for me. Ambisonic decoding to binaural signal is implemented as a module for Jamoma (Place and Lossius 2006), and is thus ready to go anytime I need it.

In January 2005 I did a workshop in collaboration with the choreographer Per Roar Thorsnes, investigating the combination of sound and dancers moving in space²⁴. For the workshop we had 8 speakers surrounding the dance space, and I mainly created sonic textures by distributing three sources using Spat so that they could be moved, shifted into the background or brought prominently onstage. In spite of some very encouraging results, I have not much further experience with Spat used for multi-speaker setups, mainly due to limitations of time and the fact that Spat has to be considered taxing on the processor, thus limiting capacity for other tasks performed in parallel.

Spat supports from 2 to 8 speakers positioned in the horizontal plane, and is flexible in terms of the exact positions of the speakers. Spat, developed in the mid 1990s were pioneering for its time, not least due to the design decision of describing reverb properties in terms of perceptual qualities rather than physical properties. Except for maintenance to keep the library running across new processor platforms, not much development has been done to the library in recent years. It leaves something to be desired in terms of documentation and support for more recent systems for dynamic control of parameters. I would like to see support for the new system for dynamic control of parameters introduced in Max 4.5, and also support for Open Sound Control (Wright, Freed, and Momeni 2003).

3.3.8 Loudspeakers as Direct Sources of Sound

The techniques discussed so far have all aimed for creating phantom images bridging the distances between spaces, allowing the auditory event to be positioned in a space or at least on a continuous circle or sphere. The loudspeakers are themselves objects present in the space. Black speakers in white spaces are generally not invisible. As discussed in section 3.2.1 auditive and visual events tend to merge, leading to a tendency of sound to gravitate towards the speakers if they are visible.

In some projects I have not aimed at being able to move sound continuously between speakers. Instead I have considered the speakers the voices of one or more choirs. Some speakers might have been grouped together, often depending on their physical position, and used for one layer of the sound. The long and narrow space

²⁴My blog posts documenting and commenting on the process can be found here, showing up in reverse order: <http://www.intermedia.uio.no/docudance/index.php/index.php?cat=25>

of Christiansands kunstforening, 60x10 meters, seemed particularly suited for this approach, and most or all of the sound for *Dialogue - Transition* was positioned in the space according to this principle, as described in section 4.2.3. This way of treating speakers as discrete sources in a space, and group them into several choirs providing different layers of the total soundscape has been inspired by the *cori spezzati*.

The Venetian polychoral style developed in the second half of the 16th century at the transition between late Renaissance and early Baroque eras. Linked to the architecture of the Basilica San Marco di Venezia, the sound delay of opposing choir lofts were taken advantage of by separated choirs, *cori spezzati*, singing in alternations at opposite sides of the cathedral (Blessner and Salter 2006).

The delay made it impossible for the two choirs to sing simultaneously, and Adrian Willaert (ca 1490-1562) solved this by writing antiphonal music where opposing choirs would sing successive, often contrasting, phrases. This was further developed by Andrea Gabrielli (ca. 1520-1586), and peaked in the last decade of the century with the style developed by his nephew, Gabrielli (ca 1555-1612/1613). Giovanni Gabrielli specified instruments, including brass, as well as dynamics, and developed the “echo” effect (Crocker 1966).

The spectacular music of San Marco grew famous and was imitated all over Europe, eventually evolving into the concertato style, which in its different instrumental and vocal manifestations eventually led to such diverse musical ideas as the chorale cantata, the concerto grosso, and the sonata. “This was a rare but interesting case of the architectural peculiarities of a single building influencing the development of a style which not only became popular all over Europe, but defined, in part, the shift from the Renaissance to the Baroque era.”²⁵

3.3.9 Distant Voices

Open Spaces

Spatialisation is not only a matter of positioning sound in the space, sound can itself contain spaces. Acousmatic music aims for reduced listening, an experience of sound isolated from outside references and visual information, presented in a darkened room (LaBelle 2006). The question remains how long, if at all, sound can manage to exist in such a contextual void. Some months ago I finally got hold of a re-release of the *Music for Films II* album by Brian Eno, first released as part of a box of eleven LPs in 1983. The record was made at the end of the era of analogue synthesizers, with the Yamaha DX7 about to happen. At first listening it was remarkable how Eno at this time was able to work on sound with no prior references, a tabula rasa of sound. Simultaneously it was quite clear to me that it would be impossible to work with the same instruments and sounds the same way today. They are no longer without a history or references, and any contemporary use somehow would have to take this into account, either consciously or subconsciously. In *La Création de Monde* (1984) by Bernard Parmegiani “reverie, myth, and fantasies of cosmic journeys abound” according to LaBelle (2006). In lack of references a fictional context seems to emerge. If the sound is not of this world, it becomes otherworldly, suggesting the supernatural world of theremin ghosts, the submarine home, a space odyssey or just another green world.

Reverberance is generally used in music production to reposition otherwise close, direct and dry sounds in a space. Most of these spaces are large and closed: the room,

²⁵Quote from Wikipedia entry on Venetian polychoral style, retrieved February 4 2007 (http://en.wikipedia.org/wiki/Venetian_polychoral_style)

concert hall, chapel, church or cathedral. Personally I am more attracted to open spaces. Discussions on how to create the impression of something happening far away in an open landscape is a pet topic in discussions with Lisbeth Bodd of Verdensteatret, in particular voices coming from afar.

Concert for Greenland started with a dialogue between Lars Øyno and the late Per Flink Basse, seemingly left in the ice. Between the actors, Lisbeth and Asle Nielsen, Håkon Lindbäck and myself we tested a wide range of solutions for how to process the signals from the microphones. It became clear that sound needed to be filtered more than having reverb applied to it. We ended up using spectral convolution to experimentally filter the sound, concluding with a solution vaguely resembling a band pass filter, completely blocking low and low mid frequencies as well as the extreme high, and a fairly complex pattern for filtering remaining frequency bands, mainly in the formant register, resulting in a thin sound sufficiently distorted for language not to be legible anymore. The noise of a defect microphone was added to the signal creating an impression of sound having been transported through a medium of some kind. The medium was of no importance to us in itself, but it had the implicit effect of enhancing the impression of distance that the signal had passed. At the time this to a high degree was a matter of empirically finding a working solution suitable to the artistic needs, but having read more about psychoacoustics some parts of it makes more sense than it used to. The filtering of high frequencies resembles filtering of sound in air. Filtering of the low and mid-low frequencies have a similar effect resembling alterations of tone color as gain levels are reduced, with deep frequencies getting less weight, as discussed in section 3.2.1. Thinking of voices heard at a distance the sound of children shouting is the first I come to think of. Maybe this is not only because children tend to shout more often than adults, but also because the nature of their voices make them more apt to traverse greater distances without being submerged in the background noise?

The microphone noise was crucial to the effect. It was first introduced accidentally due to a bad connection in one of the microphones used. When this was repaired so that the noise went away, the distance effect fell apart, making us realize that we had to add this noise from recording to reestablish the illusion. In addition to the impression of sound being mediated discussed before, I also believe that the noise helped creating an ambiance, an auditive background for the scene. With the processed voices just surfacing from this background it might have helped enforcing the illusion of them being far away, almost immersed in the background, almost a part of the auditive horizon suggested by von Békésy.

3.4 A Space for Sound

Robert Irwin described different approaches to site in terms of *site-determined*, *site-dominant*, *site-adjusted* and *site-specific* (Bishop 2005). My own projects so far have tended to fall somewhere between the first three positions. Probably because of my background from music I have an attraction towards abstract aesthetic expressions, often working at the verge where the material as representation break down and become abstract and non-referential. Often the material develop within a collaborative framework, with the implication that it at least to some extent is already defined by the time it is confronted by the exhibition space. Still the acoustic and architectural properties of the room are crucial for the final work, and I have to ensure that I still have a high degree of flexibility and openness in my material by the time I start working in the gallery space. At its best this open up possibilities for a close and fruitful dialogue and

exchange between the gallery space and my work. But depending on the room there is also a potential for the room to be difficult and challenging to work with, not least acoustically as gallery spaces seldom are built for work on sound. Adjustment might escalate to compromise, with sound ending up being site-determined for the wrong reasons.

This almost happened in the autumn of 2005 when working on the installation *White-out*, an installation of sound, video and digital prints by myself and Kurt Ralske for Visningsrommet USF in Bergen. This is a gallery that is extremely open to the outside world, with a large front of glass windows in one of the two rooms and no proper door at the entrance, just an opening to the rest of the USF Verftet, a major venue for all kinds of cultural activity. Across the passageway there is a café. For *White-out* we aimed at investigating the ambience, ambient art and the blurring between the art work / art gallery and the surroundings. The material was mainly sampled on location and further processed. Digital prints and video were based on photos taken at the gallery, including photos of the installation itself, and further processed using slitscan techniques in the custom video software Auvi, developed by Kurt Ralske.

Sound was based on recordings of the ventilation system and other ambient sounds and further processed. Piano-samples were used as impulse responses for a convolution reverb. The sound of knives and forks from the café next door reappeared as piano notes, before being further processed and abstracted. The various layers of sound were distributed to different groups of speakers in the gallery. The investigation of the sound implied an investigation of the space.

Acoustically the space turned out to be very challenging. Almost all surface were hard and reflective, mainly made from concrete. It was difficult to predict the effect of the reverberation on the sound, and any sketches I made away from the room invariably failed to work when I tested them at the location. The loudspeakers were positioned sideways at the floor, and this was probably a really bad idea, severely degrading the intelligibility of the direct signal. As soon as I started reproducing sound in the space, all surrounding ambient sound and noise seemed to increase in intensity. The space was probably more noise than I initially was aware of, but as soon as intentional sound was introduced so that the ears started paying attention, they also were forced to pay attention to the surrounding noise. The automated door just outside the gallery space was particular bad in this respect, a loud, percussive sound intruding whenever somebody passed through the door for a smoke. The fan noise of two video projectors used in the inner room also caused problems.

Twenty minutes before the opening I added yet another layer of sound, an abstracted almost white noise moving between several of the speakers, partly masking the background noise. This finally managed to solve the problems, creating a sufficiently concentrated environment to hold on to attention, not having it literally leaking out the door.

Cubic Second at Hordaland kunstsenter was in part a response or reaction to this experience. I wanted to take complete control of the space and as far as possible remove any object or information that was not intentionally chosen to be there as a part of the installation. In an almost *etude*-like manner I wanted to see how far the gallery space could be adjusted and transformed to fit my needs. I wanted to attract attention towards the elements I worked with in the installation, while rendering the rest of the room transparent or invisible.

The gallery space at Hordaland kunstsenter is a rectangular white cube (figure 3.21). For the installation the walls were painted black. An extra wall was mounted inside of the door to shelter the room from the entrance in terms of sound and light. 16



Figure 3.20: Installation views from *White-out*.



Figure 3.21: Hordaland kunstsenter. The empty gallery space May 2006. Windows were permanently sealed during the summer.



Figure 3.22: Tests of spatialisation techniques and acoustic measurements at Hordaland kunstsenter in May 2006.

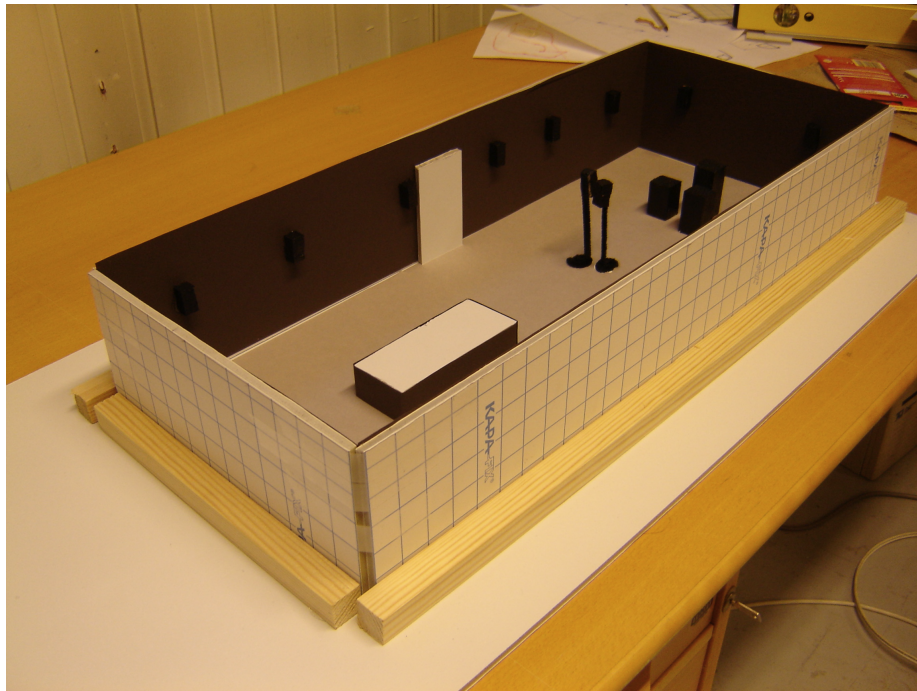


Figure 3.23: Modeled used for planning the design of the *Cubic Second* installation.



Figure 3.24: The installation was prepared in five hectic days and nights.



Figure 3.25: *Cubic Second*. The large back projection. Photo by Karen Kipphoff.

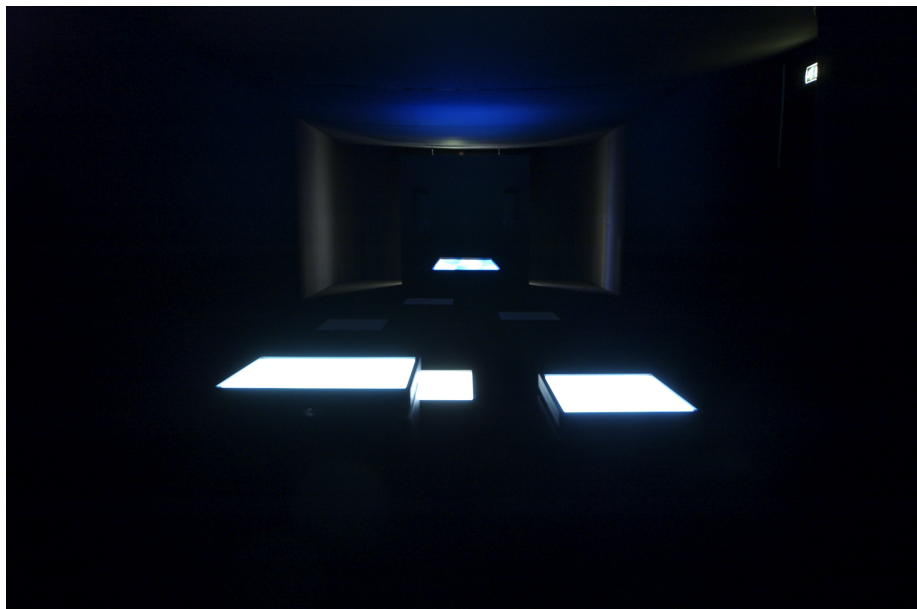


Figure 3.26: *Cubic Second*. Installation view. Photo by Karen Kipphoff.

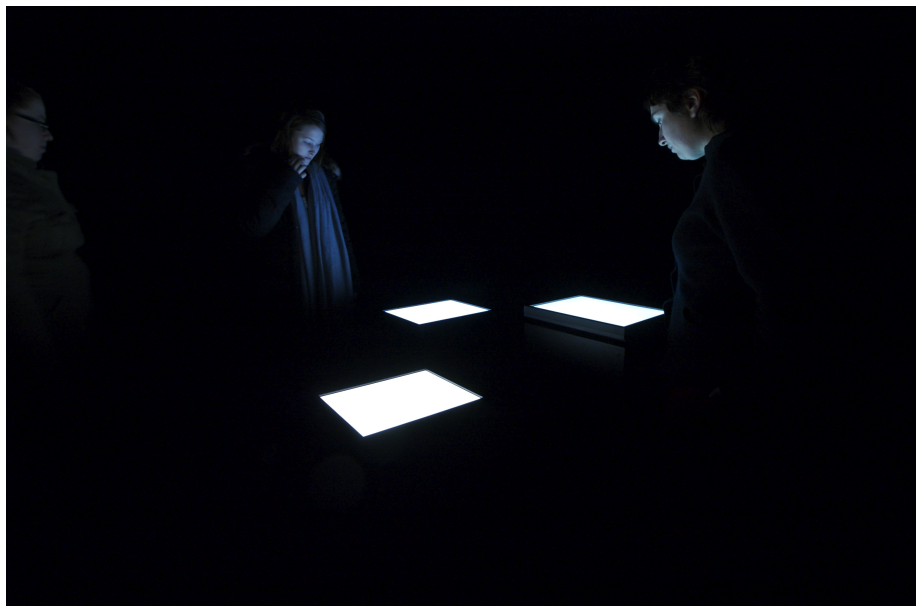


Figure 3.27: *Cubic Second*. Video on three lcd screens. Photo by Karen Kipphoff.

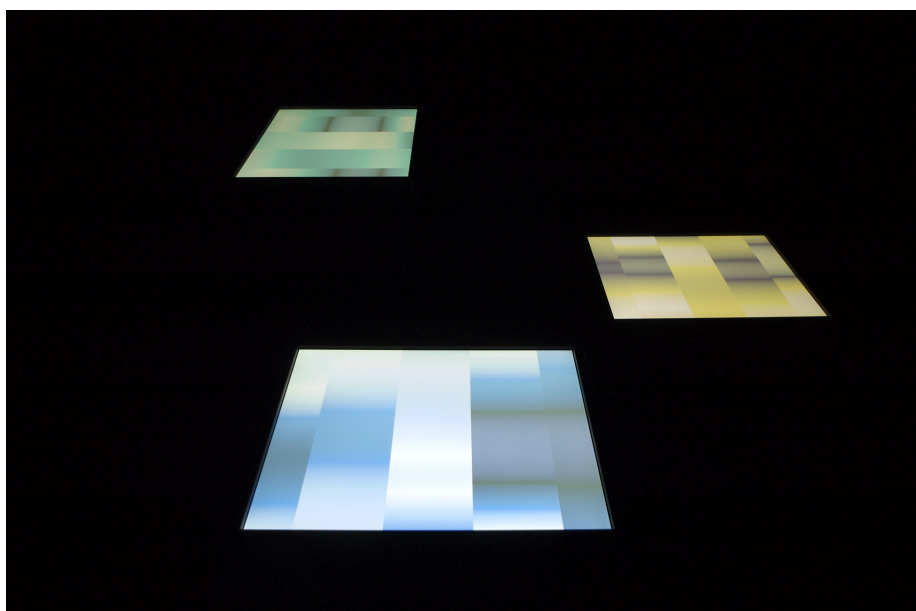


Figure 3.28: *Cubic Second*. Video on three lcd screens. Photo by Karen Kipphoff.

loudspeakers were mounted at the walls in ear height surrounding the space. A black felt carpet was used at the floor. In front of the loudspeakers, all the way around the space, semitransparent black textile was stretched from ceiling to floor, rounded at the corners, softly erasing the edges of the room. Black textile was also used in the ceiling to darken it. The textile served several purposes. It made the loudspeakers invisible, thus preventing visual cues from causing auditory events to gravitate towards the loudspeakers. At the same time it was light enough not to let the sound through without any noticeable dampening. In the ceiling it did not hinder sound from reaching the acoustic tiling, the only permanent absorbing surface of the room. In the installation the carpet aided in absorption at high frequencies. The textiles also helped hiding the technical parts of the work.

This text might leave the impression that I am very concerned about the technical aspects of the works I create, with the potential risk of the works to end up having a certain geek factor. I do spend a lot of time and energy researching technical solutions that might help me achieve what I want, but they are primarily means to an end. I consider my practice to somewhat resemble how a professional pianist works. Several hours a day are spent rehearsing scales and other exercises in order to develop and maintain a high level of technical skills. But the development of technical skills is not a purpose in itself, and it would be out of question to play scales and etudes in concerts. The skills are required in order to be on top of the material one is working on, and to be able to articulate oneself artistically without technical limitations becoming a hindrance reducing the impact that message can be delivered with.

The spatial illusions created by the various techniques used for distributing sound to the 16 loudspeakers would generally work best away from the walls. I therefore wanted to form the visual part of the installation in a way that would invite the audience towards the middle of the room. There is a cellar downstairs from the space, and the floor can open up to a stairway in the middle of the room. The idea occurred to me that I could use transform this into a horizontal back projection. That would also position the projector outside of the space, reducing fan noise.

In order to introduce more diversity in the room, inviting the audience to move around the space, I added three lcd screens mounted horizontally as a group at the far end of the room. In between the lcd screens and the large back projection three cubes meant for sitting were positioned.

The video material was abstract, generated by traversing spaces of black and white noise. The resulting gray-scaled images were mapped to algorithmically generated streams of color, created by traversing a subspace of RGB color space defined by four points. This was inspired by a book on painting I got as a child. The book started out creating paintings from a very limited palette of only two or three colors. As the number of colors used increased, the images became successively more impressive and less interesting. The level of abstraction introduced by using a limited palette interested me far more.

3.5 Data Space

Spatialisation using several loudspeakers was an inherent potential of the technology from the very moment that it became possible to capture, store and reproduce sound. Blumlein started experimenting with stereo in the 1930s (Eargle 2004), Disney's *Fantasia* used multiple channels for reproduction as early as 1940 (Kay, Ghent, Chumney, and Lutkins 1997), *Williams Mix* (1951-53) by John Cage used eight monau-

ral magnetic tapes, Stockhausen created the originally 5 track version of *Gesang der Jünglinge* in 1955-56 (Holmes 2002), and in 1958 somewhere between 300 and 450 loudspeakers were used for the spectacular Philips Pavilion (Treib 1996).

Spatialisation is not endemic to work on electronic music and sound, but a general tendency of postmodernism, a tendency of privileging space over time, flattening historical time and refusing grand narratives, according to Manovich (2001). He sees this as inherent to new media:

It is perhaps more accurate to think of the new media culture as an infinite flat surface where individual texts are placed in no particular order. (...) Any RAM location can be accessed as quickly as any other. In contrast to older storage media of book, film and magnetic tape, where data is organized sequentially and linearly, thus suggesting the presence of a narrative or a rhetorical trajectory, RAM “flattens” the data. Rather than seducing the user through a careful arrangement of arguments and examples (i.e. the rate of data streaming, to use contemporary language), simulated false paths, and dramatically presented conceptual breakthroughs, cultural interfaces, like RAM itself, bombard the user with all the data at once. (...) Time became a flat image or a landscape, something to look at or navigate through. (...) In following the general trend of computer culture towards spatialization of every cultural experience, this cultural interface spatializes time, representing it as a shape in a 3-D space. This can be thought of as a book, with individual frames stacked one after each another like book pages.

From this perspective spatialization of sound bridges towards another important principle underlying much of my work; the use of algorithmic processes as a means to create continuous flow and variation. This is further elaborated in section 4.2.1. I have been thinking of algorithmic processes as a continuous stream of realizations of a ideal sound meta-object in a platonic sense, influenced by how the composer and music critic Lambert understood the works in three parts by Erik Satie:

Satie's habit of writing his pieces in groups of three was not just a mannerism. It took place in his art of dramatic development, and was part of his peculiarly sculptural views of music. When we pass from the first to the second *Gymnopdie* we do not feel that we are passing from one object to another. It is as though we were moving slowly round a piece of sculpture and examine it from a different point of view, while presenting a different and possibly less interesting silhouette to our eyes, is of equal importance to our appreciation of the work as a plastic whole. It does not matter which way you walk around a statue and it does not matter in which order you play the three *Gymnopdies*. (Shlomowitz 1999)

This is not very different from how Manovich (2001) considers the human computer interface (HCI) to be a form of framing: “Just as a rectangular frame in painting and photography presents a part of a larger space outside, a window in HCI presents a partial view of a larger document. But if in painting (and later in photography), the framing chosen by an artist is final, computer interface benefits from a new invention introduced by cinema - the mobility of the frame.”

The sound of my installations could be said to be similarly framed, in part by means of algorithms determining how the computer programs traverse the possibilities of the

material, how the material is modified and combined in real time. On the other hand the audience by moving around the space that sound is distributed to also participate in what could be considered a customization of this framing. Depending on their position, the balance between different layers of the total soundscape and the perception of the various layers will differ.

Concluding this discussion, I find an emerging synthesis technique particularly interesting in this respect. Corpus-based concatenative synthesis is based on a database of grains of sound analyzed for their acoustic properties. If the grains are analyzed with respect to n different descriptors, the database can be considered a n -dimensional data space, with each grain representing a point in this space. Concatenative real-time sound synthesis (CataRT) (Schwarz, Beller, Verbrugghe, and Britton 2006) then is a matter of traversing this data space, using the grains positioned the closest. This can be seen as a content-based extension to granular synthesis providing direct access to specific sound characteristics.

Diemo Schwarz was kind to demonstrate CataRT for me at Ircam in May 2006, and later provided me with a copy of the code for further experiments. I first intended to use it for the installation *Cubic Second* but gradually concluded that the technique offered vast new possibilities for orchestration of electronic sound that required further in depth studying and experimenting to be able to fully take advantage of the potential. If my initial aim was to work *inside* of sound, I had not imagined that it would be possible to move inside the characteristics of sound in such a literal way as CataRT offers.

Chapter 4

Issues of the Performative

4.1 Introduction

January 1998 I found myself living in Lüderitz, a small fishing village at the Skeleton Coast in Southern Namibia. My wife had got a two year engagement there, and spent a fortune on a Macintosh computer, a synthesizer and software for music creation so that I could come along. I sat down to read manuals and learn the new toys, and quickly fell in love with Max, a program that according to one of its creators does nothing (Zicarelli 2002). Within a few months I got the idea for a sound installation, and I spent a month realizing it.

The music of Brian Eno, in particular his Ambient Music, had been a major influence for a long time already. I was more interested in working on the sensuous qualities of sound more than structural large scale compositional form, and I dreamed of making music resembling places rather than linear developments. As a student of composition I had difficulties translating this to notated music. The blank papers of sheet music got increasingly difficult to deal with, and I did not know how to walk the timeline. The music I did write mostly got poor performances by fellow students that regarded this as their student duties top priority.

In Lüderitz, while patching objects on a data screen, I felt that I had finally found the direction I had been searching for. This seemed to be a way of creating music that was more environment than development, and I was able to work directly on the shape of the sound. In addition the prospect of ridding myself of the dependency on musicians felt like a big relieve.

The transition from music compositions to installations and from notated music to algorithmic data-based processing raised fundamentally new problems to me. Often the technical compositional methods that I was trained in, such as counterpoint, harmony and orchestration, seemed difficult to apply within the new domain, and I had to search for other ways of approaching the creative process. I had to develop new (at least for me) working methods and deal with the many new artistic and technological issues continually arising. For a long while I felt that I was dealing with many different issues in parallel, a complex maze of red threads running through the projects I was involved with. Gradually I have realized that many of them are interrelated, dealing with functions and roles involved in creating and experiencing music and art in the contexts of electronic sound and music, installation art, multi-aesthetic projects, interactivity and computer-aided real-time processes. I am by no means the first to challenge the

established musical roles of composer, musician and listener of the Western art music concert practice (de Haan 2002). Still, I have had to search for my take on these problems, and in the process I might have made some observations and reflections of interest to others.

Emmerson (2001) provides a stimulating discussion of music, and in particular the music of the 20th century in terms of how it relates to body and environment, defined thus:

Body

The body generates many rhythms and sensations with cyclic periodicities lying within the duration of short-term memory. The most important are breath, pulse, and the limb movements of physical work, dance, and sex. These are a product of our biological evolution, our size, and our physical disposition in relation to the mass of the earth hence its gravitational field and would be different if we had evolved to be the size of a bat or an elephant, or if the earth had possessed a different mass.

Environment

The environment has a different time scale with both periodic and aperiodic rhythms and this is often beyond the limits of short-term memory. This often necessitates repeated listening and consignment to long-term memory, thus encouraging contemplation and consideration: water, wind, the seasons, landscape.

I have found it fruitful to consider my own work in terms of this dichotomy of body and environment, and this will be a main subject of this chapter.

The body of work I have been doing so far can roughly be divided in two categories: Works for gallery spaces, and projects presented live in a stage or concert setting, involving performers. The shift back and forth between these two paradigms has often raised opposing questions, and thus illuminated issues of interest from opposing angles. The chapter will start out discussing experience and thoughts gained from gallery-based projects and then go on to do a similar discussion concerning live art projects.

4.2 Performative Issues in Installation Contexts

4.2.1 Music as Process, Music as Weather

Music as Process

When sound is reproduced over loudspeakers, there seems to be two fundamental modes for the construction and reproduction of the sound.

It can be pre-recorded and pre-composed, existing as a multitrack, possibly synchronized to other media flows, e.g. light or video. The Philips Pavilion is an early and extraordinary example of this kind of spectacle, combining architecture, multi-channel sound, light, film and projected images (Treib 1996) In spite of being presented in a custom-built space, the linear form of the audience experience was maintained in a similar way to film, theater or concerts. The audience entered the space in groups, stayed for the 8 minute spectacle, and collectively left the space to give room for the next group. The compositional approach to time within this kind of presentations can

be considered an extension of concert music, in spite of the spatial and sonic aspects of the music by Varèse being radically new.¹

A similar approach to time can be seen in numerous video installations where the video is of fixed duration, continuously looping, starting with a title and ending with credits and copyright notes. The video signals having the form of a line segment in time, with definite start and end points. The statistical chance of entering the space of the screening as the video starts is small, and I always find myself entering in the middle of it, getting a feeling that I have already been missing something. I might stay to watch the first part having already seen the end, but the form is disturbed. In addition, once I have seen once through, I to often get a feeling of having seen it all, and decide to move on. This is in contrast to standing in front of a good painting, one can always stay there for five more seconds, and five more, and five more.

Of course there are exceptions. The videos of Jeremy Welsh for the LMW collaborations are non-narrative, abstract and synthetic (at least to a certain degree) containing no markers or indications of start or end points. I am not sure if I would have discovered that the shortest video he has come up with so far was looping at all, if it was not for the DVD player always freezing for a second at the loop point ².

The other possibility is substituting the composition for a meta-composition: An algorithm composing the actual sound in *real time* according to rules defined by the artist. Since starting to work with Max this has been my preferred mode of working, motivated both from aesthetic needs and a creative process that has felt more rewarding.

Brian Eno has long been an advocate of generative processes (Eno 1976). Influenced by minimalism and Fluxus he has also insisted on simplicity, often citing the phase compositions for tape by Steve Reich (Eno 1996b) and *The Great Learning* by Cornelius Cardew (Eno 1976) as models for his own work. The track 2/1 from “Ambient 1: Music for Airports”, released in 1978, is made up from a number of tape loops (Tamm 1995). Each loop contains a single held note, a female voice subtly processed, followed by silence. Thus each tape loop produce the same note over and over again at regular intervals, interspersed by silence. Each loop holds a different note, and the length of the loops differ. The voices combine to construct a texture, the different layers combining in ever-new ways. Sometimes the texture becomes a dense chord, at other times there is complete silence for a while.

In spite of the simplicity of the system, it already has a time structure that expands beyond most composed music. Music to be performed by more than one musician commonly need a strategy for synchronizing the performers. The beat, in addition to stirring a bodily engagement, also functions as a chronometer, ensuring that all performers share the same time. The tape loops of 2/1 on the other hand can be seen as creating several layers of independent coexisting timing. All loops are based on recordings of the same voice, further suggesting an impression of several parallel layers of time, or maybe just as much several layers of sound that have transcended time.

The LP recording of 2/1 has a fixed duration, but this can be considered an excerpt of a work of infinite duration, a system of tape recorders endlessly looping. This is one of the qualities of generative music that has been attractive to me: The ability to create

¹ Another departure from common practice for multimedia at the time was the insistence on independence of audio versus light and film. Although durations and onset were synchronized, the content of the sound was not synchronized to the content of film and light (Treib 1996). In this respect The Philips Pavilion clearly differs from another pioneering multimedia work exploring sound in space, Disney’s *Fantasia* from 1940 (Kay, Ghent, Chumney, and Lutkins 1997).

² A black and white video lasting one minute only, made for the installation in a container at the Quart Festival in 2006

soundscapes of infinite duration producing endless non-repeating permutations within a limited and controllable range, thus escaping the single repeating and predictable loop discussed above.

Algorithmically generated music do not only differs in time scale of the finished result, the compositional approach to time in the creative process also becomes radically different. Let us imagine for a while that we have access to the studio of Brian Eno. The tape recorders are all lined up and playing, and there is plenty of equipment laying around. It is tempting to tinker with the system. We could take the loop with the deepest note, cut away the silence, and turn it into an endless drone. This would cause a radical change to the work. We could slow down the speed of one of the players, altering the scale, tone color and intensity of the piece. We could do further effect processing on the signal of one or more players. Most of these changes would not take long to do, and we could then sit back and listen to the effect of it. The interesting thing is that the change would have instant effect on a global scale. It is not as if we did a change in bar 39 of an orchestral score. The change would not be perceived until bar 39 was reached. With the generative system, changes take effect immediately. From the time I started using Max I have felt that this way of working with generative systems resembles creating a painting. You apply paint somewhere on the canvas and step back to see the effect of it. The impact of the change is global, and recognized and experienced instantly. The creative process gets very different, and to me have been far more stimulating. I always had problems coping with timelines.

Interestingly, there seems to be a similar difference in working methods between the Cologne Studios and GRM in France, the work at the Cologne Studios being dominated by structured realization of pre-planned compositions, while the activity at GRM “begins with a prepared sound material, which is molded into its final form by a process of experimentation, trial and error, perhaps following unexpected paths to goals that were never foreseen initially” (LaBelle 2006). Working on the sound of music more than music as structure might inherently encourage a processual way of working.

Weather as Music

When I started working on sound installations, I was aiming for environmental rather than bodily qualities. The ambient music of Brian Eno was an important inspiration (Eno 1996a), but equally important was the accumulated experience of extended hiking in the Sandviken mountain outside Bergen over several years. I would come back regularly to the same spots, observing how they changed with the time of the day, the weather and the seasons. I wanted my installations to be experienced in a similar way, locations that could serve as points of reference, already existing by the time the spectator gets there, and continuing to vibrate after the spectator has left. I imagined that the spectator could come back several months later, noticing a slight shift, the installation having entered a different season.³

Having previously studied oceanography and meteorology, the idea of using weather data was obvious. The first attempts were made while working on *textitElektropoesia* in 2004, an installation of sound and video for the *Electrohype* biennial for computer-based arts at Malmö konsthall, made in collaboration with New York based video artist Kurt Ralske.

Tidal waves, driven by the gravitation of the moon and the sun, consists of a large number of harmonic components, with M2, the diurnal component driven by the moon,

³This idea could only have been fully realized in permanent installations. To this day all installations have been temporary, a quality that I have appreciated more and more as time has passed.

being the dominant one by far. The tidal data was therefore expected to function as an oscillator with a regular pulse, but changes over time in amplitude and equilibrium value. The data used were not calculated tidal values, but measured values for sea level, acquired from the web pages of Norwegian Hydrographic Service⁴. Water level is also strongly influenced by weather systems, and in Bergen there are a lot of them. I was expecting this to introduce additional irregularities in the signal. Data were sampled every 10 minute with a one centimeter precision. If the data were considered and treated as audio, with each measurement being one sample, the signal could be considered to have good sample rate with more than 100 samples per M2 period, but poor sample depth (somewhere between 7 and 8 bits). At a sampling rate of 44.1 KHz 3 years of data became an audio loop lasting for about 2.6 seconds.

The data was used both as control data, mapped to various parameters and as an audio signal. At a sampling rate of 44.1 kHz one year of data would result in an audio loop lasting for slightly less than one second. The data was played at somewhat lower speeds to bring the frequencies of the sound down to more interesting ranges. The sound could be described as containing a clear and penetrating deep frequency (the M2 component) with amplitude modulation and hiss and cracks at higher frequencies⁵. For *Elektropoesia* I used low and high pass filters to separate the humming sound and the bright noise, so that they could be treated as two independent layers.

One of the challenges when creating algorithmic processes is to negotiate stasis and variety in an artistically interesting way. Often the challenge is to produce variation making the result predictable in an unpredictable way. The sea level data set played back at very slow speeds functioning as control data have proven excellent in this respect. In *Elektropoesia* I used the data to control panning of sources, cross-fades between layers, gain levels, etc., often using ten minutes or more to scan through the data set. By scanning the data set at different speeds for each of the parameters controlled, a multitude of frequencies were present, never combining twice in the same way. The sea level data set has continued to be part of the box of compositional tools used for my work, and have been used again and again, most recently to control drifting spatial positions of sound layers in the installation *Cubic Second* at Hordaland kunstsenter in the fall 2006. I have so far never felt it conceptually important to communicate this method to my audience. In the next project exploring meteorological data it became much more important.

In 2005 I was invited to contribute to the touring exhibition *Generator.x*⁶, curated by Marius Watz. The *Generator.x* project is a conference, exhibition, concert tour and weblog examining the role of software and generative strategies in current digital art and design. Marius Watz started the blog well in advance of the conference and opening of the exhibition. Following it with interest, I observed that much of the generative visual arts blogged appeared as abstract, formalized aesthetic surfaces. Sometimes the lack of a context for the works felt potentially problematic, as if they were about to implode into self-referential intellectual and aesthetic games with little or no reference or relevance to the outside world, including a broader tradition and contemporary field of image and paintings within fine arts.

I decided that I would use the opportunity to revisit and reexamine my thoughts and strategies on generative processes by finally using weather data to control the installation. In the act of doing so I would also revisit important aesthetic influences and

⁴<http://vannstand.statkart.no/skjema2.php>

⁵Two sound examples have been posted to the blog (<http://www.bek.no/Members/lossius/lostblog/362>).

⁶<http://www.generatorx.no>

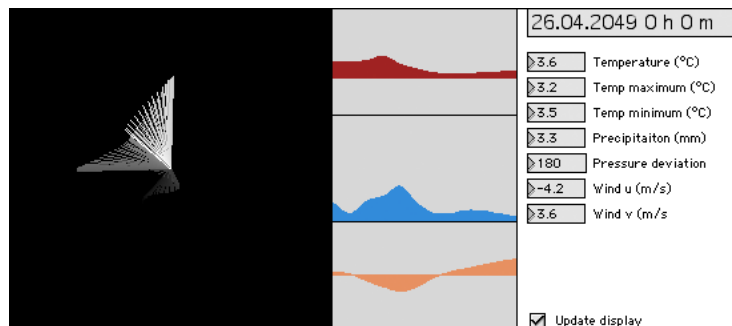


Figure 4.1: Excerpt of the Max patch for monitoring of weather data: Wind vector (x and y components), temperature, precipitation and deviation of pressure from mean value.

the relationship between generative processes and their aesthetic motivation, paying homage to John Cage with the title of the work; *music as process, music as weather* (Cage 1989). Finally I wanted the work to have a political dimension, a resignation to the fact that my desire for creating a work of art escaping the bodily, the performative and traces of human influence would be doomed to failed. Any attempts at using weather data for these reasons would be sheer escapism if the biggest human intervention ever was not taken into account: The greenhouse effect.

Climate models are computationally extremely demanding, calculating weather prognosis on a world basis with multiple iterations per day, in order to study potential changes for centuries to come. Although they are not meant to serve aesthetic purposes, they are among the most complex and advanced existing computer-based algorithmic processes. The Bjerknes Centre for Climate Research⁷ gave me access to data from their simulations studying long-term development of the climate to the year 2100 and beyond. The parameters controlling the sound of the installation were driven by a long term prognosis of the weather in Bergen, a daily prognosis for the years 2046-2065. Time was accelerated so that one day would pass in a minute, the complete data set traversed in a little less than a week. I did not aim for a straight forward sonification (sonic illustration) of the data set, rather an artistic interpretation.

The data set accessible was less detailed than initially expected, for various reasons related to concerns of storage and what data have relevance in a research context. While the numerical model integrates in steps of 30 min. only daily mean values were available for sea level temperature, pressure, precipitation and wind speed and direction (figure 4.1). In order to get a richer set of data to work from, secondary control data were generated; shift between daylight and night, the kind of precipitation (rain, snow or dry), wind turbulence and energy and estimate for cloud cover. Interpolations between daily means were done using cubic splines (Dahlquist and Björk 1974).

The sound processing used recorded material from an earlier project, a composition for mezzo-soprano, cantele and live electronic sound processing based on the thirteen-century song *Ex te lux oritur*, a hymn for the wedding of Princess Margaret of Scotland and Eric II of Norway in 1281 at Håkonshallen in Bergen, the earliest preserved profane music from Norway⁸. Thus *music as process, music as weather* in an abstract way

⁷<http://www.bjerknes.uib.no>

⁸This was a commission for the Holberg Prize Award Ceremony 2004 at Håkonshallen, December 3



Figure 4.2: View inside the custom-built cube used for the installation *music as process, music as weather*.

would be referring to both the past and future of Bergen.

The sound material was further processed in real time using granulation techniques for speed and pitch shifts and time stretches, as well as dynamic filters, reverb and delays. Ambisonics (see section 3.3.4) for distributing sound to four custom-built flat panel NXT loudspeakers⁹ with black cotton canvas surfaces. The installation was presented inside a container-shaped cube built for the touring exhibition from black molton fabric stretched across an aluminum framework (figure 4.2).

Extensive research on mapping of musical data have been carried out in the later years in relation to research on musical gestures and interfaces for musical expression¹⁰. Mapping of climate data was a combination of straight-forward one-to-one mappings (one data parameter mapped to one sound processing parameter such as wind direction driving the source position and wind speed controlling gain levels) and more complex mappings of m data parameters to n sound processing mappings using linear algebra, utilizing the Max libraries FTM for complex data sets (Schnell, Borghesi, Schwarz, Bevilacqua, and Müller 2005) and MnM for advanced mapping

2004, and realized in collaboration with Agnethe Christensen (mezzo-soprano, cantele).

⁹<http://www.nextsound.com>

¹⁰A valuable source for information on the subject is the annual conference on New Interfaces for Musical Expression (NIME): <http://www.nime.org>

of data (Bevilacqua, Müller, and Schnell 2005). MnM. I hoped that he would be able to find a number of extreme weather situations (e.g. cold and calm, warm and calm, warm and rainy, strong low pressure systems with strong wind and precipitation), set sound parameter settings for each of these, and then use the Max external `mmn.matmap` (Bevilacqua, Müller, and Schnell 2005) for continuous morphing between these preset states depending on weather data values. It was not straight-forward to achieve this, and the goal was not fully achieved, probably due to a lack of methods for optimal ways of mapping the m -dimensional set of control data to a subset of the n -dimensional space of sound parameters. I experienced that some of the parameters tended to jump back and forth between minimum and maximum values, seldom using the more subtle nuances between the extremes. This might indicate that the mapping was defined in a sub-optimal way. Normalization of the data sets and principal component analysis (Bevilacqua, Müller, and Schnell 2005) might have improved the results¹¹, but more experience and understanding of the theory is required before I am able to fully explore the potential of these mapping strategies.

4.2.2 The Memory of a Gold Fish

Dobrian (1988) suggest that application of new technology tend to progress through several stages:

The first inclination is to use the new technology to duplicate already existent functions. (...) This may be in order to demonstrate the usefulness of the technology, or it may be to eliminate the traditional (perhaps tedious, dangerous, or otherwise undesirable) method of performing the function. The second way of using technology is to perform previously unperformable but desired functions (telecommunication, for example). A third, less frequent, use of technology is to discover new, previously unconceived functions.

Duplication of already existing functions might serve another important purpose, as test cases for evaluating how well the new technology performs.

Although I initially wanted to work on sound emphasizing contemplation and distance rather than body action and involvement (Emmerson 2001) I soon felt that I would not have had much of a choice anyway. The expressive potential of the synthesizer seemed limited. As I started exploring MSP (Zicarelli 1998), the library for digital sound processing in Max, I realized that it would be challenging to create sounding results that would get anywhere near the expressive performative qualities of most acoustic instruments. I was not necessarily aiming for those qualities in my art works at the time, but I would have preferred being permitted to make the choice myself, not having technology doing it for me.

Having played violin and viola earlier on I appreciate both the expressive possibilities of the instruments and the long tradition of musicianship nurturing it. I do not subscribe to the idea that all changes are for the better, within the arts changes tend to be for something different. The changes might be required, but in the process of opening up new artistic possibilities and addressing new issues of importance, other options tend to forgotten, down-prioritized or inaccessible. At a time of transition it is worth looking back and ask what might lost in the process. The development of music

¹¹ Personal communication with Rémy Müller at IRCAM.

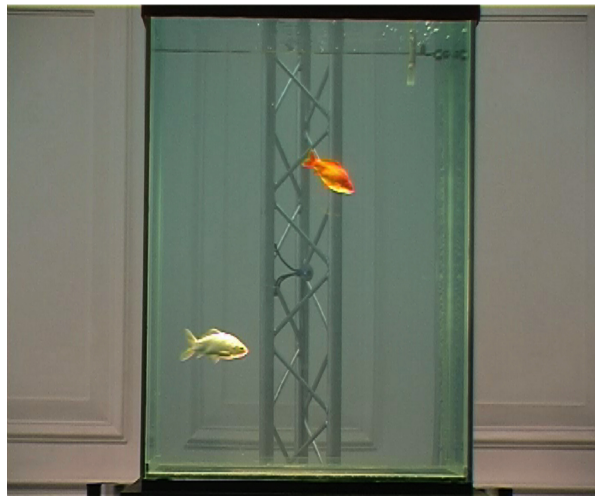


Figure 4.3: Installation view of *Ekkofisk* from Baby in Amsterdam at World Wide Video Festival 2001. One of the web cameras used for motion tracking can be seen behind the aquarium.

technology and the artistic use of it has been a continuous negotiation between expanding conceived possibilities versus limitations due to technical and financial feasibility, ever-increasing range of parameters to control versus problems concerning how to manage them, pushing and bending technology to suit artistic needs versus exploring the aesthetic implications that the technology itself seem to suggest, and the advent of research versus commercial wrapping as products by the music industry. There is nothing radically new about this. Examining the western art music one will find a continuous and similar negotiation between abstract musical ideas or structures and concerns for idiomatic instrumental writing. The important point here is to have a critical awareness concerning limitations imposed by technology so that they can be opposed or explored more efficiently, depending on artistic needs and interests. At the moment commercially available music technology generally seem to be more restrained in expressive capacity than need be, an observation supported by the extensive ongoing research on musical gestures and new interfaces for musical expression.

Ekkofisk

The first serious attempt at addressing these issues in my own artistic work was the sound processing for the installation *Ekkofisk*. The installation *Ekkofisk* was made in collaboration with Reinert Mithassel in 2000-01. As explained in section 3.3.3 two surveillance cameras were used to track the three-dimensional positions of two goldfishes in an aquarium. The resulting data were used for real-time sound synthesis, a synthesized singing voice for each fish when they were in the vicinity of each other, and bell-like sounds if they moved further away. The initial idea when we started developing *Ekkofisk* was to have a more or less instant mapping of position data to various sound parameters. Soon it became clear that this would not be possible to combine with the decision to represent data as singing voices, due to the (in hindsight) obvious fact that the act of singing is a continuous negotiation between acting at the present

moment and maintaining an awareness of how this relates to the past and future. This is probably common to most or all kinds of musical activity, as well as many other kinds of artistic and intellectual activity.

I have a general impression that many systems of interaction and real-time processing are designed with a tendency to emphasize immediate mappings, putting much less emphasis on time-span memory and causal relationships between events in time. This is instead left to the discretion of the person(s) that are to interact with the system, to be determined by how they *play* the system. In order to provide a background for the further discussion, a detailed description of the mapping process and sound synthesis used for the Ekkofisk installation will be given.

Voice Synthesis Algorithm

The voice synthesis to a large degree was done according to techniques and theories discussed by Dodge and Jerse (1997).

Pitch at any time was calculated as the sum of four factors, all expressed as fractional MIDI pitch values:

Note pitch was determined by mapping vertical coordinates onto a scale. For the sake of variety the scale changed every 60 seconds. Range was determined separately for the two fishes, one being mapped onto a male voice range, the other onto a female voice range. The extreme high and low registers were left out as the formant synthesis tended to sound strained and unnatural in those regions.

Portamento lasting for about 300 ms was implemented between note shifts within legato phases. The portamento was implemented as a hyperbolic tangent function in the pitch domain.

Vibrato was implemented using a gaussian curve lasting for the duration of the note, resulting in an expressive “swell”. Vibrato onset was delayed until the transition from the previous note (formant adjustment, volume adjustment and portamento) was completed.

Slow 1/f fluctuation was used to imitate pitch drift as recommended by John Chowning (Dodge and Jerse 1997).

The sum of these factors were translated to frequency at sampling rate, and used to control an oscillator. The oscillator had a harmonic content where upper partials related to the fundamental frequency with a -6 dB roll off per octave, imitating the audio signal created by the glottal chords. Next the signal was filtered by five parallel resonant bandpass filters to implement vowel formants. For this we used the Max object *resonators~* developed at CNMAT (Jehan, Freed, and Dudas 1999). Formant properties were controlled using values from research at Ircam for the Chant program¹² (Rodet, Potard, and Barrière 1985), as listed in e.g. Boulanger (2000). For the male voice we used the bass and tenor values, and interpolated between them depending on the actual pitch. For the female voice we used the alto and soprano values. Interpolation was

¹²We did not have access to the Chant program itself at time, and therefore had to make our own implementation of song synthesis. Anyway I wanted to construct the synthesis algorithm myself out of curiosity and a desire to learn and experiment. Later on I have been experimenting with Chant as implemented in MaxMSP. Chant is based on particle or granular formant wave-function synthesis (fonction s'onde formantique or FOF) (Roads 2001), and my impression from limited tests is that the granular nature makes the singing voice sound coarser than the algorithm implemented for Ekkofisk.

done in such a manner that the bass or alto formant values dominated at low pitches and the tenor and soprano values at high pitches. This created a more convincing voice than if static formant values were used.

Formants changed from one note to the next. Tests indicated that if we used the same formant all the time, the impression of a vowel being used would gradually diminish and the sound turn more synthetic. Vowels therefore were decided randomly with a very basic restriction: We wanted to avoid the last two vowels used. This would ensure a change compared to the previous note, and it also avoided situations where the song would shift several times back and forth between two vowels only. With a total of five vowels there would always be three to choose from. Within the relatively limited range of possibilities we managed to get sufficient variation to ensure that vowels were perceived, at the same time avoiding predictable repeating cycles of vowels.

The CNMAT objects did not enable audio rate interpolation, so we had to do the interpolation in a series of discrete steps every 10th ms. This introduced undesired discontinuities that could sometimes be heard, but on the other hand it also gave the interesting impression of being consonants of a special language spoken by fishes only.

The overall amplitude of each note depended on the distance between the fishes as determined at the onset of a new note. A transition from one note to the next was implemented using a hyperbolic tangent curve in the logarithmic (dB) level domain, and then mapped to linear amplitude domain. In addition a possible swell was implemented using a gaussian curve in a similar way to vibrato.

For the first and last notes of the phrases special care had to be taken to ensure that the phrase started and ended in a convincing way. The end of the phrases we were able to handle convincingly, but envelope tracking analysis of real singers vocalizing would probably be required to find a more realistic envelope to use as model for phrase onset. Phrasing concerns required that the start and ending had to be treated as special cases, leading to four possible classes of notes. A Markov chain was used to control how these classes could be combined: A first note could only be followed by a middle or last note, a middle note could only be followed by a middle or last note, a last note could only be followed by a pause and a pause could only be followed by another pause or a first note.

The Markov transition table was continuously modified in real time in response to tracking data to impose two additional conditions: If the fishes were close together, they would sing, and if far apart they would instead play bell-like sounds. In addition phrases should not be “too long”. This was implemented using fuzzy logics (Elsea 1995), so that the likelihood of the next note being a last note increased the more middle notes had been sung in succession.

Finally the sound was spatialized using VBAP as discussed in section 3.3.3.

To summarize the sound synthesis processing emulated the result of the physical process of singing in two steps, first producing the signal of the vocal chords, and next filtering it according to the formants of the vocal tract. The result could be compared to that of song recorded in an anechoic chamber. In order to make the sound come alive, reverb had to be added. Some of the sites used for the installation had sufficient natural reverberance to serve the purpose, but for the Nordic Interact Conference 2001 in Copenhagen artificial reverb had to be added. The performance of the synthesis engine was in part controlled in real-time by the movements of the fishes, in part the result of a synthesis-by-rule approach (Sundberg, Askenfelt, and Frydén 1983).

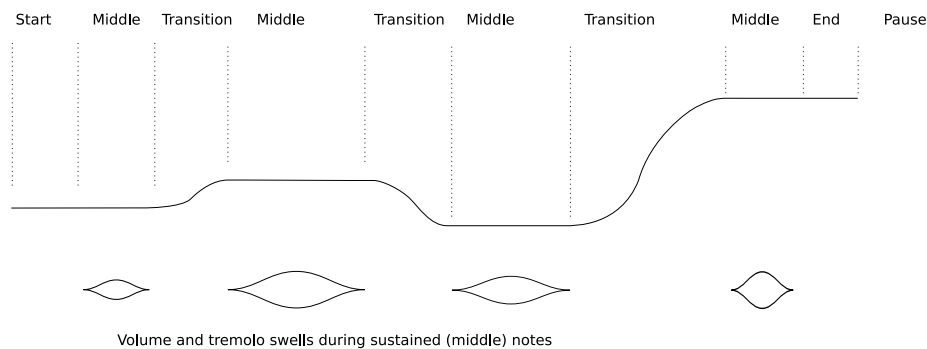


Figure 4.4: The structure of one melodic phrase is represented as a solid line. Vertical position indicates pitch height, and the horizontal represents progression in time, starting at the left. The phrase is built from a start note event, a number of middle note events with portamento transitions from one to the next and an end note event followed by a pause. Sustained notes are made expressive by means of vibrato and volume swells. If the melodic line were to be experienced as organic, decisions about the next musical evolution could only be taken at discrete points in time; at the start and end of transitions from one held note to the next as indicated by the vertical dotted lines.

Handling Memory in Real-Time Processes

The structure used to construct melodic lines in Ekkofisk is illustrated in figure 4.4. While the initial idea was to have an instant mapping of parameters to sound parameters we soon realized that this would make it impossible to create meaningful melodic phrases. Decisions could only be made at discrete points in time, first and foremost at the end of a note, deciding where to go next in terms of pitch, volume and vowel, or alternatively to end the phase. The decision on duration of the next note could be postponed to after the completion of the transition. For the transitions to appear intentional they had to be constructed with a clear idea of where they were coming from and where they were heading. Any attempts at changing ones mind along the way would change the impression from melodic song to endless and aimless synthetic glissandi. Similarly it was imperative to know the duration of a note at the onset of it, in order to give swell of volume and vibrato the right proportions, filling the duration of the note, no less or more. Once the decision was made no new decisions could be made before the note had been performed. The alternative would create either abrupt changes or notes that would seem to get stuck, both breaking the organic illusion aimed for.

The generative process driving the singing engine can be considered an automated improvisation machine, running in real-time. The making of Ekkofisk illustrates that a musical development can not be constructed simply as a stack of amnesic moments. A memory of the past and a plan for the future is required, in spite of the urban myth claiming that gold fishes have no memory.

If this problem is translated from song to language, it becomes evident that as I write this paragraph I have a clear idea that at some point I will be writing *THIS* word, and then find a way of ending the paragraph so that it appears to be semantically meaningful. On a longer time scale a similar awareness is needed concerning how the

phrase relates to the rest of the section, the rest of the chapter and the overall structure of the text. This is the very same problem about how to deal with short and long term form that any improvising musician continuously negotiate during the performance.

Many user-system interactive art works that I have experienced have tended to emphasize the immediate response and the one-to-one correlation between audience input and system response. If the system simultaneously is claimed to represent a democratizing approach, presented as an “instrument” requiring no prior skills or training, I am generally skeptical, fearing that the work might have limited abilities for offering deeper rewarding commitment and engagement, or that the quality of the experience might depend on the degree of sensitivity of the participating audience in the interact.

Erotogod, an installation I did in 2001 in collaboration with Ståle Stenslie, Asbjørn Flø and Knut Mork Skagen seemed to suffer somewhat in this respect. At the opening at Henie Onstad kunstsenter I witnessed the audience lining up to try it out, entering a stage or alter for a seven minute ride wearing a body suit with sensors and sensoritory feedback controlling 3 surrounding video projections and 16 channel surround sound. To me one audience performance stood out from the rest, that of Jøran Rudi, director of NoTAM¹³. A seasoned musician and well experienced with technology, sensor systems, electronic music and multimedia he seemed able to instantly decode the system and how it responded to his input. He had an ability of listening and playing the system that created a different degree of presence, expression and depth to anyone else that I witnessed. If the interactive system requires this kind of performance to shine, I believe that spectators with a prior training in music, theater, dance or other kinds of performance are likely to have a heightened sensitivity to the situation that makes them predisposed to be better players in spite of facing an instrument that they have never seen before.

Returning to the gold fishes, in the process of developing an expressive algorithm for the song, we also faced the risk of a break down of perceived connection between the movement of the fishes and the resulting sound. Decisions concerning the song could only be made at discrete points, often several seconds apart. This was further complicated due to latency of the system introduced by time consuming video processing with limited frame rate and further latency due to network communication from the computer used for video video tracking to the computer used for sound synthesis. The parameter that turned out to be crucial for maintaining a connection was the one parameter that we were able to update continuously, namely panning of the sound according to the position of the fishes.

The treatment of melodic form in Ekkofisk have to be considered very basic. Any form on a longer time scale than 15-30 seconds were entirely left to the performance of the fishes¹⁴. Still it points at some fundamental problems worth looking into concerning the dealing with time and memory in real-time processing and interaction.

Expressive Limitations of MIDI

Recordings using the Theremin, one of the earliest electronic instruments, still stand apart for their remarkable expressive qualities. John Cage might have criticized

¹³<http://www.notam02.no>

¹⁴The performance and mood of the installation in fact varied markedly over the four installments done so far. To me the stronger performances were the ones at Bergen kunstforening in 2000 and World Wide Video Festival, Amsterdam in 2001. The fishes used at Bergen kunstforening did not seem comfortable with the situation, and one of them had a habit of swimming upside down at the top of the aquarium for days. The installation sounded mellow and mournful. In contrast the fishes used in Amsterdam were very active, producing music of a much lighter and more cheerful mood.

the choice of repertoire used by performers of its day to expose the instrument (Cage 1961), but there are few commercial synthesizers available today that are able to match this instrument in terms of expressive capacity. When the first commercially available synthesizers started appearing in the late 1960s, the organ keyboard was chosen as model for the musical interface. This, combined with the introduction of voltage control, provided precise and consistent control of the instrument (Holmes 2002), but in the process of doing so the decision was made to mimic one of the simpler musical interfaces available. Technically speaking the organ keyboard is a layout of on/off press buttons. Disregarding the ability to control the registers of the organ for a moment, the only tone control available to the performer is deciding when notes are to start and end¹⁵. This is in stark contrast to most melodic instruments such as bowed instruments, woodwind and the singing voice where the tone can be moulded continuously by physical interaction with the instrument. Synthesizers generally improve the situation slightly by introducing key velocity sensitivity at note onset. The MIDI specification also offers polyphonic key pressure messages for individual control of each MIDI note after onset (MMA 1996), but very few keyboard controllers implement this. This critique of the MIDI specification was raised early on, e.g. Loy (1985) quotes William Buxton:

A major limitation of synthesizers to date, especially recently, is that they constrain the performer to expressing ideas through a limited set of gestures. (Ironically, some electronic instruments from the 1930s to the 1960s were more flexible in this regard.) This straightjacket of most over-the-counter systems (for example the piano-type keyboard synthesizer), has meant that in many cases the medium of expression is totally at odds with the musical idea. To follow on this, then, if gesture and idea are tied, and the device is the instrument for capturing the gesture, then the range of input devices could be as diverse as the range of musical ideas.

When aiming to capture expressive melodic lines in real time using a MIDI keyboard, similar concerns arise to the discussion on melodic construction in *Ekkofisk*. The MIDI protocol is unable to communicate in real-time the information that is required to produce sustained notes with a unfolding of swell and vibrato depending on duration, as the duration of the note is not known until the key is released, and by then it will be too late.

Keyboard-controlled synthesizers instead tend to implement envelope (and vibrato envelope) as attack, decay, sustain and release (ADSR).¹⁶ The evolution of the note event is predetermined as part of the programming of the sound, the only degree of freedom left being the duration of the sustained part. This is often instantly felt when playing a synthesizer voice for the first time: The instrument itself suggests what the duration of the note events should be, and when the key should be released. Attempts at violating this feeling often leads to notes sounding as if as cut off too early or held too long with a dead and static section before the release. Without having done any systematic listening tests to confirm the theory I am fairly convinced that if a survey of pop music from e.g. the 1980s utilizing synthesizers was carried out, one would find that for most synth sounds used within each track the range of note durations would

¹⁵Blessner and Salter (2006) discuss how reverberant rooms allows the organist some added control over the notes, as short notes will not last long enough to fully excite the reverberation of the room, and thus sound quieter. Synthesizers are generally portable, so this does not necessarily apply to them.

¹⁶ADSR was first implemented by Robert Moog in response to a suggestion by Vladimir Ussachevsky of Columbia University (Pinch and Trocco 2004).

tend to be very limited, in striking contrast to a similar survey of note durations for individual instruments in a orchestra score from the classical era.

An alternative approach is to use the modulation wheel common on most MIDI keyboards for real-time continuous expressive control. This overcomes the limitations of the note-on message discussed, but introduce other limitations. In analogue instruments the energy required to set air in vibration so that sound can appear has to be provided by the musician in the act of playing. The physical gestures involved are not only way of utilizing the body to express musical ideas, but the amount of energy, pressure, movement etc. is often directly proportional to the desired intensity of the music. The physical act of playing integrates with the aesthetic act of shaping music. Musicians that have been playing their instrument for years experience that the boundary between body and instrument gradually blur, and the physical action fuse with the expressive engagement.

In electronic music the energy is pulled from a plug in the wall. The production of sound is much less depending on physical labour, and instead a major deciding factor when constructing interfaces is the consideration of what is technically and financially feasible. Added to this is the notion that everything is possible, often with the inevitable corollary that everything should be done all of the time. One of the consequences is that interfaces have a tendency of requiring the musician to perform through a series of parallel and disintegrated physical actions. The right hand controls pitch while the left hand controls the modulation wheel. Even if the playing of stringed instruments could be considered to resemble this somewhat, I still feel that the playing of those instruments is a much more integrated act. The awkwardness of the MIDI keyboard used this way is further illustrated by the fact that it would be the role of the right hand to select the pitch, while the left hand controls the expression. When playing stringed instruments one generally says that the left hand is the craftman and the right hand the artist, as the sensory motoric abilities of the right hand are considered more fine-tuned. Maybe all keyboard players are left-handed.

Even if the modulation wheel is truly mastered, there are indications that the expressive capacity will remain limited. A recent software synthesizer, Synful Orchestra, emulates the melodic instruments of the symphonic orchestra using a novel technique named "Reconstructive Phrase Modeling" (RPM). RPM is based on additive synthesis with a noise component. The interesting aspect of the method is that control data for the synthesis is retrieved from a RPM phrase database, a database containing analyzed data of "musical phrases for each instruments. These are not recordings of isolated notes but complete musical passages that represent all kinds of articulation and phrasing (...), the many ways each instrument can move from one note to the next. (...) When Synful Orchestra receives MIDI input (...), it looks at the pitch, velocity, amount of separation or overlap between notes, note duration, volume, pitch, and mod wheels to determine what kind of phrase is being played. Synful Orchestra searches the RPM Phrase Database for fragments that can be spliced together to form this phrase. These fragments represent transitions between notes, slurs with portamento, lightly tongued transitions, aggressive fast bowing, rapid runs, long sustain regions with graceful vibrato, and noise elements like flute chiffs and bow scratches. Synful Orchestra stretches and shifts these fragments in time and pitch so that they combine to form the final output" (Lindemann 2006).

Synful Orchestra can operate in two modes, with or without a one second delay for expression. When played real-time the delay for expression has to be disabled to avoid serious latency, but when used for playback of a recorded track in a sequencer the delay for expression "allows Synful Orchestra to examine the incoming MIDI data and

make more sophisticated and expressive synthesis choices”. This implies that although Synful Orchestra use both MIDI note-on and expression messages, this information is insufficient to shape the expressive qualities of transitions as the transitions depend on an awareness of past, present and future.

To conclude the discussion I believe that there is a potential risk in real-time processing of underestimating and not taking into account the importance of progression and causality in time. A mapping system providing instant mapping of control parameters from sensor interfaces or similar is in itself unable to negotiate how every instant is to position itself in a longer time-span. Either this will have to be dealt with by a performer, as is the case with the Theremin, the technology has to presuppose a limited expressive regime, as is the case for the synthesizer ADSR envelope, or it might run the risk of becoming amnesic.

4.2.3 Installations with Sound of Their Own Making

LMW has been a framework for a series of installations of painting, video and sound in collaboration between Jon Arne Mogstad, Jeremy Welsh and myself. Jon Arne Mogstad is a painter at a time of video and photography (Jaukkuri 2004) while Jeremy Welsh investigates video within an expanded field of painting and digital media (Welsh 2005). For myself sound in relation to moving and still image, exploration of sound in a gallery and installation context and the shift from performed to exhibited sound have been key subjects in this collaboration. For all of us LMW has functioned as an experimental playground for investigating how the different media and their traditions might differ, relate and fuse. So far we have done three gallery exhibitions, an installation out- and inside a container at the Quart music festival, a workshop/work in progress presentation at the Bergen National Academy of the Arts and three smaller presentations as part of larger events. In addition Jeremy Welsh and Jon Arne Mogstad did an exhibition at Visningsrommet USF in November 2006, with a modest contribution of sound from myself, as it opened a few weeks after the opening of my installation *Cubic Second* at Hordaland kunstsenter. As part of the exhibition at Visningsrommet USF documentation from earlier LMW projects were presented.

Dialogue - Transition

When I started preparing material for the first LMW exhibition, *Dialogue - Transition* at Christiansands kunstforening in 2004, I wanted to continue work on algorithmic seemingly performed melodic sound from the Ekkofisk installation, at the same time substituting the singing voice for more abstract and synthetic voicing. I made good progress developing algorithmic principles for generating melodies, getting musically convincing, expressive and at times intense and almost aggressive results, but the “*Whys*” started haunting me. The more musical the sound material got, the less it seemed to belong to the project. A blog post dated February 29 2004 provides a snapshot of thoughts and unresolved questions at the peak of a minor crisis less than a week before the opening¹⁷:

Making electronic music for installations is radically different from writing music to be performed by musicians. When sheet music is interpreted by good musicians, you get a human presence that is at times missing in

¹⁷<http://www.bek.no/Members/lossius/lostblog/88>

electronic music. Kraftwerk and the Detroit techno reacts to this by exploring the machine-like world where human presence is removed altogether. “I want to be a machine” as Ultravox put it.

To me, one of the great achievements of Eno on *Another Green World* is that he creates a machine-like music that is longing for a soul. Some of the tracks implies a very complex and ambiguous emotional landscape.

Personally, I would like to keep the complexity and warmth of performed music. I have been banging my head into a lot of walls while working on sounds for the installation in Kristiansand. I’m able to get very interesting sounds generated by the video material of Jeremy, but end up with a lot of whys. It either ends up sounding inhuman or I start adding expressive qualities that are very hard to justify. The question I have kept asking myself is: *Why should sounds still have a human presence and expressiveness when there’s no one performing it anymore?* I know what I’m aesthetically attracted to and want to do, but I need to be able to answer some very fundamental questions in order to do it in an artistically convincing way.

Who do you meet when you’re listening to music? If you experience a human presence, is that the presence of the composer, the musician or yourself? If you see an exhibition of paintings at a gallery space, the experience is independent of the artist being at the gallery. E.g. paintings of van Gogh and Pollock bear a strong testimony of their physical movements. Is it possible for me to capture in an analogue way my own gestures and use that as part of the algorithms creating the soundscape?

Is interaction nothing but a cheap workaround for this problem, leaving responsibility of presence to the audience? I have often been frustrated and unhappy about interactivity. I feel that I create an instrument, but the one playing still have to do so in a musical way. That requires awareness and training; not necessarily practice at using this instrument, but a mental awareness and sensitivity. Most of the audience does not have that, and are unable to experience the true potential of the instrument/art work they are facing.

In written music the composer indicates gestures, but it’s the responsibility of the musicians to actually create them. During a presentation at BEK last year Francisco Lopez stated that music is currently going through a shift of paradigm, getting rid of the performer as mediator between the composer and the audience. I believe this is a valid and relevant way of reading current changes in music practice. Still, for me as a composer, “reclaiming” the right to the physical expressiveness feels like a major step. Awareness of past music practice makes it impossible to do this without being acutely aware of what is left out and behind.

A few weeks ago, Jeremy sent me photos of paintings by Jon Arne Mogstad. Spray paint is an important part of his techniques, and to me there is a lot of musicality in his images. I will try to track his physical gestures, and map them to musical parameters. (...)

Figur 4.5 is a sample of the paintings I was studying at the time.

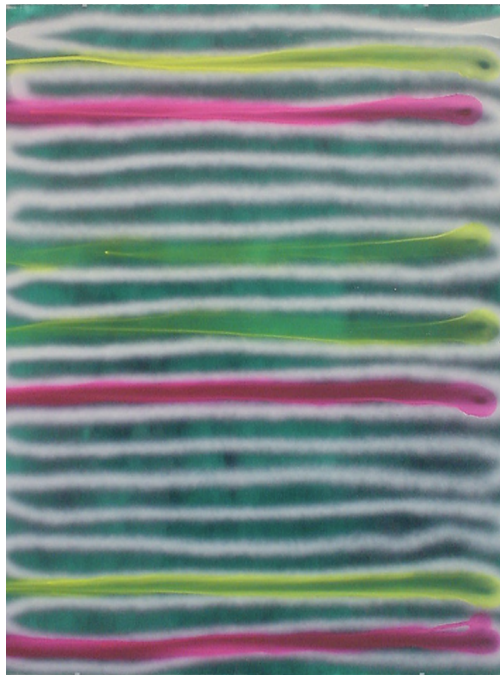


Figure 4.5: Hypnotic #1. Painting by Jon Arne Mogstad.

Video as Sound Source

In the end all material aiming at melodic constructions and performative musical qualities were abandoned in a frustrating process of elimination; starting out with very little, and removing more and more, in particular anything aiming for bolder musical gestures. The more I removed, the more I saw that what was left began to flourish. I was commenting on this in a presentation at a *Sound in Focus* seminar at the Bergen National Academy of the Arts a few weeks later. David Toop remarked that a similar shift could be seen in a lot of electronic and electro-acoustic music during the 70s, a shift from modernist expressionism towards the minimalistic, as if composers were coming to terms with the inherent aesthetic language of the new medium of electronic and computer-based music and sound instead of insisting on a compositional approach derived from analogue composed music. Personally I believe that if electronic and computer-based music and sound currently has an inherent tendency towards minimalism, it has a lot to do with limited interfaces and limited capacity for advanced control of the new technology. While minimalism is artistically interesting, I believe that the expressive limitations of technology should be challenged, as they might prevent other artistic possibilities. In the case of the *Dialogue - Transition* installation, I believe the real issues causing difficulties with the original approach were to be found elsewhere, but I only recognized them a few weeks after the opening.

In the remaining process up until the opening I reexamined gestural, expressive and time-related qualities in the work of Jeremy Welsh and Jon Arne Mogstad in order to position my own contributions relative to that, either moving in the same direction as their material or introducing opposing forces. Core to this process was an investigation

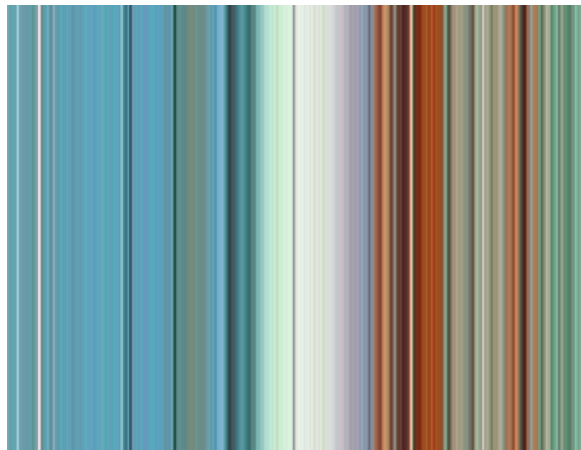
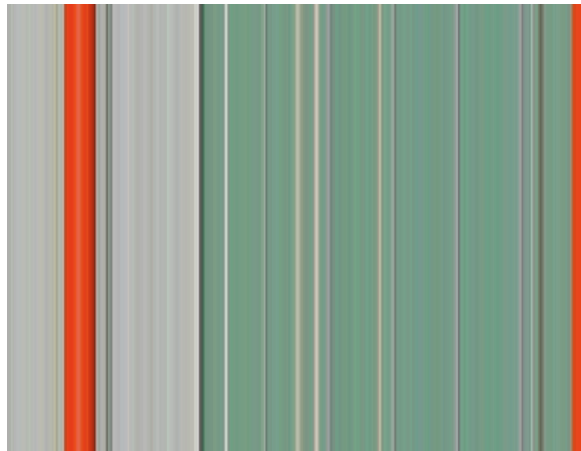


Figure 4.6: Vertical stripes. Snapshots from video material by Jeremy Welsh.

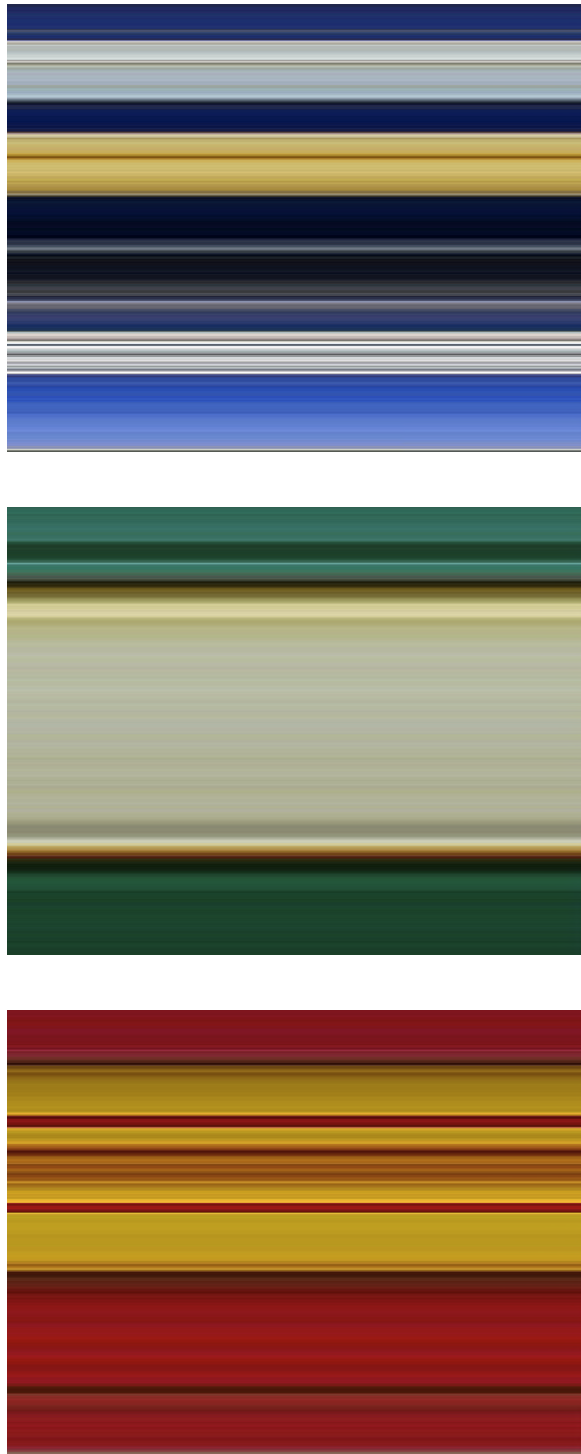


Figure 4.7: Horizontal stripes. Snapshots from video material by Jeremy Welsh.

of the qualities of their material and how and if it could be mapped and reused for the sound processing. Three sound sources ended up being used for the installation:

Misreading video files as audio The raw sound I ended up using for the installation was based on misreading video files as audio. Several of the videos had strong rhythmic and repeating patterns, in particular the video sections made up from vertical or horizontal stripes (figures 4.6 and 4.7). I expected the data files for these videos to show repeating patterns, as the same pixel color combinations occurred again and again for each row of the frames.¹⁸ Trying different possibilities I ended up reading them as mylaw-compressed audio files sampled at $n \cdot 441$ Hz, using varying values for n : $1/8$, $1/4$, $1/2$, 1 and 2 . 441 Hz is $1/100$ of the sample rate used for CD quality audio playback. This resulted in a semi-pitched, semi-pink-noise-like sound with amplitude disturbances that I found interesting. The different values of n came across as octave transpositions.

Audio generated from live video feed Another source of audio was created by mimicking one of the principles for video processing used by Jeremy Welsh: From the images a single row or columns of pixels was extracted and stretched to fill the full height or width of the image, so that color would be constant along each of the pixels columns or rows of the image (Jaschko 2004). In the processed videos that Jeremy was working on images were shifted at a fast rate, approximately every second frame of the video (12.5 times a second).

A video camera was set up to point towards the audience and from the image one row of pixels was extracted. Motion tracking was performed on this row using Max and Jitter, and the resulting image presented on a data screen. Colors were boosted, resulting in an abstract live interactive image displayed at a computer screen. When no-one was present or moving, the image would be black. As the audience approached the camera, a small bright vertical stripe would appear, growing larger until it filled the whole of the screen when the audience were in front of it.

At the same time the resulting image was scanned in the horizontal direction and used as a lookup table for audio wavetable scanning synthesis. This was done at very low frequencies, typically 1 to 4 Hz, resulting in a crisp discontinuous and quasi-rhythmic noise, to be further processed. This audio source would only come alive as someone approached the section of the gallery space that the video camera was capturing.

The interactive video and audio processes were not meant to be immediately recognized as a major and obvious interactive response of the room, but rather giving the audience a more subconscious impression of the room somehow responding to their presence with a slightly intensified sound image.

Resonance modeling A recording of viol consort music by the English early Baroque composer John Jenkins was used as input exciting a physical model of a resonating pipe using the program Modalys-ER (Polfreman 1999). The result was

¹⁸In hindsight I do not think that the visual content of the videos had much to say. Most video compressions use some form of spectral Fourier transformation of the data set, so any rhythmic patterns would be lost in the process. Instead I believe that the sound quality mainly was an artifact of linear interpolation over several samples between sample values that from an audio perspective seemed more or less random. One of the externals (`lp.frrr`) from the Litter Power Pro package of stochastic objects for Max/MSP are able to produce fairly similar results (Castine 2002).

a pitched and tonal but still abstract continuous sound, suspended in a perpetuated state, resembling an extended harmonic cadenza that is never resolved. The inclusion of this sound source might seem an odd choice, but the sound file was part of material resulting from the *Living room* workshop in Trondheim in 2003 (further discussed in section 3.3.6). *Living room* was the first collaboration between Jeremy Welsh and myself. During that workshop he started exploring the video processing methods that would be used for *Dialogue - Transition*. On a more personal level this sound file had already become part of the history and references of the project. Of course the material was also felt to work well with the rest of the sound material, providing the right balance between blending in and being a contrasting textural layer.

Sound sources were further processed using a bank of parallel band-pass filters, with bandpass frequencies set so that they formed harmonic ratios respectively to each other and to the sample-rates used for playback of the video. This processing offered a wide range of colorings of the timbres of the incoming sounds, varying from abstract and unpitched noise to fairly clearly pitched sounds.

In addition a multichannel granulated delay algorithm was used, based on the “Ben-nies” library for Max/MSP by Benjamin Thigpen, available as part of the IRCAM Forum real-time composition library for Max/MSP¹⁹. This is a fairly simple real-time granulation algorithm with feedback and offers a wide range of effects; all the way from time stretches or single or multiple delays with or without feedback to synthetic delay or reverb effects with time distortions where the incoming sound is pulverized into particles as it vanishes.

Control Data Generated from Image Analysis

Jon Arne Mogstad uses spray paint extensively. One of his paintings (figure 4.8) dominated by white spray paint against a blue background was analyzed using Max and Jitter and used to control amplitudes of the various sound layers. This was done by reducing the image to black and white according to brightness, and interpreting gray values as dB sound volume. White was maximum volume, and black silent. The image was then scanned along one vertical line (one pixel column) in a similar way to how Jeremy had been generating the stripe images in his videos. This scanning was performed in time, so that a dynamic amplitude curve would result. The horizontal lines of spray paint would result in breath-like amplitude swells, one *crescendo* and *decrescendo* for each line. For different sound layers, the image was scanned at various speeds, ranging from 1 sec. to 6 minutes. This created anything from a shaky tremolo to slow breath-like sweeps as the scan algorithm moved from one stripe to the next. Reading the image like this was a means to create a repeating pattern that due to the nature of the sprayed stripes would be rhythmic but still contain variety and a certain level of organic unpredictability as compared to using a simple cyclic mathematical function.

When all of the line was scanned from bottom to top, a new scan would begin, at a random pixel column in the middle region of the image. A Welch window function was superimposed on the scan. Window functions are used for a number of DSP methods, most notably granulation, Fourier transforms and wavelets. Most window functions share the general idea of suppressing values at the tails of the region, so that only the middle region is audible and emphasizing²⁰. The window function served to make

¹⁹<http://forum.ircam.fr>

²⁰For a discussion on the topic, refer to (Roads 1996) pp. 1099-1104.

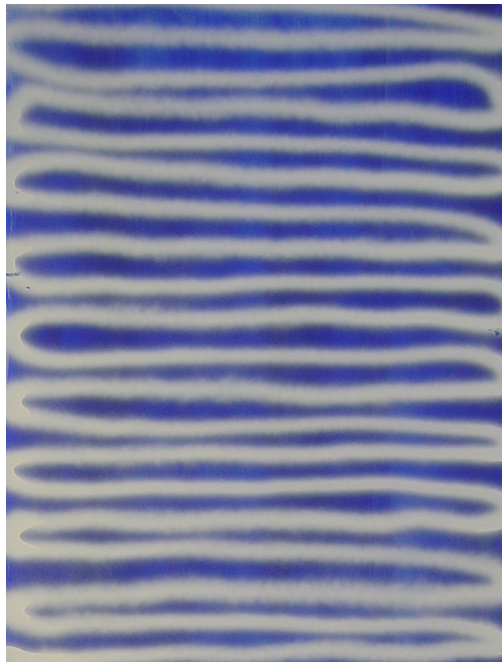


Figure 4.8: Hypnotic #2. Painting by Jon Arne Møgstad.

the different layers of the total soundscape come and go instead of being ever-present, depending on whether amplitude was read from the middle part of the image or close to the edges. This caused a less dense soundscape with larger amounts of drift and variation in the total texture as all of the layers would not be present all of the time.

An Extension of Painting Space

Christiansands kunstforening is long and narrow gallery, approx. 65 x 10 meters. A total of 12 loudspeakers were used, spread in the space, in addition to the speakers of a wide screen TV used for displaying videos. The sound for the installation consisted of multiple layers, all working together to form a coherent soundscape. The positioning of speakers and sound drew inspiration from the Venezian Renaissance school of concertante music, exploring the room by using groups of speakers as choirs positioned at different positions. Speakers were positioned at various heights, some pointing towards the roof or walls so that the direct signal immediately was interacting with the acoustics of the space. The overall acoustics of the space was much less reverberant than initially feared. This was probably due to the rectangular space having large total surface of walls, floor and roof compared to volume, thus reducing reverberation time according to Sabine's equation (section 3.2.2). In addition the floor was covered with felt carpet dampening high frequencies, and the large glass surfaces would provide further damping of deeper frequencies. While all sound was heard in all of the space, one still experienced the soundscape as sufficiently dry to be perceived as concise, and one had a clear notion of sound coming from sources nearby or at a distance.

Each of the elements of the sound would be distributed to a number of loudspeakers. The granulated delays create textures with spacial richness and movement. Different



Figure 4.9: Installation views *Dialogue - Transition*.



Figure 4.10: Installation views *Dialogue - Transition*.



Figure 4.11: Installation views *Dialogue - Transition*.

textures were distributed through different subgroups of loudspeakers, and depending on the movement of the spectators in the room, different layers would be prominent or in the background. Still, as all elements could be heard in all of the room, sound ensured a continuous auditive awareness of all of the space, functioning as a perceptual glue transforming the experience from an exhibition of a number of individual works to become an installation, in a similar sense to how Allan Kaprow considered environments an extension of painting space (Reiss 2000). The intensity of each layer of sound would vary in time, and occasionally one would experience the other end to be very active and attention-seeking. The interactive sound and video, positioned at the far end of the gallery space relative to the entrance made that end of the room more active when there were spectators in the vicinity. Spectators in other parts of the room would recognize the presence of audience at the far end due to the heightened intensity of sound they would induce. When looking across the room video would function in much the same way, providing movement and shifts of intensity that would contribute to an overall awareness of the size of the space and the total content of painting, video and sound in the installation.

Looking back I believe there are at least two reasons why the initial attempts at creating sound with performed and musical qualities turned out to be difficult. The musical performance suggests the performer or musician, and loudspeakers are used instead, the situation might introduce a sensation of displacement, as the music seems to originating at an earlier time and possibly somewhere else. The video of Jeremy Welsh and the paintings of Jon Arne Mogstad both escape this kind of trapping in time. The videos containing a flickering flow of inter-independent images at a high frame rate with no immediate narratives, causality or development hint at a database of visual information scanned in a more or less random sequence, an overload of information that after a while starts suggesting a different viewing mode, ignoring the frantic surface, instead probing for common denominators through some sort of visual or perceptual low-pass filtering, not unlike the experience of listening to minimalist music by e.g. Steve Reich or early Philip Glass.

The expressive action of painting is clearly visible as broad physical gestures in the paintings by Jon Arne Mogstad. But these are frozen motions, time collapsed or packed onto a split second, freed from the axis of time, coherent statements caught by the spectator at a glimpse instead of depending on a time span to unfold and reveal their information. Similar qualities were required for the sound in order to work in this installation context, and that was probably part of the reason why the sound material paved itself a very different direction from what I had initially planned.

A few weeks after the opening at Christiansands kunstforening, Maia Urstad exhibited the installation *Lydmur*²¹ in Bergen. This was a physical wall built from older portable radios, cassette and CD players, sized 8x2 meters, (figure 4.12) with a 18 minutes sound composition triggered as the audience enters the space, reproduced over many of the CD players (Sekkingstad 2004). Experiencing the installation for the first time I got the impression that at times an enlarged speech-resembling sound was sweeping through the wall, as if the it was possessed by a haunted ghost.

The *Dialogue - Transition* installation had no such object or body that sound could have been bound to, and hence sound with strong bodily gestures ended up be foreign to the context. Considering the shape and size of the gallery space and the visual material to be used, what was required for it to come together was not the introduction of an object or body in the space, but rather sound to reinforce the gallery as a an

²¹The title could translate as "Sound Barrier" or "Wall of Sound".



Figure 4.12: *Lydmur #3* av Maia Urstad. Electrohype Biennale 2004 - 2005 at Malmö Konsthall. Photo by Lars Gustav Midtboe. Used by permission.

environment, a site to be enacted by the audience.

A Can of Stellar Stripes and Other Vessels

In a lecture at the d!sturbances symposium in Copenhagen in 2003 Heiner Goebbels said that he often introduced opposing or contrasting elements with no attempt at bridging or morphing the material, instead leaving a tension or gap that was left to be resolved by the audience. This could be said to be the main *modus operandi* of the *Dialogue - Transit* installation, as the title of it itself reveals. As the LMW collaboration has continued, the separation of the different media has been increasingly blurred, at its best merging to become one multi- or metamedium.

Although I am still inclined to do further investigations of algorithmic processes leading to performative and music-like expressive qualities, that stopped being a concern within the framework of the LMW collaborations. Instead I began using the performer at hand, doing audio recording of Jon Arne in the act of painting, preferably spray-painting as that resulted in recordings with ample possibilities for further processing.

Preparing for the *A Can of Stellar Stripes and Other Vessels* installation or exhibition at Trondhjems kunstforening fall 2004 I recorded Jon Arne and Jeremy applying spray paint at top of big sheets of paper with photo emulsion. These later became part of a collage in one of the rooms of the gallery (figure 4.13).

The resulting sound was processed using filters and granulated delays with feedback in the same way that material for the previous installation, resulting in sound on seven channels. The delay was set up so that the sound feeding into the granulation system would be gradually more delayed for each of the seven channels. The seven channels were distributed to four loudspeakers positioned sideways in the windows of the room as illustrated in figure 4.14, so that channel one was routed to the first speaker, channel two panned midway between speaker one and two, channel three routed to the second speaker, etc. Filter parameters as well as playback speed of the sound recording was occasionally altered so provide sufficient variation in the sound that one would



Figure 4.13: Installation view from installation in Trondheim 2004.



Figure 4.14: Installation view from installation in Trondheim 2004. Speakers in windows intersected by paintings.

not grow weary of it. Sound playback was changed between ratios 1:1, 1:2 and 2:1. In the same way as for the previous installation this would come through as octave transpositions.

I have found when working with sound bordering between noise and pitch that the introduction of transpositions and pitch shifts tend to radically alter the way sound is perceived. As soon as the pitch of a sound is changed, perception of the sound seems to change from listening to the sonic qualities of the sound to focusing on relations between pitches. I have found that one way of varying pitch remain in a state of listening to sonorous qualities rather than pitch intervals, is by moving up and down in octaves. This kind of changes could be considered an extension of the idea of the drone, in spite of the change everything remains fundamentally the same, or stated slightly differently, everything maintain the same relationship to the fundamental frequency.

The resulting sound discussed here felt energetic and restless sound, and was played back at a higher volume than I tend to use for installations. The grains of the audio processing literally resembled the fragments of paint drops constituting a stripe of sprayed painting, and continually created motions starting at the entrance of the room and running along the wall, being dissolved as they approached the montage at the far wall.

What Goes Around, Comes Around

In the summer of 2006 LMW contributed to a group exhibition at Seljord kunstforening with the installation *What goes around, comes around*. By now the working method seemed to have changed for all three of us. Instead of combining material prepared in advance, more or less everything, painting, video and sound, was made while working together at the site within the time span of a hectic week. This was partly based on a workshop earlier the same year at Bergen National Academy of the Arts, used to experimenting with new ways of working and integrating the various media (figure 4.15 and 4.16).

I started out by recording while Jon Arne Mogstad spray painted on the walls using a compressor (figure 4.17). In addition to recording him painting, a recording was made while he gradually let air out of the compressor before taking a break. From these two recordings the sound of the installation as constructed. Recordings were treated as input to physical models of resonating pipes similar to what I had been doing with the music by Jenkins earlier on, and further treated to become a vibrating and organic layer of metallic sound repositioning itself to new random positions in the space at random intervals. The two raw recordings were presented as is in alternation as separate layers of sound. The recording of air released from the compressor sounded like pink noise with amplitude and spectral fluctuations and a gradual large scale fade out lasting for almost six minutes. When played back, the source of this sound was hard to recognize, and it behaved like an abstract sound object. On the other hand the recording of Jon Arne painting was easily recognizable, with occasional speech, noises from working with cans of color, spray painting, stepping back and forth, etc. Levels were set so that it would surface occasionally in the mix, coming and going. Ambisonics were used to distribute the sources to the 5 speakers used, so that the sound would rotate slowly around the room, in accordance with the title of the installation.

Sometimes, when the sound of Jon Arne spray painting passed over a zone of large spray painted stripes (see the lower photo in figure 4.18) the earlier time when the installation was in the making fleetingly seemed to coexist with the “now” of the finished work experienced by the audience. It did not necessarily represent a conceptual riddle to the same extent that (LaBelle 2006) read *Box with sound of its own making* by

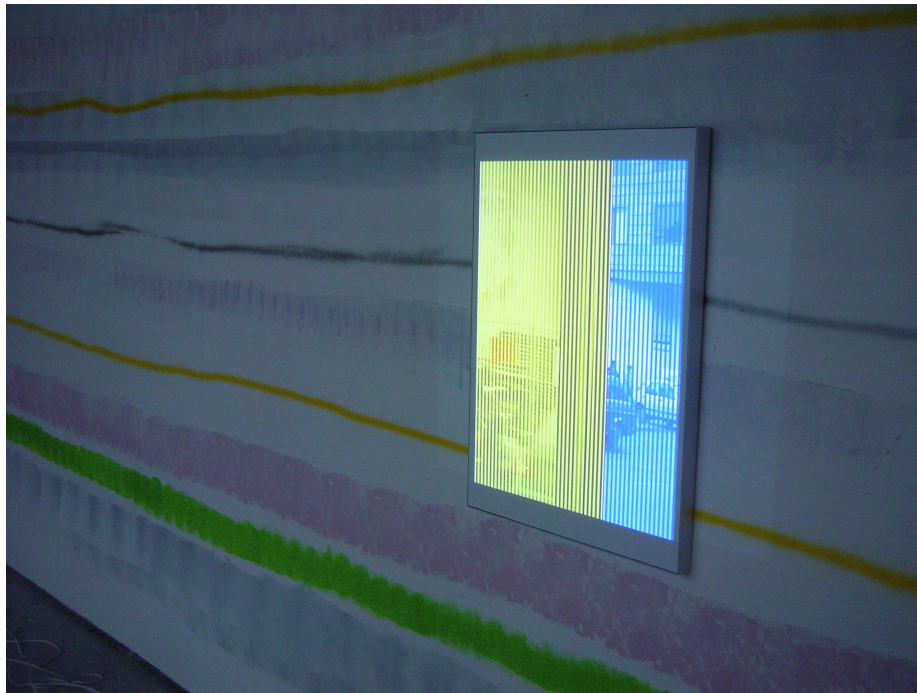


Figure 4.15: Video projection on top of flat panel loudspeaker on top of wall painting. *What goes around, comes around* workshop and work in progress presentation at KHIB.



Figure 4.16: Installation view from *What goes around, comes around* workshop and work in progress presentation at KHIB.



Figure 4.17: Recording while Jon Arne is working.

Robert Morris, but it shared a duality of process and result. Painting starting folding back out in time, and as one heard Jon Arne discussing the result with himself, another dimension of openness entered. Not only did the spectator become integral to the completion of the work (Bishop 2005), as installations represent “a shift from art as object to art as process, from art as a ‘thing’ to be addressed, to art as something which occurs in the encounter between the onlooker and a set of stimuli.” (de Oliveira, Oxley, and Petry 1994). But the physical realization of the work itself opened up as process and ceased to be an absolute. It could have been conceived and realized slightly or radically differently, and as spectator one was free to start imagine these other possibilities.

Improsynth/jc87

As fits a touring band LMW travelled directly from Seljord to Kristiansand to do an installation at the Quart festival²². The installation was to happen in and outside a container situated near the entrance to the festival area. In some respects this was a revisiting of points of departure for the LMW collaboration as we were back in Kristiansand. In addition a lot of the video material by Jeremy Welsh was based on photos taken of containers and container terminals, representing floating and temporary architecture. Still the setting was in all respects very different from earlier projects. In this work sound and the situation that the audience found themselves in when entering the container seemed to take the lead roles.

Sound was based on recordings on the site, an impromptu performance hitting, knocking, touching, kicking the walls of the container. This was further processed, mainly using granulation and reverb, to create multiple layers of sound with varying degrees of smoothness or more tactile, rough and edgy surfaces. In the processing of the sound, a major concern was to maintain and enhance frequencies of the original recordings to ensure that once played back, the sound would resonate with the container, and make it vibrate.

Outside the logo LMW was painted in large letters down one side of the container, and rasterized patterns and stripes of color painted on the outer end. Inside the con-

²²<http://www.quartfestival.com>



Figure 4.18: Installation views *What goes around, comes around.*



Figure 4.19: *Improsynth/jc87*, Quart Festival 2006. Photo of internal taken while preparing the work, as photo documentation of the completed work was difficult. Unorthodox use of ambisonics was employed, not to recreate a realistic ambisonic sound field, rather to use the collective sum of speakers to push enough air to excite resonances in the metallic walls, roof and floor of the container.

tainer, eight loudspeakers were suspended along the side wall and a small lcd screen was attached to the end wall. Dark drapes were hung in the entrance so that the only light inside came from the seven inch lcd screen (figure 4.19).

Entering the sun backed container, it felt small, hot and uncomfortable, of uncertain depth, the floor vibrating under the feats as the sound set up resonances. The fairly loud volume sound, moving back and forth along the length of the container, enhanced the ambiguity as to how deep the container was.

4.2.4 A Shared Space

My wife recently picked up a Furby²³ for the kids to play with. This toy epitomizes everything that was wrong about interaction at the end of the 1990s, claiming to be a brave new emoto-tronic toy, “the ultimate combination of advanced robotic technology, puppetry, and realistic lifelike form-factors”. Instead of “an entire gamut of emotions fully realized in a plush toy” you get a manual to read, and a scheme of stimuli and responses that quickly makes you wonder if this is the latest trick of Pavlov, only this time you are the dog. The first day the kids played with it for 20 minutes, the next day 15, and since then it has been hibernating.

Shulgin and Manovich have both accused interactive installations for a tendency towards being manipulative and totalitarian (Manovich 1996). Fortunately this has often been safeguarded against by adding mediocrity. Personally I have seen few installations with spectator-system interaction that I have been enthusiastic about, but through collaborations with the composer Asbjørn Flø and the sound artist Jana Winderen I have gradually recognized other potentials for interaction that I find more appealing.

The installation *Lydspor* (Soundtracks) by Asbjørn Flø and myself investigated technological failures, processing and deconstructing the sound of digital glitches, crashing computers generating sound and cheap lo-fi speakers pressed to and beyond their limits. Sound was played back using a 40-speaker setup mounted in a steel grid floor with 24 push buttons in groups of 3 for interactive triggering of sounds. With all sounds turned to eleven once activated this was not the occasion for the subliminal. We had orchestrated the various layers of sound to differ according to parameters such as frequency, rhythms, smoothness, etc. to be audible when coexisting, but prior to the opening I was unsure about how the installation would appear and sound to the audience and how and if they would use it.

As the audience entered at the opening, first a few and then more and more, the installation responded with a crescendo from relatively quiet sleep mode to the frantic, staying there for the next one and a half hour. For the first time I understood the idea behind the moment form of Stockhausen where attention is on the “now”, on the eternity that exists in each present moment. There would be changes, new moments, organized by the spectators randomly moving from one group of buttons to another, but apart from the introductory *crescendo* as audience arrived and a similar *decrecendo* at the end of the event, the installation itself refused the production of any large scale form or narratives. In his discussion of *Kontakte* by Stockhausen (Dack 1999) suggest a narrativity of the “antinarrative”. Each moment seems self-contained with no connection to succeeding or preceding moments, but individual sections or moments still have a number of narrative curves. In *Lydspor* these curves seemed be caused by and between the actions of the audience.

As the evening drew to a close a few stray spectators entered. An elderly man with

²³<http://www.hasbro.com/furby>

a walking-stick entered the empty room, went across the floor to the middle, turned around and headed for the door again. He had apparently not seen the buttons on the floor, but on the way out he accidentally touched the last one with the cane. He literally turned on the stick and spent the next five minutes investigating all of the buttons using his feets and the stick.

A little after three middle-aged upper-class ladies entered. Not seeing me hidden in a corner of the room, they let go of their cultivated and reserved outward appearance, and climbed around the floor laughing and playing. By the time they left they seemed ready for a different night on town than originally planned.

The room seemed able to function as a catalyst for this kind of “found” performances. The poetic of the unfolding situations did not necessarily have much to do with sound at all in the end, but the room encouraged impromptu situations where spectators would reveal different sides of themselves in ways that to me were beautiful. A project originally conceived to address failures and shortcomings of technology ended up with a rather humane appearance.

Kunsten å gå på vannet (The art of walking on water) was an interactive sound installation produced in collaboration between Jana Winderen, Jørgen Træen and myself. Technically the installation consists of a number of floor mats with built in burglar alarms, connected via MIDI to a computer. MaxMSP is used for sound processing in real-time, played back using monitors positioned at the four corners defining the area of the installation. The mats function as a giant keyboard, inviting a different and more physical interfacing with the computer (figure 4.20). When stepped on, sounds are triggered, one for each mat. It is possible to shift between several groups of sound. Another more continuous and diffuse layer functions as a sonic backdrop. The work was a commission for Turnéorganisasjon for Hedmark²⁴ as part of “The Cultural Rucksack”²⁵, and over the last few years Jana Winderen has visited almost all primary and secondary schools in the Hedmark county, setting up the installation and inviting pupils to experience it in small groups of four or five at a time. The exchange between herself and the children has been an integral part of their experience. The installation has also been presented in other contexts outside Hedmark.

From the descriptions of her experience when touring, the interface encourages the children to move in different ways, anything from careful walking to the extravert. Just as important is the interaction between the movements and sound, and the interaction between the children in terms of how they respond to movements and sound. The installation becomes a social space, a vehicle for negotiations of relations between the children, between the children and the artist, and between the children and the collective sound they produce.

One issue kept nagging me as being partly unresolved in this project; the potential contradiction between encouraging the children to listen attentively, an act that usually requires a degree of quietude, and simultaneously requesting physical movements to generate the sounds to listen to. When invited to exhibit at Galleri F15 as part of the sound art exhibition *The Idea of North* in 2005 we²⁶ decided to start from scratch to solve this. The physical interface of mats were kept, but all sounds, the programming involved, the construction of interaction, and the conception of the room was defined anew. The floor was painted blue, and a bench introduced inviting both physical activity and attentional listening and contemplation. In the interactive design we aimed for a

²⁴<http://www.turneorganisasjonen.hm.no/>

²⁵<http://www.denkulturelleskolesekken.no>

²⁶Jana Winderen and myself.



Figure 4.20: Kunsten å gå på vannet. Photo by Jana Winderen.

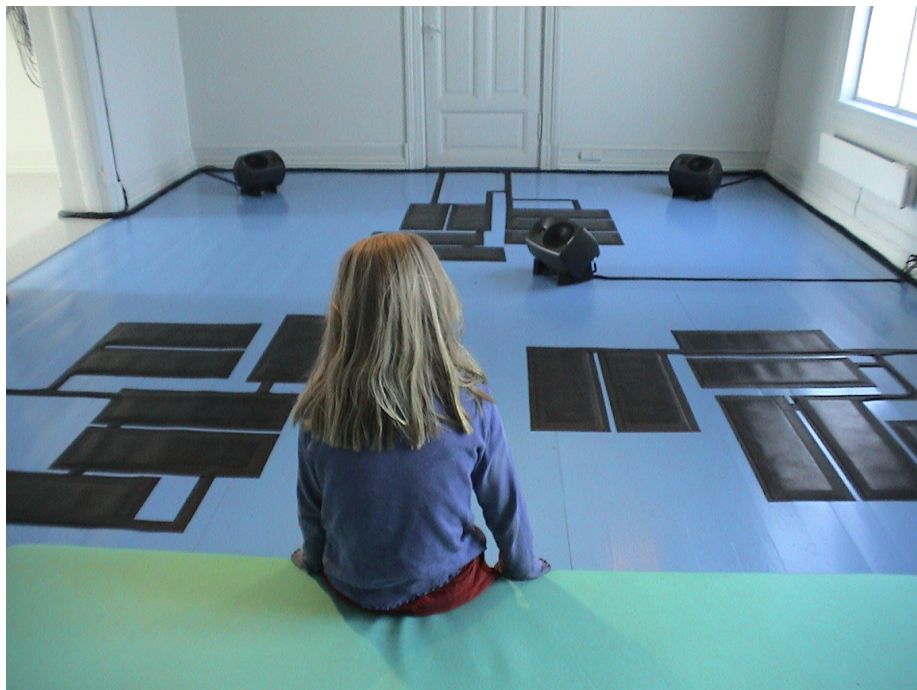


Figure 4.21: Quiet and Relaxed, but Alert. Photo by Jana Winderen.

sense of quietness, a balance between providing sufficient feedback for the interaction to be sensed, while restraining means to encourage a slower, calmer and less extrovert experience than the very physical active situations often occurring in *Kunsten å gå på vannet* (figure 4.21). The technical equipment used was mounted on the wall in the room next door.

The mats were organized in three groups, with a layout hinting at the black keys of the piano. This provided a structure for several simultaneous layers of audience-generated sound. Sound was very minimalistic, based on three field recordings done by Jana Winderen in a forest; a brook, footsteps in a puddle and footsteps in snow. Each recording was of short duration and continuously looping, but the loops were of different lengths in a similar way to the track *2/1* from *Ambient 1: Music for Airports* by Brian Eno, discussed in section 4.2.1, and as such the three loops would continuously combine in new ways. Sound processing for the interactivity was very simplistic: Each loop was associated with one of the groups of mats, and when a mat was activated the sound was delayed, band pass filtered and boosted to remain audible before being added to the dry signal of the loop. The impression was one of pitched sound, a different pitch for each mat in the group. The textual surface of the processed sound was different from one group to the other. The brook was the most instrument-like, a continuous whistle-like sound, the steps in snow having a rougher and crunchy surface while the processed sound of steps in a puddle become slower almost rhythmic pulses. The three groups also differed in register of the band pass filters, an orchestration to maintain transparency when several mats of several layers were combined.

In *Quiet and relaxed, but alert* spectator-system interaction became subordinate to the spectators interacting with each other. The installation became a space for shared experiences as one would play the mats while others listened and responded, two played duets using one group of mats each, or someone sat down to watch others in activity.

Bourriaud (1998) claims that art has the ability of tightening spaces of relations. Unlike TV, literature, theatre, cinema and presumably also live music performances there is a possibility of an immediate discussion. “Art is the place that produces a specific sociability.” To a large degree I believe this to be culturally conditioned. The concert audience has not always refrained from coughing, commenting, cheering or booing in the middle of the performance. Also, reading the statement by Bourriaud the devil’s advocate ponders if the embrace of sound art and the immersive might eventually reduce this social ability of the gallery. When one have to keep the voice down at the vernissage to prevent that the sound installation drown in idle chatting, the gallery might become another arena for the artists monologue.

When I first started working on installations, I envisaged works experienced in solitude. At the time I was listening to music by Salif Keita, and the lyrics for one of the songs at the *Folon* album, *Nyanyama*, was provocative to me:

We artist, we play the Kora, the Balofon
And even the herdsman’s Krinan

Making music is easy for us
And a noble profession
We are the needles which sew the social fabric
And music is our thread

I did not find it particularly easy to make music at the time, moving in a new direction that implied abandoning the use of instruments. The resulting works also

seemed to redraw from filling a social function, possibly the oldest and original purpose and context of music. Installations as shared spaces do not depend on interaction, but I first became aware of the possibility of embedding such qualities in the works through my collaborations with Asbjørn Flø and Jana Winderen, probably because interaction seemed to strengthen a potential for relations, but also because this was an important aspect of the art practice of Jana Winderen, both prior to and after our collaborations.

One of the qualities of the installation *Cubic Second* that I had not fully foreseen, but was very pleased about, was its ability to function as a social and shared experience. Jon Arne Mogstad commented how the spectators, standing around the horizontal video displays, reminded him of the paintings of Georges de Latour, as their faces were lit up by the light. The installation seemed to invite a concentrated sensing among the collective of people present at any time. As such I feel that *Cubic Second* for myself might point towards possibilities for new and different contexts as compared to the concert hall, while maintaining both the focused setting and possibly strengthen the ability for the experience to take place within a social context.

4.3 Performative Issues in Live Art

4.3.1 Works for Stage

Interaction in stage-based works is radically different from audience interaction, as the interaction is left to trained musicians, actors or dancers. In the staged-based projects I have been involved with the interaction design has been developed in close collaboration with the performer(s). The choice of processing algorithms and how to use them is a response to initial input by the performers. Simultaneously my live electronic processing alter the outcome of their actions. The processed recorder becomes a radically new instrument, requiring a different way of playing to activate the new expressive possibilities. A dancer with a sensor system is no longer moving in response to preexisting sound, instead movements produce sound. An actor no longer have complete control of her own acting, as acts of expression are radically transformed by technology.

This is often challenging to the performers. They partly loose control of their own expressive language as they are suddenly playing a completely different instrument than they are used to, and this instrument remains partly out of their control, as it is also controlled by the man with the mouse. In the development process the instrument itself changes continuously, as patches and structures are altered and refined. If interaction is based on input from microphones, stage monitoring might cause feedback, but on the other hand avoiding stage monitoring limit their ability to hear and judge the result of their own enacting. Getting comfortable with this way of working and performing require time, curiosity and patience.

A question coming up regularly is how important it is for the performers to understand the logics of the processes they are exposed to. There is a potential danger of the performers focusing too much on the technical aspects, working so hard to cognitively understand and control the processes that they loose the presence in the moment so important to a good performance. On the other hand an understanding of the processes involved might help performers to use it to the maximum. The real challenge here is to ensure that the performers by the end of the development phase truly master the processes they are exposed to the same way a good musician master her instrument. Ideally the borders between body and process should blur, so that a process fully

understood can be played intuitively.

Tsalal and Concert for Greenland

Verdensteatret is a Oslo-based collective of artists working on stage productions and other related projects, headed by Lisbeth J. Bodd and Asle Nilsen. My first involvement with Verdensteatret was in the autumn 2000 contributing to an installation based on the stage production *Règla*. At this time Verdensteatret was increasingly incorporating technology into their projects. In 2001 we started development of a new production that eventually would become *Tsalal*.

Quite a few problems and issues were raised in the early stage of development concerning music and sound for stage. In general sound for film and stage sound tend to be divided into a foreground and background layers. The foreground layer mainly consist of sound produced by the actors: speech, foot steps, etc. The background layer provides ambience, situation the action in a sonic environment. Most of the time music functions as an additional component of this layer, although music can get a more prominent role, in particular at transitional stages of a story.

For *Tsalal* we wanted to explore the use of live processed sound to challenge this dichotomy. The voices of all actors were amplified and processed in various ways, most of the time resulting in speech being incomprehensible, a digital onomatopoeia. Of special interest were ways of transforming background to foreground and vice versa: At one point one of the actors, Lars Øyno, spit water on the stage. The burst of sound is repeatedly looped, and gradually transformed using filters and feedback. Twice more he spits water. Sound accumulates and gradually transform into the sound of a train. Through transformations like this we tried to create a continuous field from foreground to background, enabling the voices of the actors to create the background layer, appear out of the background layer to become more prominent or fade from foreground into background.

The above approach was continued in the next production, *Concert for Greenland*, but in this production the approach seemed to expand to all parts of the production, blurring the boundaries between actors, scenography, light, shadow, video projections, voices, sound and music.

Hacker, Cracker, Tracker

2003-2006 I have been doing a series of performances combining dance, music and sound processing in collaboration with the dancer Gitte Bastiansen and the recorder player Frode Thorsen, using technology to create new and extended possibilities for interaction between the three of us. The dancer were using the DIEM Digital Dance system (Siegel and Jacobsen 1998), a gesture tracking system using bend sensors attached to the body, with wireless communicating of data as MIDI control change messages. The sound of the recorder captured using a microphone and further processed in MaxMSP. Other layers of electronic sound were added. The parameters of the various processes could be controlled dynamically by the dancer. Thus everyone had the possibility of responding to, interact with or interfere with the activity of the others.

The first production, *Hacker*, was commissioned by Rikskonsertene²⁷ and toured primary schools in the Oppland county spring 2003. The experience gained during this toured formed the basis for a production meant for adult audience presented at the contemporary music festival Autunnale in Bergen as *Cracker*. For this event the video

²⁷<http://www.rikskonsertene.no>

artist Gisle Frøysland joined in, doing real time video processing based on live video feed of the dancer and sensor and sound analysis data.

All of the material was completely reworked once again for further tours in Oppland county in 2005 and 2006, the project now being labeled *Tracker*. Tracker became a framework in terms of material, form and development with ample room for improvisation as well as more structured changes and adjustments from one show to the next.

A continuous challenge of this collaboration has been how to get meaningful information from the sensor system. Technically the system have been very reliable in the sense that all components are working, not failing technically once in close to hundred shows. But the information coming in is not necessarily the most meaningful in terms of tracking the movements of a dancer. When design such a system one should start out observing and filming dancers in movement, and analyze for how they move, what movements are of importance and where in the body movements originate from. This should be done in close dialogue with the dancers in order to get access to their experience as much as possible. My impression from observing Gitte Bastiansen was that the belly seemed to be a focal point of the body. Movements often seemed to radiate from the belly out towards hands, feet and head. If so information about the movements of the belly, sides and back would be more valuable than information on the movements of elbows and knees. The DIEM Digital Dance System seems instead to have been designed in response to the two questions ‘What technology do exist?’ and ‘Where can we put it?’ We ended up adding three bend sensors attached to belly and sides and analyze them for information on movements in frontal and sideways directions.

Another problem was inconsistency in the data received from the sensors. Bend sensors were supposed to provide values in the range from 0 though 127. In practice only a small part of that range was active. Sensors would differ in sensitivity, and also seemed to change during performances. Movements of the dancer could also cause them to move, thus shifting what range they reported. This was solved using statistics: Running means and standard deviation values were calculated (Castine 2002), and used for auto-scaling of the incoming sensor values.

4.3.2 Working with Real-Time Technology in Live Projects

Workflow

None of the projects for stage that I have been involved in have been developed according to a preexisting script and a prepared plan for scenography or choreography. Development has been process-oriented, material developed and rehearsed through a flow of improvisations, discussions and suggestions from the artists involved, as one of the aims of the production has been to explore new artistic possibilities offered by new technology. It is hard to venture into uncharted land if you insist on sticking to the map.

Work in these projects differs from development for installations. The production is convoluted, with many artists (and possibly technicians) present on and off stage at the same time, involved in a collective shared process. Artists working with technological mediums have to fill at least three often conflicting roles at the same time; contributing creatively as artists, develop technological solutions and play them technically and artistically in rehearsals. With one eye on the stage, the second on the screen and the third deep buried in a manual or looking for replies to questions asked at a mailing list, rehearsals tend to be hectic. In installations I usually develop my patches either

working alone, or sitting next to some of the collaborating partner, both immersed in our respective tasks. Even when there is continuous exchange concerning ideas and material, the nitty-gritty details of realizing the respective shares of the work is usually clearly divided according to competences and what medium we have main responsibility for.

In stage productions it is very hard to find this kind of time and space alone for development, testing, fool-proofing and debugging. Technology has been central to the process in the projects I have been involved in, with almost all other aspects of the production depending on it. If technological problems are encountered, or the artists controlling technology are absent from rehearsals for some reason, the process gets severely hampered. While the processing in programs such as MaxMSP is carried out in real-time, the development of patches can hardly claim to happen in real-time. Coming up with a new idea might take three seconds, but developing the patch required to realize it might take three hours or even three days, and in the mean-time everything risk coming to a crawl. Most of the time it is hard to say how long the development or alteration of a process or algorithm will take. Everything might seem easy to achieve beforehand and afterwards, but in the middle of it you never know what unforeseen problems might arise. This makes it hard for others to plan other tasks to work on in the meantime, and the actors, musicians or dancers to often end up waiting. When that happens for the second time in a day, energy and motivation is drained, and they struggle to mobilize the presence required for their contributions to the project. So far I have never found ways of avoiding this from happening, and in new collaboration with performing artists without previous experience from similar projects, I have to explain that this unfortunately has to be accounted for and considered part of the required working methods, although I of course try to reduce it as far as I can.

Such processes also becomes financially expensive, and commonly the artists working off-stage tends to be the ones working the hardest. In Verdensteatret productions, when the actors leave after seven hours, it is lunch time for the rest of us; we still have half of the working day left.

It is a continuous struggle to find the right balance between working fast so that the general flow of the production is not hindered more than necessary, and investing time in quality assurance, making sure that patches are sufficiently flexible, structured, documented and tested so that they do not cause problems further down the road. I have seen repeatedly when working with others that the importance of this has to be learned the hard way. If the long term quality of the technical development is down-prioritized, it will strike back as things grow more complex and is stress-tested, in other words the last week before the opening. In the worst case this might undermine the need for rehearsals of the completed work, or even threaten the ability to do the show altogether.

Potential Limitations of Max

In Verdensteatret several structural and methodological problems concerning using Max for live productions have been surfacing repeatedly. Currently we are three to four artists using Max extensively to control sound, video, light and robotics. According to Miller Puckette, the initial developer of Max, the program is a continuous encouragement of unstructured approaches²⁸. Every artist and developer using Max

²⁸His response to the presentation of Jamoma at ICMC 2006 was that we had to be crazy attempting to propose a protocol for standardized API for designing modules in Max, as Max is a continuous encouragement of unstructured approaches.

seems to find their own take on how to structure patches, or not. While a lot of low-level functionalities are developed and shared among third party developers²⁹, relatively few higher-level modules are available. Most likely few have found ways of structuring more complex patches so that they can be easily adopted by others. Within Verdensteatret this introduces obstacles in joint development, exchange of patches, and in designing a system for controlling patches that are transparent to everyone involved. Patches are easily privatized, reducing the potential for collaboration or at least introducing an element of inefficiency. It also makes the ensemble extremely dependent on the person(s) controlling patches in performance. In case of illness they might be more or less impossible to get someone to stand in.

A Max patch can be described as an algorithm waiting for some kind of input, and when receiving, it is performing a specific task on it producing some kind of output. If the exact same input is received over and over again, it is likely to produce the same output repeatedly. Random processes might be introduced complicating this description, but the inherent structures of Max patches is a sort of static engine making them prone to encouraging artistic results of a static nature. While this quality is one of the reasons why I was initially attracted to the program, and is a common feature of my installations, it can implicate artistic limitations when used for concerts and stage. If the production has development, narratives, or abrupt contrasts, one needs to be able to dynamically alter the processes and the output they produce. This will have to be done by a combination of changes in input to the patch, and changes in the parameters of the patch. The latter can itself be considered a kind of input to the patch.

A limiting factor in the *Hacker* and *Cracker* productions was the limited access to parameters of the patch when only the laptop monitor and mouse was used for interfacing with the patch. Only one parameter could be changed at a time, and this inevitably made transitions slow. If several parameters were required to change, it had to be done by altering them one by one. *Cracker* ended up appearing as a continuous morphing or drift through material. In some respects this worked well, but the ability of more rapid and abrupt changes were sorely missed as a contrasting possibility. At the time Max did not offer any real ready to go systems for managing cues so that changes of several parameters could be aggregated into more complex cues.

When making a patch, it is initially often conceived for a specific occasion; a specific processing performed to a specific input at a specific point during the concert or play. The underlying algorithm might be more flexible with more parameters that can be controlled, but as the implementation is tailor-suited to the task at hand, everything not of initially use gets buried deep down in the patch, out of sight and out of hand later on. If more parameters could remain available in the interface, the patch could turn out to be more flexible, more expressive and maybe also be used in other ways during the performance. It is a challenge to design patches so that the full flexibility of the patch remains available.

Development of new material in Verdensteatret often happens through a series of improvisations. If interesting material surfaces we try to hold on to it by attempts to reproduce so that it can be integrated into material already existing. This is combined with a continuous de- and reconstruction of material. Improvisations by their very nature are of a fleeting nature, depending on a continuous and partly subconscious sensitivity to surroundings and the others involved. With a lot of technology involved the situation becomes highly sensitive to small alterations in the parameters of the electronic media, minor changes can create enough of a disturbance to make the stage a

²⁹<http://www.maxobjects.com>

different place. The technical media are often depending on many parallel parameters. Reconstructing the progression of parameters from an improvisation can be daunting. A possibility for taking snapshots capturing current settings for all parameters, or even record all changes, so that they can be recalled later on would make the technology much more flexible, dynamic and manageable as an instrument for live improvisations and performance.

Developing Jamoma

I believe that the best way of finding a sustainable solution addressing the issues discussed here is by developing a firm but flexible methodology for structuring patches so that it is ensured that these matters are consistently solved in a way that makes patches manageable in development and controllable in performance. The need for developing a structured approach has been the motivation for my involvement with Jamoma.

In 2003 I was organizing a workshop on advanced control of sound installations hosted by NoTAM in Oslo. From prior e-mail exchanges I had the impression that Tim Place of Electrotap LLC had been into similar problems and started finding approaches that might be worth looking into. After the workshop his stay in Norway was extended with a residency at BEK, giving us more time to discuss these topics. He had started developing a framework for Max patches named Jade, appearing as a stand-alone application. By the time Verdensteatret started development of what would become *Concert for Greenland* Jade was still considered premature for use in the production, but in March 2005 Tim decided to open-source the modular structure as Jamoma (Jade Modules for Max), and I joined in as a developer. Jamoma has been used for all art projects I have done from then on, the first being the *Tracker* tour in April 2005. Jamoma development has to a high degree been informed by practical experience and needs in artistic projects, a close integration of artistic work and technological development. Additional developers have joined the project, and Jamoma is now developed in collaboration between Tim Place, myself, Alexander Refsum Jensenius, Pascal Baltazar and Dave Watson with additional input and suggestions from Alexander Fogar, Mathew Aidekman, John Hudak, Thomas Grill and Jeremy Bernstein. Jamoma has also been used for research in musical gestures at Department of Musicology, University of Oslo and Schulich School of Music, McGill University (Jensenius 2006) and (Marshall, Peters, Jensenius, Boissinot, Wanderley, and Braasch 2006).

Jamoma consists of two parts: A recommendation and an implementation of that recommendation. An important part of the recommendation is the structural subdivision of modules into graphical user interface, logical algorithm and a system for parameter and state handling, based on the Open Sound Control protocol (Wright, Freed, and Momeni 2003).

Jamoma offers a compelling set of benefits to users. These benefits include fast and flexible interchange of modules, patch-building and module construction, as well as possibilities of advanced control of the modules in performance. Jamoma modules may encapsulate any type of functionality that can be performed by Max, MSP, Jitter, its components (such as Java or JavaScript), or any third party objects. Jamoma might significantly ease development, control and maintenance of patches in large projects involving several artists using Max.

Modules might be controlled in a number of different ways. A system for remote communication enables a set of control modules for controlling other modules. Several such modules are implemented. *jmod.cuelist* loads a text-based script of event cues,


```
#####  
CUE sweep  
#####  
  
# Module filter  
/filter/cf 3000 3000  
WAIT 6000  
/filter/cf 200 3000
```

Figure 4.22: A cuescript. When the cue named *sweep* is called, center frequency of the filter module ramps to 3000 Hz over 3 seconds, holds for 3 seconds, and then ramps down to 200 Hz over 3 seconds.

and is able to control all modules provided that they have been given unique names. The cues can be executed in arbitrary order. A *WAIT* syntax can set the execution of a cue on hold for a specified amount of time, opening up the possibility for scripting of complex events evolving over time. The current state of all modules can be queried, and used to create new cues. Figure 4.22 illustrates a simple cue controlling one parameter of one module.

Other control modules exist for mapping of parameters between different modules, and for enabling network communication using Open Sound Control. Such modules are able to control themselves as well, offering possibilities for complex recursive generative systems.

Chapter 5

Concluding Remarks

Sound is energy, disturbances of pressure levels of the atmosphere. For sound to exist air, or space, is required to propagate through, as well as an initial disturbance releasing the energy. Sound is an indication of dynamic action; “Sounds signify events taking place” (Blessner and Salter 2006). This could be restated as “Sounds signify events taking place in space”.

The many problems and issues that I have been working with as an artist over the last decade, currently seems to be in a process of convergence. Throughout the last three years “body” or “the performative”, or possibly more correctly, how sound and humans relate, have evolved as a common denominator. Likewise the relationship between sound and space has emerged as a fundamental issue. Increasingly these two issues seems to be the flip sides of a Möbius strip.

Ten years ago I would not have expected this to happen. At the time my concern was very much about creating an environmental space to be experienced in solitude. I have probably changed quite a bit myself, as a consequence of the accumulated experience of collaborations with other artists, the experience of working at BEK, a position where serving other artists and getting involved in and assisting their work was a main component, and other more private events.

Békesy suggested the idea of the auditive horizon; As sound sources move away from us, there is a threshold beyond which we are no longer able to determine changes in distance only from the signal of the isolated sound source reaching our ears. The car passing by never really vanishes, but eventually it is immersed in an ever-present bed of background noise.

This background noise currently fascinates me. The arts have long striven for the sheltered spaces: the decoupled concert hall, the infinite black box and the neutral white cube. I consider *Cubic Second* the most concentrated and focused space I have managed to create for my works so far. Paradoxically, due to the sheltered quality of the room, it seemed to prevent me from letting sound disappear into the horizon as there were not enough background noise to be masked in. Labelle and Roden (2000) speak of works that “do not cut themselves off from location, interference, or unwanted noise, but rather embrace these elements as an important compositional source”. The challenge I am currently pondering is how to allow a background layer, with a high degree of richness in information and possibilities for integration with or continuum towards the foreground. The drone, the stasis of minimalism and the backdrops of field recordings all have limited possibilities for dynamics, change and interference, and I would like to work on how to extend or move past these models.

I am also pondering how to do so while being able to maintain a degree of fidelity or depth of information in the work. In some respects artistic work resembles science. Both are focusing on a subject or topic of some kind, and a series of assumptions and simplifications have to be applied in order to isolate the subject sufficiently for something meaningful to be expressed. The dark, seemingly borderless space at Hordaland kunstsenter was the condition forming the basis for the *Cubic Second* installation, the prerequisite for a concentrated sensuous experience of sound and vision in the center of the space. How to maintain a situation maintaining this kind of focus while being based on less strict assumptions, allowing more integrated and complex surroundings?

Chapter 6

Artistic outcome

List of works realized within the framework of the research project 2003-2006.

6.1 Installations

Solo exhibitions

Dialog/Overgang (Dialogue/Transition) (LMW 1) Installation of video, digital prints, paintings and sound. Trond Lossius, Jon Arne Mogstad and Jeremy Welsh. Christiansands kunstforening, Kristiansand, Norway, March - April 2004

Lydspor (Sound Track) Interactive sound installation. Asbjørn Flø and Trond Lossius. Norsk Form, Ultima 2004, Oslo, October 2004

A Can of Stellar Stripes and Other Vessels (LMW 3) . Installation of sound, video, prints, light, paintings, objects. Trond Lossius, Jon Arne Mogstad, Jeremy Welsh. Trondhjems kunstforening, Trondheim, Norway, September 2004

White-out (Elektropoesia II) Audiovisual installation. Trond Lossius and Kurt Ralske. Visningsrommet USF, Bergen, Norway, September - October 2005

Cubic Second Installation. Trond Lossius. Hordaland kunstsenter, Bergen, Norway, November 2006

Touring Exhibitions

Trickster Anaglyphical stereo-photographic prints, sound. Andrea Sünder-Plassmann and Trond Lossius. Vestlandsutstillingen (West Norway Touring exhibition) 2005

Music as process, music as weather Sound installation. Opening at Kunsthallen på Tulinløkka, The National Museum of Art, Architecture and Design, Oslo, Norway, September - October 2005

Group exhibitions, etc.

Kunsten å gå på vannet (The art of walking on water) Interactive sound installation. Jana Winderen, Jørgen Træen and Trond Lossius. Commissioned for Den kulturelle skolesekken - Turnéorganisasjonen for Hedmark. Touring schools in the Hedmark region, Norway, 2004-2006. Huddersfield Art Gallery - Huddersfield Contemporary Music festival, Huddersfield, UK, 2005,. Festspillene i Bergen - Nystemt - Grieghallen, Bergen, Norway, May 2005. Barnas verdensdag Oslo, 2005. Kunst torget/Riksutstillinger, Norsk Form, 2005.

Elektropoesia Audio visual installation. Trond Lossius and Kurt Ralske. Electrohype - Nordic Biennale for Computer-based art. Malmö konsthall, Malmö, Sweden, November 2004 - January 2005

Quiet and relaxed, but alert Interactive sound installation. Jana Winderen og Trond Lossius. Galleri F15, Moss, Norway, May - August 2005

Staged Spaces Media installation. Karen Kipphoff og Trond Lossius. Bergen kunsthall, Norway, November - December 2005. The National Autumn Exhibition, Kunsternes Hus, Oslo, Norway - September 2006

What goes around, comes around Installation of video, paintings and sound. Trond Lossius, Jon Arne Mogstad and Jeremy Welsh. Seljord kunstforening, Seljord, Norway, July - August 2004

Improsynth / jc87 Container installation. Trond Lossius, Jon Arne Mogstad and Jeremy Welsh. Q-art, Quart Festival, Kristiansand, Norway, July - 2006

6.2 Stage works

Konsert for Grønland (Concert for Greenland) "Concert-installation-performace" by Verdensteatret. Premiered at Black Box, Oslo, March 2004. Trondheim kunstmuseum/Trondheim Matchmaking, Trondheim, Norway, 2004. Kulturhuset Kilen Stockholm, Sweden, September, 2004. Ultima/Black Box, Oslo, Norway, October 2004. BiT Teatergarasjen/Detox, Bergen, December 2004. Operahgskolen, Kunsthgskolen i Oslo, 2005. Stamsund Internasjonale Teaterfestival, Stamsund, Norway, 2005. The International Theatre Festival Divadelna Nitra, Nitra, Slovakia, 2005. Tou Scene. Stavanger, Norway, 2005. PS 122, New York City, 2005. Dom im Berg, Steirischer Herbst Festival -06 Graz, Austria, 2006 de Singel Antwerpen, Belgium, 2006. Trafo House of Contemporary Art, Budapest, Hungary, 2006

Tracker Performance/concert with sound, music and dance. Trond Lossius, Gitte Bastiansen, Frode Thorsen. Rikskonsertene, Touring schools in the Oppland region, Norway, 2005-2006

6.3 Music commission

Ex te lux oritur commission for Holberg Prize Award Ceremony 2004. Trond Lossius (sound processing) and Agnethe Christensen (mezzo-soprano, cantele). Haakonshallen, Bergen, Norway, December 3 2004

6.4 Various projects and events

Dialog/Overgang - remix (LMW 2) Installation of video, sound and painting. Trond Lossius, Jeremy Welsh, Jon Arne Mogstad. Part of the Beta event at BiT Teatergarasjen, Bergen, May 3 2004

Dialog/Overgang - remix (LMW 4) Installation of video, sound and painting. Trond Lossius, Jeremy Welsh, Jon Arne Mogstad. Bergen Open/Bergen National Art Academy, December, 2004

...different from the one you are in now... (net based sound art). Trond Lossius. IMPACT ME'05, Palestina and Israel, 2005.

What goes around, comes around (LMW - 5) Installation of video, sound and painting. (work in progress). Trond Lossius, Jeremy Welsh, Jon Arne Mogstad. Bergen National Academy of the Arts, Bergen, Norway, January 13 2006

6.5 Prices

2005-2006 New York Dance and Performance Awards (A.K.A. The Bessies) Installation & New Media category. Received by Verdensteatret for *Concert for Greenland* performances at P.S. 122

Bibliography

- Baalman, M. (2004). Application of Wave Field Synthesis in electronic music and sound installations. In *Proceedings of the 2nd International Linux Audio Conference*.
- Bevilacqua, F., R. Müller, and N. Schnell (2005). MnM: a Max/MSP Mapping Toolbox. In *Proceedings of the 2005 Conference on New Interfaces for Musical Expression*.
- Bishop, C. (2005). *Installation Art: A Critical History*. Tate Publishing.
- Blauert, J. (1996). *Spatial Hearing - Revised Edition: The Psychophysics of Human Sound Localization*. The MIT Press.
- Blessner, B. and L.-R. Salter (2006). *Spaces Speak, Are You Listening?: Experiencing Aural Architecture*. The MIT Press.
- Borgdorff, H. (2006). The Debate on Research in the Arts. *Sensuous knowledge. Bergen National Academy of the Arts 2*.
- Boulangier, R. (Ed.) (2000). *The Csound Book: Perspectives in Software Synthesis, Sound Design, Signal Processing, and Programming*. The MIT Press.
- Bourriaud, N. (1998). *Relational Aesthetics*. Les Presse Du Reel, France.
- Cage, J. (1961). *The Future of Music: Credo*. In *Silence: Lectures and Writings*, pp. 3–6. Wesleyan University Press.
- Cage, J. (1989). *An Autobiographical Statement*. <http://www.newalbion.com/artists/cagej/autobiog.html>. Retrieved 2007-01-19.
- Castine, P. (2002). *Litter Power Package Documentation*. DSP Audio.
- Caulkins, T., E. Corteel, and O. Warusfel (2003). Wave Field Synthesis Interaction With the Listening Environment, Improvements in the Reproduction of Virtual Sources Situated Inside the Listening Room. In *Proceedings of the 6th International Conference on Digital Audio Effects (DAFx-03)*.
- Craven, P. and M. Gerzon (1977). Coincident Microphone Simulation Covering Three Dimensional Space and Yielding Various Directional Outputs, U.S. Patent no 4042779.
- Crocker, R. L. (1966). *A History of Musical Style*. McGraw-Hill, reprinted by Dover Publications, Inc. 1986.
- Dack, J. (1999). Karlheinz Stockhausen's Kontakte and Narrativity. *eContact*. On-line: <http://econtact.ca/> 2(2).
- Dahlquist, G. and Å. Björk (1974). *Numerical Methods*. Prentice-Hall.

- de Haan, S. (2002). The relationship between composer, performer and listener in 20th Century music making. http://www.brisinst.org.au/resources/dehaan_simone_compose.html. Retrieved 2007-02-07.
- de Oliveira, N., N. Oxley, and M. Petry (1994). *Installation Art*. Thames & Hudson Ltd.
- Dobrian, C. (1988). Music Programming: An Introductory Essay. <http://music.arts.uci.edu/dobrian/CD.MusicProgramming.htm>. Retrived 2007-01-21.
- Dodge, C. and T. A. Jerse (1997). *Computer Music: Synthesis, Composition, and Performance*. Schirmer.
- Eargle, J. (2004). *The Microphone Book. 2nd Edition*. Focal Press. Elsevier.
- Elsa, P. (1995). Fuzzy Logic and Musical Decisions. Technical report, University of California, Santa Cruz.
- Emmerson, S. (2001). From Dance! to "Dance": Distance and Digits. *Computer Music Journal* 25 (1), 13–20.
- Eno, B. (1976). Generating and Organizing Variety in the Arts. *Studio International*, 226–233. Reprinted in *Audio Culture: Readings in Modern Music*, ed. C. Cox and D. Warner, The Continuum International Publishing Group Inc. 2004.
- Eno, B. (1996a). *Ambient Music*. In *A Year with Swollen Appendices*, pp. 293–297. Faber & Faber Ltd, London.
- Eno, B. (1996b). *Generative art*. In *A Year with Swollen Appendices*, pp. 330–332. Faber & Faber Ltd, London.
- Essl, K. (1995). Lexikon-Sonate. An Interactive Realtime Composition for Computer-Controlled Piano. In *Proceedings of the "II Brazilian Symposium on Computer Music" (Canela 1995)*.
- Everest, A. F. (2000). *Master Handbook of Acoustics. 4th edition*. McGraw-Hill/TAB Electronics.
- Farina, A. and F. Righini (1997). Software Implementation of an MLS Analyzer with Tools for Convolution, Auralization and Inverse Filtering. In *Pre-prints of the 103rd AES Convention, New York, May*.
- Farina, A. and E. Ugolotti (1998). Software Implementation of B-Format Encoding and Decoding. In *Pre-prints of the 104 rd AES Convention, Amsterdam, 15 - 20 May, 1998*.
- Gerzon, M. and G. J. Barton (1992). Ambisonic Decoders for HDTV. In *Pre-prints of the 92nd AES Convention, Vienna, March*.
- Hammer, E. (2005). Hosted/Visited: Two perspectives on La Monte Young & Marian Zazeela's Dream House. Master's thesis, University of Bergen.
- Holmes, T. (2002). *Electronic and Experimental Music*. Routledge.
- ITU (1993). Recommendation BS. 775: Multi-channel stereophonic sound system with or without accompanying picture. Technical report, International Telecommunications Union.
- Jaschko, S. (2004). Segmentation and reassembly of the digital moving image: an artistic method of depiction of temporal and spatial nonlinearity. In *Proceedings of the 12th International Symposium of Electronic Arts (ISEA2004)*.

- Jaukkuri, M. (2004). *Juxtapositions, slippages, breaks, and junctures*. In Jon Arne Mogstad: *Exhibition Catalogue*, pp. 18–29. Kunstnerforbundet, Oslo.
- Jehan, T., A. Freed, and R. Dudas (1999). Musical Applications of New Filter Extensions to Max/MSP. In *Proceedings of the 1998 International Computer Music Conference*.
- Jensenius, A. R. (2006). Using Motiongrams in the Study of Musical Gestures. In *In Proceedings of the 2006 International Computer Music Conference*.
- Jot, J.-M. (1996). Synthesizing Three-Dimensional Sound Scenes in Audio or Multimedia Production and Interactive Human-Computer Interfaces. In *5th International Conference: Interface to Real & Virtual Worlds, Montpellier, France*.
- Kahn, D. (1999). *Noise, Water, Meat: A History of Sound in the Arts*. Cambridge, The MIT Press.
- Kay, J., K. Ghent, B. Chumney, and E. Lutkins (1997). *Film Sound History: 40's*. <http://www.mtsu.edu/smp/forties.html>. Retrieved 2007-01-19.
- LaBelle, B. (2006). *Background Noise: Perspectives on Sound Art*. Continuum International Publishing Group - Academi.
- Labelle, B. and S. Roden (Eds.) (2000). *Site of Sound: of Architecture and the Ear*. Errant Bodies Press.
- Laborie, A., R. Bruno, and S. Montoya (2003). A New Comprehensive Approach of Surround Sound Recording. In *Pre-prints of the 114th AES Convention, Amsterdam*.
- Lindemann, E. (2006). *Synful Orchestra Users Guide* (2.3.2 ed.). Synful LLC.
- Litovsky, R. Y., H. S. Colburn, W. A. Yost, and S. J. Guzman (1999). The Precedence Effect. *Journal of the Acoustical Society of America* 106(4 Pt 1), 1633–1654.
- Lossius, T. (2005). Sound Installations and Other Inter-Diciplinary Projects. In N. Malterud, P. Crabtree, P. Kvist, and E. Wiig (Eds.), *Årbook 2003-04*, pp. 96–103. Kunsthøgskolen i Bergen.
- Loy, G. (1985). Musicians Make a Standard: The MIDI Phenomenon. *Computer Music Journal* 9(4, reprinted in *The Music Machine*, ed. C. Roads, Cambridge, The MIT Press, 1989.), 8–26.
- Malham, D. (2003). *Higher Order Ambisonic Systems. Abstracted from the Mphil thesis Space in Music - Music in Space*. Ph. D. thesis, University of York.
- Malham, D. and A. Myatt (1995). 3-D Sound Spatialisation using Ambisonic Techniques. *Computer Music Journal* 19(4), 58–70.
- Malham, D. G. (1992). Experience with Large Area 3-D Ambisonic Sound Systems. In *Proceedings of the Institute of Acoustics Autumn Conference on Reproduced Sound* 8.
- Manovich, L. (1996). On totalitarian interactivity (notes from the enemy of the people). <http://www.manovich.net/text/totalitarian.html>. retrived 2007-01-16.
- Manovich, L. (2001). *The Language of New Media*. The MIT Press.
- Marshall, M. T., N. Peters, A. R. Jensenius, J. Boissinot, M. M. Wanderley, and J. Braasch (2006). On the Development of a System for Gesture Control of Spatialization. In *In Proceedings of the 2006 International Computer Music Conference*.

- Mathews, M. (1999). *What is Loudness?* In *Music, Cognition, and Computerized Sound: An Introduction to Psychoacoustics*, ed. P. Cook, Chapter 6, pp. 71–78. The MIT Press.
- McKeeg, A. and D. S. McGrath (1997). Using Auralization Techniques to Render 5.1 Surround to Binaural and Transaural Playback. In *102nd AES Convention. Preprint 4458*. Audio Engineering Society.
- McKenzie, A. M. (2006). What and Why. <http://www.brainwashed.com/h3o/whatis.html>. Retrived 2007-01-31.
- Merimaa, J. and V. Pulkki (2004). Spatial Impulse Response Rendering. In *Proceedings of The 7th International Conference on Digital Audio Effects (DAFx'04)*, pp. 139–144.
- Microsoft (2001). Multiple Channel Audio Data and WAVE Files. Technical report, Microsoft Corporation.
- MMA (1996). MIDI 1.0 Detailed Specification. Document Revision 4.2. Technical report, The MIDI Manufacturers Association.
- Momeni, A. and D. Wessel (2003). Characterizing and Controlling Musical Material Intuitively with Geometric Models. In *Proceedings of the 2003 Conference on New Interfaces for Musical Expression (NIME-03)*, Montreal, Canada.
- Moore, B. C. (2003). *An Introduction to the Psychology of Hearing, Fifth Edition*. Academic Press.
- Moreau, S., J. Daniel, and S. Bertet (2006). 3D Sound Field Recording with Higher Order Ambisonics: Objective Measurements and Validation of a 4th Order Spherical Microphone. In *Pre-prints of the 120th AES Convention, Paris*.
- Mumford, L. (1934). *Technics and Civilization*. First published New York, Harcourt Brace,.
- Pakes, A. (2004). Art as Action or Art as Object? The Embodiment of Knowledge in Practice as Research. Working Papers in Art & Design. <http://www.herts.ac.uk/artdes1/research/papers/wpades/vol3/apabs.html>. Retrived 2007-02-07.
- Pinch, T. and F. Trocco (2004). *Analog Days: The Invention and Impact of the Moog Synthesizer*. Harvard University Press.
- Place, T. and T. Lossius (2006). Jamoma: A Modular Standard for Structuring Patches in Max. In *Proceedings of the 2006 International Computer Music Conference*. New Orleans: International Computer Music Association.
- Place, T., N. Wolek, and J. Allison (2005). *Hipno: Getting Started*. Cycling'74 and Electrotap.
- Polfreman, R. (1999). A Task Analysis of Music Composition and Its Application to the Development of Modalyser. *Organized Sound* 4(1), 31–43.
- Puckette, M. and D. Zicarelli (1990). *MAX - An Interactive Graphical Programming Environment*. Menlo Park: Opcode Systems.
- Pulkki, V. (1997). Virtual Sound Source Positioning Using Vector Base Amplitude Panning. *Journal of the Audio Engineering Society* 45(6), 456–466.
- Pulkki, V. (1999). Uniform Spreading of Amplitude Panned Virtual Sources. In *Proceedings of the 1999 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*.

- Pulkki, V. (2000). Generic Panning Tools for Max/MSP. In *Proceedings of International Computer Music Conference 2000.*, pp. 304–307.
- Pulkki, V. (2006). Directional Audio Coding in Spatial Sound Reproduction and Stereo Upmixing. In *The Audio Engineering Society 28th International Conference.*
- Pulkki, V. and C. Faller (2006). Directional audio coding: Filterbank and STFT-based design. In *120th Audio Engineering Society Convention.*
- Reiss, J. H. (2000). *From Margin to Center: The Spaces of Installation Art.* The MIT Press.
- Risset, J. C. (2003). Computer Music: Why? <http://www.utexas.edu/cola/france-ut/archives/Fall2003/risset.pdf>. Retrived 2007-02-08. Online pdf document.
- Roads, C. (1996). *The Computer Music Tutorial.* The MIT Press.
- Roads, C. (2001). *Microsound.* The MIT Press.
- Rodet, X., Y. Potard, and J.-B. Barrière (1985). The CHANT Project : From Synthesis of the Singing Voice to Synthesis in General. *Computer Music Journal* 8(3), 15–31. Reprinted in C. Roads, ed. 1989. *The Music Machine.* MIT Press.
- Rudi, J. (2003). Norge - Et Lydrike, Norway Remixed: A Sound Installation. *Organised Sund* 8(2), 153–162.
- Rumsey, F. (2001). *Spatial Audio (Music Technology) (Music Technology Series).* Focal Press.
- Schacher, J. C. and P. Kocher (2006). Ambisonics Spatialization Tools for Max/MSP. In *Proceedings of the 2006 International Computer Music Conference.*
- Schafer, R. M. (1977). *The Soundscape: Our Sonic Environment and The Tuning of the World.* Destiny Books.
- Schnell, N., R. Borghesi, D. Schwarz, F. Bevilacqua, and R. Müller (2005). FTM: Complex Data Structures for Max. In *Proceedings of the 2005 International Computer Music Conference.*
- Schwarz, D., G. Beller, B. Verbrugghe, and S. Britton (2006). Real-Time Corpus-Based Concatenative Synthesis with CataRT. In *Proceedings of the 9th Int. Conference on Digital Audio Effects (DAFx-06), Montreal, Canada, September 18-20.*
- Sekkingstad, S. (2004). Kunstens nye stemme: noen betraktninger om lyd i samtidskunsten. Master's thesis, University of Bergen.
- Shlomowitz, M. (1999). Cage's Place In the Reception of Satie. As part of his Ph.D. at the University of California at San Diego, USA. <http://www.af.lu.se/fog-wall/article8.html>.
- Siegel, W. and J. Jacobsen (1998). The Challenges of Interactive Dance: An Overview and Case Study. *Computer Music Journal* 22(4), 29–43.
- Spors, S., H. Teutsch, A. Kuntz, and R. Rabenstein (2004). *Sound Field Synthesis. In Audio Signal Processing for Next-Generation Multimedia Communication Systems, ed. Y. Huang and J. Benesty, pp. 323–344. Springer.*
- Sundberg, J., A. Askenfelt, and L. Frydén (1983). Musical Performance: A Synthesis-by-rule Approach. *Computer Music Journal* 7(1), 37–43. Reprinted in C. Roads, ed. 1989. *The Music Machine.* MIT Press.

- Tamm, E. (1995). *Brian Eno: His Music and The Vertical Color of Sound*. Da Capo Press, Inc. New York.
- Treib, M. (1996). *Space Calculated in Seconds: The Philips Pavilion, Le Corbusier, Edgard Varese*. Princeton University Press.
- Welsh, J. (2005). The Bridge : The Expanded Field of Painting and Digital Media. In N. Malterud, P. Crabtree, P. Kvist, and E. Wiig (Eds.), *Årbook 2003-04*, pp. 88–95. Kunsthøgskolen i Bergen.
- 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.
- Wright, M., A. Freed, and A. Momeni (2003). Opensoundcontrol: State of the art 2003. In *Proceedings of the 2003 Conference on New Interfaces for Musical Expression (NIME-03), Montreal, Canada*, pp. 153–160.
- Zicarelli, D. (1998). An Extensible Real-Time Signal Processing Environment for Max. In *Proceedings of the 1998 International Computer Music Conference*, pp. 463–466. San Francisco: International Computer Music Association.
- Zicarelli, D. (2002). How I Learned to Love a Program that does Nothing. *Computer Music Journal* 26 (4), 44–51.