The Composition and Performance of Spatial Music

A dissertation submitted to the University of Dublin for the degree of Doctor of Philosophy

Enda Bates

Trinity College Dublin, August 2009.

Department of Music &

Department of Electronic and Electrical Engineering

Trinity College Dublin

Declaration

I hereby declare that this thesis has not been submitted as an exercise for a degree at this or any other University and that it is entirely my own work.

I agree that the Library may lend or copy this thesis upon request.

Signed,

____

Enda Bates

Summary

The use of space as a musical parameter is a complex issue which involves a number of different, yet interrelated factors. The technical means of performance, the sonic material, and the overall musical aesthetic must all work in tandem to produce a spatial impression in the listener which is in some way musically significant. Performances of spatial music typically involve a distributed audience and often take place in an acoustically reverberant space. This situation is quite different from the case of a single listener at home, or the composer in the studio. As a result, spatial strategies which are effective in this context may not be perceived correctly when transferred to a performance venue. This thesis examines these complex issues in terms of both the technical means of spatialization, and the compositional approach to the use of space as a musical parameter. Particular attention will be paid to the effectiveness of different spatialization techniques in a performance context, and what this implies for compositional strategies which use space as a musical parameter. Finally, a number of well known works of spatial music, and some original compositions by the author, are analyzed in terms of the perceptual effectiveness of the spatialization strategy.

The results of a large number of listening tests and simulations were analysed to determine the fundamental capabilities of different spatialization techniques under the less than ideal conditions typically encountered during a performance. This analysis focussed on multichannel stereophony, Ambisonics, and Wavefield Synthesis. Other methods which are orientated toward a single listener are not addressed in this thesis. The results indicated that each spatialization scheme has particular strengths and weaknesses, and that the optimum technique in any situation is dependent on the particular spatial effect required. It was found that stereophonic techniques based on amplitude panning provided the most accurate localization but suffered from a lack of spaciousness and envelopment. Ambisonics provided an improved sense of envelopment but poor localization accuracy, particularly with first order Ambisonics systems. Consequently it would appear that stereophony is preferable when the directionality and focus of the virtual source is paramount, while Ambisonics is preferable if a more diffuse enveloping sound field is required. Ambisonics was consistently preferred for dynamically moving sources as this

technique eliminated the panning artefacts exhibited by amplitude panning as the source moves from a position at a loudspeaker, to one inbetween a pair of loudspeakers. The decoding scheme and order of the Ambisonics system also has a significant effect on the perceptual performance of the system, particularly at off-centre listener positions. A single-band, max-r_E decoding scheme was found to be the most suitable approach for a distributed audience, and increasing the order of the system was shown to improve the performance at all listener positions. It is recommended that an octagonal array be adopted as a minimum standard for performances of multichannel spatial music, as this arrangement can be utilized for higher order Ambisonics and can also be readily implemented with digital audio hardware.

Wavefield synthesis (WFS) was found to be quite distinct from multichannel techniques such as stereophony or Ambisonics. The spatial aliasing frequency is a critical aspect of any WFS system and localization errors and timbral distortions significantly increase if this parameter is too low. The ability of WFS systems to position virtual sources both behind and in front of the loudspeaker array was shown to be extremely difficult to achieve, particularly if the listener's position is fixed or if the performance space contains significant early reflections and reverberation.

In the latter half of this thesis, a number of landmark works of spatial music were presented and analysed in terms of the perceptual validity of their approach to spatialization. It was shown that many composers have used spatial distribution to improve the intelligibility of different layers of material, and this approach was found to agree with the findings of scientific research in the area of auditory cognition. The use of recognizable spatial motifs was shown to be highly difficult to implement, and complex, abstract spatial designs are only indirectly related to what is eventually perceived by the audience. A gestural approach to spatial music has its origins in the practice of diffusion, yet this approach is equally applicable to other aesthetics and would seem to be highly suitable for mixed-media electroacoustic works. The use of augmented instruments which map the actions of the performer to a spatialization algorithm would seem to be well suited to performances of mixed-media spatial music. In addition, the use of flocking algorithms to control spatialization and sound synthesis also appears to be a novel and effective techniques for the creation of spatially dynamic, electronic sounds.

Acknowledgements

I am deeply grateful to my two supervisors, Dr. Dermot Furlong and Mr. Donnacha Dennehy, for all their support and guidance over the past four years. Without their incredible knowledge and encouragement, this thesis would not have been possible.

I would also like to particularly thank Dr. Fionnuala Conway and Gavin Kearney for their help over the years, all those who took part in listening tests, and my colleagues at the Spatial Music Collective.

Finally I am very grateful for the support of my family and friends, and all the students and staff of the Music and Media Technology course in Trinity College.

Table of Contents

List of Figures

List of Tables

1 Introduction

One of the defining characteristics of electroacoustic music has been the use of technology to expand musical boundaries, whether through electronic processes in conjunction with traditional instruments, or through electronic processes alone. This allowed previously neglected musical parameters such as timbre to come to the fore, while other parameters such as rhythm, melody and traditional harmony were relegated to the background. The use of space as a musical parameter is another novel aspect of this artistic movement. However, space is in many respects, fundamentally different from these other parameters. Timbre relates to the spectral and temporal relationship between the components of an individual sound object, while rhythm, melody and harmony involve temporal and spectral relationships between sound objects. Space as a musical parameter is, however, much broader and more difficult to define. It incorporates the dimensions of individual sound objects, the relationships between sound objects and even the relationship between the sound objects and the acoustic space in which they are heard. Furthermore, audible space itself can only become apparent through the temporal development of sounds themselves. Space as a musical parameter is therefore all-encompassing and yet difficult to define. It encompasses every aspect of a sound and yet the spatial aspects of a sound are often not consciously perceived as separate from the sound itself. In our evolutionary development, the *where* is dealt with immediately by our subconscious while the *what* and *why* become the primary focus of conscious attention.

1.1 Spatial Music: A Personal Perspective

"Men fight for freedom, then they begin to accumulate laws to take it away from themselves."
Thomas Jefferson

The most difficult question I encounter when composing a piece of spatial electroacoustic music is why, specifically, am I doing what I'm doing? Why am I placing a sound at this location, or why am I moving a sound in space in this particular way? Why use a certain from of audio synthesis or this particular spatialization technique? As an Irish composer in the early 21st century I see no necessity to pledge allegiance to any one aesthetic or style. Indeed, as someone whose musical origins are in punk rock, I find the idea of an overarching musical ideology which orders and informs every composition to be distinctly unappealing. The most attractive aspect of electroacoustic composition, at least to this composer, is the complete freedom to write and perform music using any style or genre, any instrument or object, and the most sophisticated and cutting-edge technology. However, the fact remains that in the absence of an overriding ideology or aesthetic, the questions presented at the start of this section prove troublingly difficult to answer. Of course intuition and instinct are a critical and often underestimated aspect of the compositional process. The initial choice of material and basic idea for a piece are generally intuitively chosen, and many of the specific decisions made during the composition may also be instinctual. However, the development of the initial idea into the larger structure of the overall work generally requires some form of intellectual process. When walking, each individual step may be taken intuitively, but in order to get anywhere, we need to consciously think about where we are going, and why.

1.1.1 What Now in the Age of Disillusionment

> *"Order without liberty and liberty without order are equally destructive."*
> Theodore Roosevelt

The basic philosophy outlined in the previous Section is not in any sense unusual and it could in fact be considered as quite symptomatic of our times. The title of this section has been used to refer to various time periods following the First World War, however this description is as appropriate now as in any other period of history. The English speaking world has lost a great deal of faith in science, religion, politics, and the media (often it should be said with very good reason) and while this can be in many respects quite liberating, it may also lead to a certain loss of direction. In a musical context, this dichotomy has been further magnified by the further development of digital technology which fundamentally alters the relationship between the composer and their instrument. In the past, the sheer expense of hardware-based music technology meant that composers were by necessity limited in the technological options available to them. However, this in turn forced composers to be highly creative in their use of these limited devices, and to develop sophisticated aesthetical frameworks which justified their particular mode of composition. In contrast, when a modern composer starts an application such as Cycling74's Max/MSP, Csound or Supercollidor, they see an empty screen, which may then be filled in a bewilderingly variety of ways. Virtually any sound, acoustic of synthetic, is readily and often freely available, and these sounds may be played or transformed using virtually any technique ever invented. Complex sounds and processes can therefore be rapidly produced simply by opening presets and randomly adjusting parameters, however, this approach quickly leads to aimless composition and cliché. This dichotomy between freedom and order is also apparent in the contrast between the creative possibilities offered by this software, and the logical and highly structured approach required to actually write a program which can achieve these artistic goals. The two sided nature of this process is mirrored in my own work as a composer. I desire some form of logical framework to guide my inspiration to a finished piece of music, but this framework is not some overarching philosophy, merely the particular, and importantly, the most appropriate aesthetic for this particular piece.

1.1.2 Why Spatial Music?

Electronic music outside of the electroacoustic tradition consists of a vast array of different styles and genres, many of which are strongly dance orientated. In this context, the performers ability to work the crowd is perhaps as important as their interaction with their musical instrument, which may be a laptop, hardware controllers, or even turntables and vinyl records. In less dance-orientated styles, the situation is much closer to that of a traditional performance, as the audience is more concerned with observing the performer. However, there is a fundamental difference between observing a performance by a violinist, and a performance by a laptop musician. In the latter case, it is often difficult to relate the visible actions of the performer to the audible result, as there is no clearly defined relationship between the physical input and the resulting sound. The visual actions of the performer will be the same if they are synthesizing and creating sounds in real-time, or if they are simply triggering the playback of pre-recorded tracks. When I encountered spatial electronic music for the first time, I was immediately struck by the contrast between the dynamic movement of sounds in space in this music, and the static nature of the sounds and the performers at the concerts of electronic music I had encountered up to this point. I very quickly formed the opinion that spatialization was a critical and necessary aspect of electronic music, as the spatial movement of sounds provided a physicality and dynamism that greatly enhanced the listening experience. This opinion was reinforced when I realized that the earliest practitioners of electronic music had also believed in the absolute necessity of spatialization, particularly in a performance context (see Chapter 8).

As with many other composers, this early enthusiasm was quickly tempered by the experience of hearing my work performed for the first time outside of the studio over a large loudspeaker array. The piece *Discordianism*, which is discussed in detail in Chapter 10.2, is a clear example of a work which functions quite well for a single listener, but rapidly degrades as the size of the listening area increases. This is particularly true of the dynamically moving noise drone which is the focal point of the third and final movement. Amplitude panning and a quadraphonic system simply could not effectively reproduce this movement when the distance between the loudspeakers was increased to cater for a larger audience. However, having created a stereo mix of this piece for a CD release, I remained convinced of the benefits of spatialization, as much of the dynamism and clarity of the work was lost in the reduction to two channels. As I began to explore these issues it quickly became apparent that a gap existed between empirical research on spatialization techniques and the issues faced by a composer of spatial music during a live performance. The majority of the scientific research conducted in this area has focused on the ideal conditions of a single listener in an acoustically treated space. In addition, much of the writing by composers of spatial music has discussed both the single listener and performance contexts and this seemed somewhat anachronistic considering the drastic decline in sales of music media in the modern era and the continued widespread enthusiasm for live performances. These very practical concerns were the initial motivation for much of the technical research discussed in this thesis. However, the experience of composing *Discordianism* also raised other, broader questions about the nature of spatial music composition. While the organization of material in this work can be analysed using Denis Smalley's theory of spectromorphology (see Chapter 8.5) this was not consciously implemented at the time. Certain aspects of the work are quiet effective, such as the spatial distribution of different layers of rhythmic material for example. However, the way in which spatial movement is used to support gestural interplay between the different layers of material is quite inconsistent, as there was no underlying rationale guiding the compositional process in this regard. In addition this argument could perhaps also be applied to the mix of different aesthetics contained within the piece. As I began to examine how other composers had dealt with the use of space as musical parameter, it quickly became apparent that this issue could not be treated in isolation, as the way in which these composers used space is intimately connected to their overall musical philosophy. A study of the use of space in music, rapidly lead to debates which are as old as electronic music itself. Spatial music is in many respects a microcosm of electroacoustic music which can refer to many of the different styles within this aesthetic, but is not tied to any one in particular. The study of the aesthetics of spatial music and the musical use of space as a musical parameter therefore appeared to be a good way to indirectly approach electroacoustic music composition and the performance of electronic music in general.

1.1.3 Why Talk About Spatial Hearing?

> *"While we are free to choose our actions, we are not free to choose the consequences of our actions."*
> Stephen R. Covey

The two, seemingly distinct issues discussed in the previous Section, one quite practical and technical, the other more conceptual and artistic, emerged at the same time from a single piece of music, and have been ever present since. It therefore felt entirely natural to me to approach the study of spatial music via these distinct, yet interrelated topics. The musical use of space exploits a fundamental aspect of hearing, namely our ability to locate sounds in space. However unlike other parameters such as pitch, rhythm and timbre, the movement of sounds through space is usually not an intrinsic aspect of the sound itself, but is instead an illusion created through the careful blending of a number of static sources. A composer of spatial music cannot therefore treat this parameter in the same way as pitch or rhythm, and the technical details of how this illusion is created and maintained are clearly a fundamental and necessary part of the composer's craft. While practical problems remain, there exists a clear empirical approach to solving these issues, and this will allow us to impose some order on space by revealing just how far the illusion can be pushed before it falls apart. This in turn will provide us with a firm basis upon which we can reliably exploit space as a musical parameter. The empirical and systematic examination of these techniques will indicate their particular strengths, capabilities and weaknesses and will in effect function as a form of orchestration guide for spatial, electroacoustic music

1.1.4 The Imaginative Use of Empirical Thinking

The composition of electronic music requires both imagination and technical knowledge and therefore, is inherently both an empirical and a creative process. This is equally true of spatial music, except in this case, it is not sufficient that the composer just knows how to program the relevant software or hardware. As space is so much broader and harder to define than other musical parameters, the composer must also consider how the often illusory nature of this attribute will be perceived by the audience. The first half of this thesis therefore focuses exclusively on the perception of spatial audio and particularly on the weaknesses and limitations of the techniques involved. Although much of this discussion may appear at first to be quite negative, it can also be viewed in a much more positive light. The first act of any composition is the rejection of certain instruments or aesthetics in favour of the instrumentation or style chosen for this particular piece. In the same way, a clear understanding of the limitations and weaknesses of each spatialization technique will provide clear guidance as to the most appropriate technique for a particular piece of spatial music, or indeed a particular type of movement within the piece. Rather than being a negative finding, the discovery of the limitations of these techniques may in fact provide some much needed direction for a composer of spatial music who must choose between a myriad different techniques and applications. In addition, if we can consider this empirical research as form of orchestration guide for spatial music, then statements about the limitations of a particular technique are no more negative than a statement that a violin cannot produce a pitch lower than G3!

The second half of this thesis concentrates on the artistic use of space and the aesthetics of spatial music, however, this discussion is now informed by the results of the empirical analysis conducted earlier. Various works of spatial music are assessed, not in terms of their artistic validity, but in terms of the perceptual effectiveness of the use of space as a musical parameter. The discussion is therefore primarily concerned with examining whether the intentions of the composer in terms of the use of space were realistic or achievable, and not with the artistic merit of the overall musical aesthetic. In addition, a number of original compositions by the author are also presented. Some of these works, particularly early pieces such as *Discordianism*, are good illustrations of the practical and technical problems common to works of spatial music and which are examined in detail earlier in the thesis. However, later works such as *Auto Harp* and *Rise* hopefully illustrate the benefits of this research and a more positive use of the empirical examination of spatialization techniques. The algorithmic control of space using flocking algorithms represents an even tighter merging of empirical and creative processes. In this case, it is the algorithm and not the composer which dictates the movement of sound through space and works such as *Flock* and *Rise* use this relatively new technique to explore the implications of this ceding of control for the compositional process.

1.2 The Research Question

The way in which space is used as a musical parameter is influenced greatly by a number of different, yet inter-related factors. The choice of medium, the means of performance, the choice of material, and the overall musical aesthetic are intertwined, and choices made in one area influence possible choices in other areas. One significant choice is based on technological factors, namely whether to use a collection of different pairs of loudspeakers, or a regular, symmetrical array of matched loudspeakers. The use of a collection of pairs of different loudspeakers, a loudspeaker orchestra, is primarily associated with the style of *acousmatic music,* which has been extensively developed and refined through the years, with a particular amount of work taking place in centres in France, the U. K. and Canada. Composers such as Smalley, Emmerson, Wishart, Barrett *et al* have examined the use of space in this aesthetic, often from the point of view of music analysis. As space is only revealed by the temporal development of the sounds themselves, this analysis has often focussed initially on the nature of the sound object itself, the intrinsic or extrinsic connotations of the sound, before then moving onto the conceptual spatialization and then finally the practical means of spatialization. The relegation of the practical means of spatialization to the end of the discussion is understandable, as the performance, in this case live diffusion to a loudspeaker orchestra, is highly focussed on adapting the work to the particular venue, and preserving the stereo image for the entire audience.

An alternate approach based on multiple tracks of audio played through a symmetrical array of loudspeakers has its origins in a different aesthetic. This approach allows for the creation of much more sophisticated spatial effects than can be achieved with manual diffusion to a disparate orchestra of loudspeakers. However, this approach often neglects to examine how these abstract spatial designs are perceived by the audience. This problem arises from the significant differences between the perception of a single listener, such as the composer in the studio, and a distributed audience in a performance setting, and is often further exacerbated by a lack of intervention (such as that of a diffusionist) during the performance. In general, it is difficult to adapt these works for the acoustic or technical setup of a particular performance space. The question which this thesis therefore attempts to answer is, "what are the limitations, strengths and weaknesses of the most commonly used spatialization techniques, and what does this imply for the performance and composition of spatial music?"

1.3 Aims and Objectives

The main aims of this thesis are to examine the perceptual effectiveness of various works of spatial music in terms of the technical means of spatialization, and also the compositional approach to the use of space as a musical parameter. Particular attention will be paid to the effectiveness of different spatialization techniques in a performance context, and what this implies for compositional strategies which use space as a musical parameter. In this way, the thesis may function as a sort of guide to spatial orchestration, which covers both the technical operation of different spatialization techniques and how this relates to the different aesthetics of spatial music.

1.4 Methodology

This thesis begins with an examination of the perceptual mechanisms related to spatial hearing and a scientific evaluation of the perceptual capabilities of the most commonly used spatialization schemes, namely stereophony, Ambisonics and wavefield synthesis (WFS). The perceptual performance of these systems under the less than ideal conditions typically found in a performance is examined in detail through a series of listening tests carried out by the author. The results of these tests are then incorporated into a meta-analysis of the existing research in this area which summarizes the results of a large number of other listening tests and simulations. The conclusions drawn from this meta-analysis are then used to assess the validity of the various spatial strategies adopted by composers of spatial music such as Charles Ives, Karlheinz Stockhausen, Iannis Xenakis, Denis Smalley and Pierre Boulez. Finally, this research was utilized in the composition of a number of original works of spatial music, the development of a spatial music instrument and an implementation of the Boids flocking algorithm for spatial music composition for the Csound synthesis language.

This particular methodology was adopted so as to ensure that the real technical and perceptual limitations of the practical means of spatialization are fully considered. This emphasis on the limitations of these systems is perhaps somewhat negative, yet if space is to be used effectively as a musical parameter then these practical issues must be fully appreciated. This is particularly true in the case of spatial music performances as while a particular technique may be effective for a single listener, it may be much less effective for a distributed audience.

1.5 Motivation

The use of space in electroacoustic music composition and performance is perhaps one of the most unique aspects of this artistic movement. However, the dissemination of this music via fixed media and domestic audio technology is still a significant challenge. This problem has been further exacerbated by the drastic decline in sales of fixed media music and the increase in online distribution. Yet despite these difficulties, the public's appetite for live musical performances is undiminished, and in fact has significantly expanded over the last decade. The musical and social experience of attending a concert is fundamentally different from listening to music on an mp3 player, on the internet, or at home, and this is particularly true of live performances of spatial electroacoustic music. The experience of listening to music performed with a large and carefully configured loudspeaker array or loudspeaker orchestra provides a unique selling point and an experience which cannot be easily reproduced elsewhere. This aspect of electroacoustic music is the primary motivation for this thesis, which concentrates exclusively on the performance context. Binaural technology and other techniques specifically for the delivery of spatial audio to a single listener will therefore not be considered.

1.6 Outline

This thesis is broadly divided into two parts. The first part deals with auditory perception and the various perceptual mechanisms related to directional hearing. This provides a perceptual basis for an assessment of various spatialization techniques in terms of their perceptual performance in a performance context. The second half of this thesis examines the use of space as a musical parameter through the analysis of various works of spatial music. In each case, the compositional and spatialization strategy is assessed in terms of its perceptual effectiveness, based upon the findings presented in the first half of this thesis.

Chapter Two summarizes the perceptual cues associated with spatial hearing. The perceptual mechanisms which allow a listener to determine the direction and distance of a source signal are presented and the effect of acoustic reflections and reverberance on spatial hearing are discussed. Finally the perception of multiple, concurrent sources is discussed in terms of Bregman's theory of auditory scene analysis (ASA).

Chapter Three examines the technique of stereophony which for the first time allowed for the creation of virtual sources that are not fixed at the physical location of a loudspeaker. The development of this technique is presented from its conception in the 1930s, to modern, multichannel formats such as 5.1. The perceptual basis of the stereophonic principle is also investigated along with various theories of stereophonic localization.

Chapter Four introduces more recently developed spatialization techniques such as Ambisonics and wavefield synthesis (WFS). The development of Ambisonics from Alan Blumlein's work with coincident microphone techniques is presented and the perceptual optimization of ambisonic decoders is discussed. The final part of this chapter addresses the theoretical background of the new technique of WFS and some of the practical issues related to this technically demanding method are discussed.

Chapter Five investigates different approaches to the simulation of distance and the dynamic motion of virtual sources. Well known algorithms by John Chowning and F. R. Moore are discussed and compared with the perceptual mechanisms involved in the localization of a real source, which were discussed earlier in Chapter Two.

Chapter Six analyses the results of a wide variety of tests conducted with the spatialization techniques presented in the preceding two chapters. Various stereophonic systems are first evaluated in terms of localization accuracy and the perception of dynamically moving stereophonic sources. Various ambisonic decoding schemes are then evaluated, particularly in terms of their performance at off-centre listener positions. A number of tests which compare and contrast stereophony and Ambisoncs are then discussed in terms of the perceptual differences between these two techniques. The effect of acoustic reflections and reverberance on source localization is then examined and the results of a series of listening tests conducted by the author are presented. Finally, the results of a number of listening tests carried out with WFS systems are presented and the perceptual performance of this technique is assessed.

Chapter Seven is the opening chapter of the second half of this thesis which focuses on spatial music composition and aesthetics. The history and development of European antiphonal choral music is discussed, and works of acoustic spatial music by Charles Ives and Henry Brant are analysed. The use of spatial distribution to increase the intelligibility of different independent layers of material is also discussed.

Chapter Eight charts the development of electronic music in mid-twentieth century. Various landmark works by Karlheinz Stockausen, Iannis Xenakis and Denis Smalley are analyzed along with two original compositions by the author. The legacy of *Musique Concrète* and *Elektronische Musik* is assessed in terms of their effect on the development of electroacoustic spatial music. Abstract compositional systems such as serialism are assessed in terms of their perceptual effect and the performance practice of diffusion is discussed. Finally, Smalley's theory of spectromorphology and the gestural use of space is presented and its use as a compositional structuring principle is assessed.

Chapter Nine focuses specifically on mixed media electroacoustic music, i.e. music for live instruments and spatial electronic sound. Landmark works by

Stockhausen and Boulez are examined in terms of their technical and artistic approach and the specific difficulties associated with this form of spatial music are discussed.

In Chapter Ten, a number of original works of acoustic, electronic and mixed-media spatial music are presented and analyzed. An original work of choral spatial music is discussed in terms of the spatial choral music discussed previously in Chapter Seven. Two works of electronic spatial music are presented which illustrate the divergent approaches to electronic music composition discussed in Chapter Eight. Finally an original mixed-media composition by the author is analyzed and the spatialization approach adopted for this work is evaluated.

Chapter Eleven examines various musical instruments which can be used for the live performance of spatial music. Various augmented instruments such as the hypercello are introduced, and the use of the hexaphonic guitar as a spatial music instrument is discussed. Finally the technical implementation of a hexaphonic system is presented along with an original electroacoustic composition for the hexaphonic guitar.

Chapter Twelve discusses the use of flocking algorithms such as Boids for sound spatialization and synthesis. Real-time and off-line applications are evaluated along with two original compositions which illustrate these different approaches to the use of flocking algorithms in electroacoustic composition.

2 Spatial Hearing

Traditional musical parameters such as pitch, rhythm and timbre are perceived with a relatively high degree of accuracy. Various studies have shown that a change in pitch of a fraction of a semitone is quite perceptible and our ability to temporally segregate an audio signal is similarly precise. The cross-modal perception of spatial locations is also relatively accurate but is reduced significantly when the visual element is removed. Parameters such as pitch, timbre and rhythm are often directly related to the physical structure of the instrument or the actions of the musician. However, the use of space in music often relies on electronic processes which can only simulate the effect of spatial movement. The ability of these processes to satisfy the various perceptual mechanisms involved would appear to be crucial if this aspect of the work is to be successful. However, before these spatialization processes can be assessed it is first necessary to understand the various perceptual mechanisms involved in normal spatial hearing. By necessity, the mechanisms which allow the location of a real sound to be determined must be first understood, before the illusion of sounds moving in space can be created.

In this thesis, spatial locations will be described using the spherical coordinate system illustrated in Figure 2.1 [Blauert, 1997]. The azimuth angle indicates the angular position in the horizontal plane (with zero degrees being straight ahead), the elevation indicates the vertical angle of incidence in the median plane, and the distance is measured from a point at the centre of the head, directly to the source.

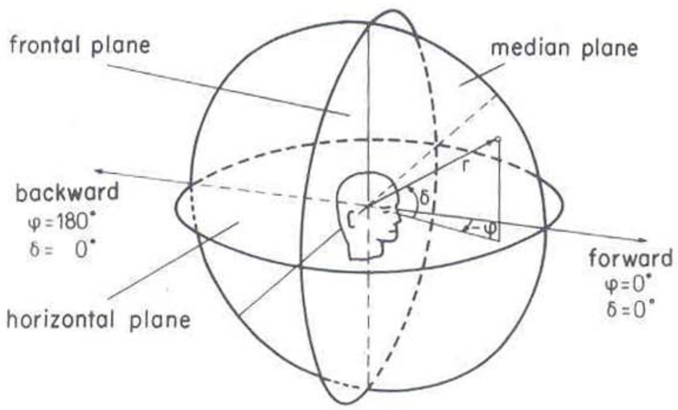

Fig. 2.1 Spherical coordinate system

The study of spatial hearing can be broadly divided into two categories, which typically focus on either certain characteristics of the source signal or of the acoustic environment. A typical audio scene may contain sources that are relatively discrete and localizable (localization refers to the perceived direction and distance of the source signal), however, the same scene may also contain reflections which are diffuse and not easily localizable and provide information about the acoustic environment. This chapter will begin with the more straightforward scenario of a single source under free-field conditions. This implies that the contribution of reflections of the source signal from nearby surfaces is negligible and that the influence of the acoustic environment can be ignored. While this rarely occurs under normal circumstances, outdoor locations such as an open field or mountain top can be considered as approximately equivalent to free-field conditions.

2.1 Directional Hearing

Under ideal conditions it has been shown that the region of most precise spatial hearing lies in the forward direction with frontal hearing having a localization accuracy of between 4.4^0 and 10^0 for most signal types [Blauert, 1997]. Accuracy decreases as the source azimuth moves to the sides, with the localization blur at $\pm 90^0$ being between three to ten times its value in the forward direction. For sources to the rear of the listener, localization blur improves somewhat but is still approximately twice that for frontal sources.

It is reasonable to assume that the following three factors must influence to some extent our ability to localize the position of a sounding object;

- The audio signal produced by the source
- The body, head and ears of the listener
- The acoustic environment containing both the source and the listener

It is also reasonable to assume that these three factors must interact in some fashion to produce an impression in the listener of the spatial location of the source signal.

Consider now a simple example of a laterally displaced sound source positioned in front of a single listener as shown in Figure 2.2. For now the discussion is limited to the horizontal plane and free-field conditions. As the signal produced by the source spreads out and progresses toward the listener the wavefront will first arrive at the right ear, before then diffracting around the head to reach the left ear.

This ability to hear binaurally, i.e. with two ears, is an important aspect of the human auditory system. Localization mechanisms are often distinguished as being either interaural, i.e. related to the differences between the signals at the two ears, or monaural which is related to attributes of the signal that are perceived equally with both ears.

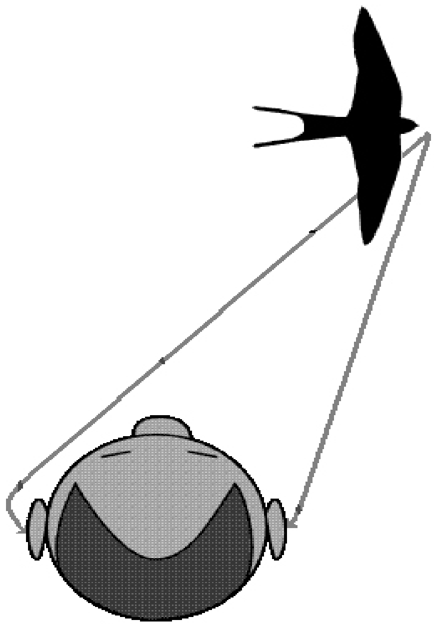

Fig. 2.2 Lateral source example

The preceding example illustrates how a laterally displaced source will result in a time delay between the two signals arriving at the ears. This interaural time delay (ITD) is one of the principal cues used to determine the source azimuth and an extensive amount of experimental work has been carried out to examine its influence on source localization. Various experiments have shown that both the spectral content of the source signal and the attack portion of the signal envelope can produce ITD cues, depending on the context [Blauert, 1997].

If the source in the preceding example consisted of a continuous signal then the interaural delay will result in a phase shift between the two ear signals. This localization cue will be referred to here as an interaural phase shift (IPD) to distinguish it from the more general ITD cue. It should be noted that this phase shift can only be determined by comparing the relative phase of the two ear signals. As the frequency of the source signal increases it eventually produces a phase shift between the two signals that is greater than 180^0. At this point the IPD cue becomes a less reliable indicator of azimuth as it is impossible to determine which signal is leading and which is lagging. So clearly the IPD cue is frequency dependent and in practice has been found to become unreliable at about 700-800Hz, and completely ineffective above 1.6 kHz [Blauert, 1997].

ITD cues can also be derived from the attack portion of the wavefront at the two ears even when phase differences are inconclusive. Experiments have shown that the lowest frequency at which the attack portion of the signal provides useful estimates of azimuth is proportional to the steepness of the attack [Elfner *et al*, 1968]. Generally the ITD cue is the more prominent temporal cue above 1.6kHz while the IPD is more prominent at lower frequencies [Blauert, 1997].

The preceding example illustrated how a laterally displaced source arrives first at the closest ear and then diffracts around the head to reach the other ear. The head shadowing which arises as a result of this diffraction results in an amplitude difference between the two ear signals which is also related to the azimuth of the source. This localization cue is commonly referred to as the interaural level difference (ILD). A signal whose wavelength is comparable to the size of an obstacle will diffract easily around the obstacle. Therefore, when the frequency of the source signal in our example is low enough so that its wavelength is comparable to the spacing of the ears, little head shadowing will occur. Calculating this frequency using an average head diameter of 180mm and a speed of sound in air of 340m/s results in a figure of 1.89kHz [Wiggens, 2004]. Experimental results indicate that ILD values change smoothly with increasing frequency and that the more high frequency content in the signal, the greater the average level difference. In general it has been shown that there is a strong weighting for ILDs with high frequency signals and poor weighting of ILDs with low frequency signals. The converse is true for the ITD. For wideband stimuli, the ITD has been shown to be the dominant localization cue [MacPherson, 2002]. Neither cue is particularly strong in the region of 2kHz.

The preceding discussion describes the interaural localization cues which provide information on the likely azimuth of the source. However, the example shown in Figure 2.3 demonstrates that some other process must also be involved. This example contains the same source as before with a duplicate source positioned symmetrically to the rear. Clearly in this case, both sources would create exactly the same ITD, IPD and ILD cues and that this information would not be sufficient to resolve the two potential source directions. This cone-of-confusion [Begault, 1994] is thought to be resolved using both head movements and monaural localization cues created by the filtering effect of the shoulders and ears.

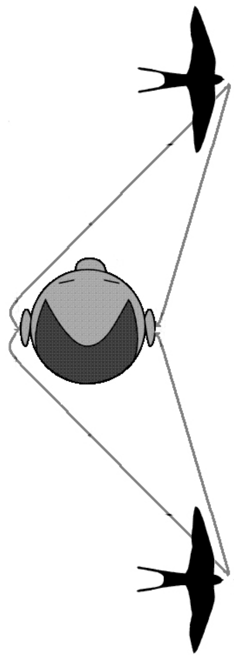

Fig. 2.3 Cone-of-confusion example

The complex shape of the ear pinnae acts as an acoustic filter which introduces time and level differences between the individual spectral components of each ear input signal. As the shape of the upper body and pinnae is irregular and highly idiosyncratic, this filtering effect would change relative to the spatial location of the source signal. The total filtering effect of the pinnae, head and shoulders is often measured and characterized as the Head Related Transfer Function (HRTF). As the ear pinnae are orientated toward the front, the two source signals in the preceding example would be filtered in different ways, allowing the listener to distinguish between the two source locations. It is still not entirely clear as to how the brain distinguishes between the spectral content of the source signal and spectral changes introduced by the HRTF. This could be achieved using slight movements of the head, as this would introduce changes in the HRTF independently of the source signal and, indeed, numerous experiments have shown that head movements allow listeners to determine whether a source is positioned in front or behind or above or below [Blauert, 1997; Spikofski, 2001]. Rotating the head toward the source will alter the various interaural cues in such as way as to resolve the cone of confusion and the ITD and ILD will decrease as the head turns towards a source positioned in front but increase for a source to the rear. In the same way, head movements will result in a change in the filtering effect of the head and pinnae which could also help to resolve the cone of confusion. Of course, head movements can only be effective if the signal duration is long enough to allow for their effect on the different cues to be evaluated.

In a number of experiments with narrowband sinusoidal signals, certain frequencies were found to correlate with specific angles of elevation, independently of the actual source position [Blauert, 1997]. This suggests that HRTF cues not only help to resolve the cone of confusion in the horizontal plane, but also contribute to the localization of elevated sources and, in fact, HRTF cues have been shown to be crucial in determining the position of a source anywhere in the median plane. Studies suggest that spectral cues in both low (below 2kHz) and high frequency (above 5kHz) regions are used to resolve front-back confusion while the prominent spectral cues for the judgement of elevation are derived solely from the high frequency components (above 5kHz) [Asano *et al,* 1990].

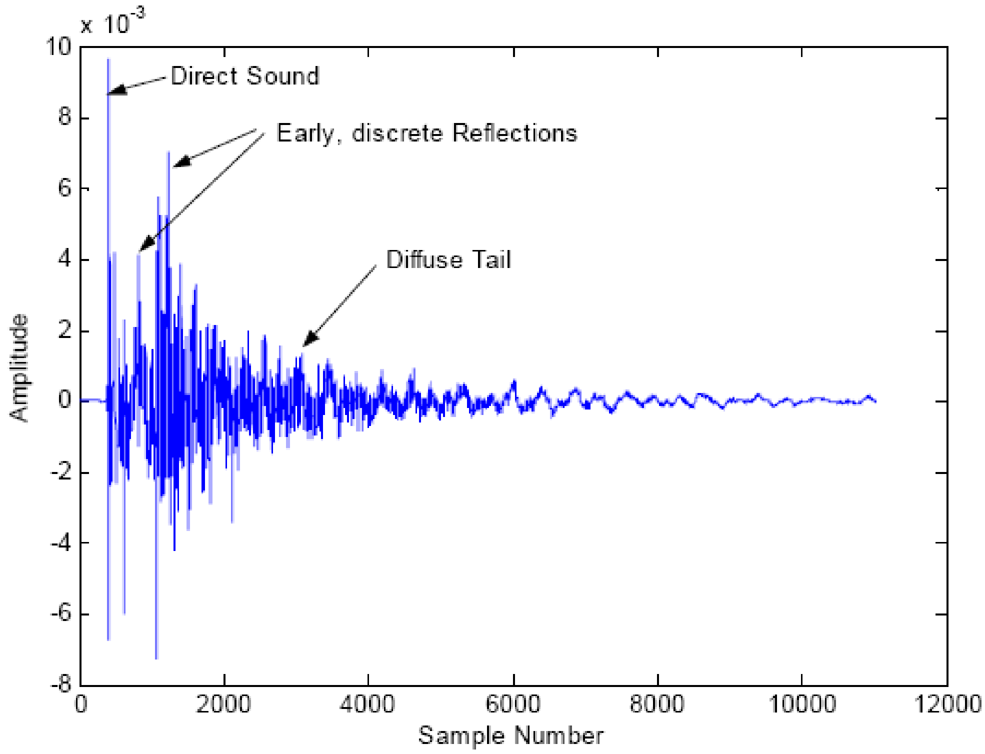

Fig. 2.4 Impulse response of a room with a relatively short reverb time

The preceding discussion on directional hearing has focussed on free field conditions where the effect of the acoustical environment is neglected. However, an environment such as this is quite different from the acoustic spaces in which most music is heard. In an enclosed space, a sounding object will produce a wavefront that will travel directly to the ears and also indirectly via multiple reflections from the floor, walls and ceiling of the space. This indirect signal is generally divided into early reflections which arrive relatively shortly after the direct sound, and the later arriving, more diffuse reflections or reverberance (see Figure 2.4 [Wiggens, 2004]). Despite the presence of these additional wavefronts arriving from multiple different directions it is still quite possible to accurately localize a sound in an enclosed space. [Bates *et al*, 2007a] The law of the first wavefront, or precedence effect is thought to account for this ability and as shall be seen later, is an important consideration in multichannel loudspeaker systems.

2.2 Directional Hearing and Acoustics

As has been seen, in a reverberant environment, sounds reach the ears via several different paths. Although the direct sound is followed by multiple reflections which would be audible in isolation, the first-arriving wavefront dominates many aspects of perception, including source localization [Litovsky *et al*, 1999]. This theory, known as the precedence effect, or the law of the first wavefront, states that when multiple coherent signals are presented at the ears, localization is primarily based on the earliest arriving signal. One study [Begault, 1994] showed that for delays up to approximately 0.6ms the source moved laterally toward the non-delayed side of the listener's head. For delays greater than approximately 0.6ms and less than 35ms the source position remains unchanged but some timbral coloration may occur. Even greater delay times result in two distinct sources as the delayed signal is perceived as a distinct echo. It should be noted that the precise time values at which these perceptual effects occur is highly dependent on the nature of the source signal [Litovsky *et al*, 1999]. Other studies have also shown that the presence of significant early reflections can have an effect on localization accuracy [Hartman *et al*, 1985]. Lateral reflections from side walls were found to be particularly detrimental for azimuthal localization while the effect of early reflections from the floor and ceiling was less conclusive. Hartmann speculates that as the angular direction of these reflections matched that of the source signal, reflections from the floor and ceiling may reinforce horizontal localization [Hartmann *et al*, 1985]. In general, it has been found that signals with strong transient characteristics are localized independently of the room reverberation time, but may depend on the specific room geometry which can result in significant early reflections [Hartmann, 1983].

Early reflections and reverberation also has a significant effect on both the perceived size of the source, and the spatial impression of the acoustic environment. The term spaciousness has been used to refer to both these effects, and other terms such as spatial impression, ambiance, apparent source width, immersion and envelopment are also frequently used, sometimes interchangeably, and sometimes with more specific definitions. A strong correlation has been found between the degree of coherence between the two ear signals and the lateral component of the ear signals, and this is though to influence the spatial impression of both the source and environment [Blauert, 1997; Plenge *et al*, 1975]. Early studies of spaciousness described this characteristic in terms of the lateral and frontal components of the ear signals [Barron *et al*, 1981]. However, what is meant by spaciousness is different depending on whether this lateral component is derived from early reflections alone, or from both early reflections and later reverberation. The addition of lateral early reflections results in a change in the perceived size of the source and the apparent source width (ASW) is generally used to refer to this source-specific measure [Beranek, 1996]. Early reflected energy arriving within approximately 80 ms of the direct sound results in an increased ASW and the extent of this broadening effect depends upon the ratio of the total energy to the energy of the lateral component [Barron *et al*, 1981]. Blauert introduced the term "locatedness" as a measure of the degree to which an auditory event can be said to be clearly in a particular location [Blauert, 1997] and this is clearly related to ASW. Localization refers only to the perceived direction and does not therefore directly relate to the locatedness or ASW, although clearly the source direction will be difficult to precisely determine in the case of a very large ASW. The important distinction between these two measures will be discussed in more detail in Chapter Six.

Later arriving reverberation alters the spatial impression in a different way which is primarily related to the acoustic environment. The term spaciousness is often used in this case, as opposed to ASW which is primarily related to the spatial impression of the source. The terms spaciousness and envelopment are often considered equivalent, although occasionally slightly different definitions are given. When distinguished, spaciousness is described as the sense of open space in which the source is located while envelopment refers to the sense of immersion and involvement in a reverberant sound field which fully surrounds the listener [Rumsey, 2001].

Numerous studies have shown that envelopment and spaciousness are generally desirable qualities in a concert hall [Schroeder *et al*, 1974; Barron *et al*, 1981; Beranek, 1996] and many concert hall designs have attempted to increase the amount of lateral reflected energy directed to the seating area for this reason. The acoustician David Griesinger has suggested that the importance of the distinction between early arriving lateral reflections and later reverberation is often overlooked in this context [Griesinger, 2009]. It has already been shown that early arriving reflections reduce localization accuracy and Griesinger suggests that an increase in lateral energy of this sort will negatively impact clarity. This clearly contradicts previous studies which have stressed the importance of ASW. Griesinger suggests that the direct sound and later lateral reverberation should be emphasized to improve clarity (which corresponds to improved localization) and also spaciousness (meaning the desirable spatial characteristics of the hall).

Rumsey has similarly pointed out the difference between acoustic research, which suggests that large values of ASW are preferable, and studies of spatialization techniques which emphasize localization accuracy [Rumsey, 1998]. Griesinger has noted a similar contradiction between the differing levels of reverberation typically found in performances and recordings of classical music [Griesinger, 2009]. A clear distinction can be made between the three-dimensional spatial experience of a live instrumental performance in a concert hall and a two-channel stereo reproduction which must reproduce both the direct and reflected sound from the front. Improved localization accuracy is perhaps desirable in the latter case in order to distinguish sources in the foreground from background reverberation [Rumsey, 1998]. However, the situation is much more complicated in the case of electronic spatial music performances which often utilize multi-channel loudspeaker arrays and electronic spatialization techniques within a concert hall. The distinction between ASW, spaciousness and envelopment introduced earlier may also be hard to maintain in this situation, as a multichannel reproduction of a single source from multiple positions around the audience will provide some sense of envelopment, but by the direct sound and not the reverberant field. In this situation, the sense of spaciousness may be associated with the source rather than the acoustic environment, and the distinction between these terms becomes harder to define. In addition, it is hard to predict the effect of a lateral source on perceptual aspects such as ASW or spaciousness.

The preceding discussion indicates the significant effect of acoustic reflections on the perception of an auditory source. In addition, reverberation provides a significant amount of information regarding the size and composition of the environment within which the source is situated. The degree of attenuation of acoustic reflections provides an indication of the nature of the reflecting surfaces while the time and duration of the diffuse late reverberation can indicate the dimensions of the space. Room reflections and reverberation can also provide information on another highly important aspect of spatial hearing, namely, the distance of the source from the listener and this will be examined in more detail in the next section

2.3 Distance Hearing & Moving Sources

The most obvious clue to the distance of a sounding object is its amplitude. It has been shown that for a constant point source in a free sound field, the sound pressure level falls of by six decibels with every doubling of the distance [Blauert, 1997]. However, amplitude can only serve as an absolute measure of distance if the listener is familiar with the source signal as the amplitude of an unfamiliar or synthesized sound can only provide a relative estimate of a change in distance [Mershon, 1997; Sheeline, 1983]. For example, it is easy to determine whether someone is speaking quietly nearby or shouting loudly far away because of the familiar and characteristic dynamics of the human voice, but when the source signal is unfamiliar this distinction cannot be made. Various studies under free-field conditions have found that for a source with constant amplitude, there is no relationship between the actual and perceived distance [Nielsen, 1993; Mershon *et al*, 1975], indicating that, in this case, the signal amplitude can only indicate how the relative distance is changing. For very close sources (< 1m), it appears that the auditory system also uses additional binaural localization cues to estimate the source distance. Brungart analysed a number of HRTF measurements which showed that very near sources result in substantial changes in ILD [Brungart *et al*, 1999], although the ITD remains largely unchanged.

It has long been known that the perception of distance is also influenced by the effect of acoustic reflections in the listening environment. The level of the direct and early reflected sound will change substantially as the distance from the source to the listener changes. However, the level of diffuse reverberation is largely independent of the position of the listener in the room. Therefore, as the source distance increases, the direct sound will decrease while the reverberant sound remains constant. Beyond a certain distance, the reverberant signal level will be greater than the direct signal level, and the perceived distance becomes fixed and independent of the actual source distance. This critical distance is indicated in Figure 2.5 [Howard *et al*, 1996]. Various studies have shown that in reverberant rooms, the perceived distance of a real source is independent of the source level [Nielsen, 1993], which suggests that the ratio between the direct and reverberant signals, the D/R ratio, is a significant distance cue in real rooms. This theory was first proposed in the 1960s and this simple ratio is still commonly used by sound engineers and producers to control the depth of different sources in two-channel stereo mixes.

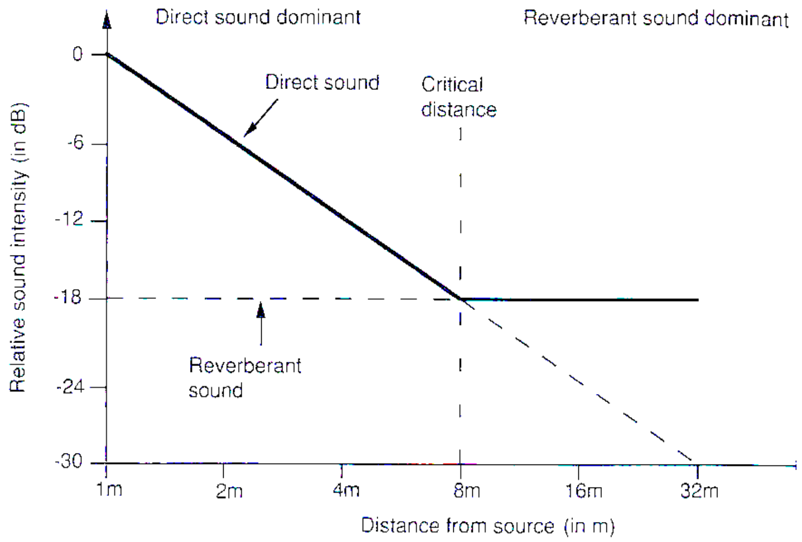

Fig. 2.5 Source distance v sound intensity

The D/R ratio can provide a relative sense of distance but this simple ratio ignores the fine spatial and temporal structure of the reflected indirect signals. Experiments by Kendall reported that a strong impression of distance was perceived when listening to dry test signals augmented solely with a limited number of artificial early reflections, even when these reflections were restricted to those which followed the direct signal by 33ms or less [Kendall *et al*, 1984]. Michelsen carried out a similar test which also found that a better distinction of distance was achieved when simulated early reflections were added instead of solely diffuse reverberation [Michelsen *et al*, 1997]. Neher investigated the perceptual effect of different early reflection patterns and found that listeners were unable to distinguish between an early reflection pattern comprised of accurately panned reflections, and one that was physically identical except that each reflection was simply reproduced by the nearest available loudspeaker. This suggests that although spatial differences in the early reflections pattern are perceptually salient, the actual angles of incidence of reflections may not be crucial [Neher, 2004].

Michael Gerzon presented a similar model of distance hearing, based on a theory originally proposed by Peter Craven [Gerzon, 1992b]. The Craven hypothesis assumes that the apparent distance of sounds is derived from the relationship between the relative time delay and amplitude of the early reflections and the direct signal. Gerzon and others have suggested that closely-spaced or coincident microphone techniques having a substantially omnidirectional total energy response will reproduce the absolute source distance better than microphones with a more directional response [Gerzon, 1992b; Theile, 1991]. Gerzon also points out that although it is now known that the simple direct/reverberant ratio does not provide an absolute measure of distance, it is still a useful subsidiary cue for relative distance, and is thus preferably made consistent with the apparent distance [Gerzon, 1992b].

In general it has been found that the perceived distance of a sound source in a room is compressed, as it increases virtually linearly with source distance at short range, but converges to a certain limit when the source distance is increased beyond the critical distance [Mershon *et al*, 1975; Nielsen, 1993]. There is therefore a non-linear relationship between the perceived and actual source distance. Bronkhurst suggests that this non-linearity arises not only due to the different mechanisms involved for different source distances, but also because the auditory system is not always able to accurately separate the direct and reflected signals [Bronkhurst, 2002]. In a number of listening tests, the effect of early reflections on perceived source distance was assessed in terms of the angle of incidence of the reflected sound. When the lateral walls were made completely absorbent, the source was perceived to be close to the head, virtually independently of the actual source distance. When lateral reflections were introduced, the perceived distance more closely matched the actual source distance, although larger distances were still underestimated [Bronkhurst, 2002]. This suggests that the direct to reverberant ratio is estimated by the auditory system using directional binaural cues to separate the direct and reverberant signals, although further tests are needed to confirm the validity of this hypothesis.

While clearly the amplitude of the source signal and the specific relationship between the direct and indirect signals have been shown to be dominant cues in distance perception, other secondary cues also provide some indication of relative distance. In general, an increase in source distance results in a reduction of the high frequency spectral content of the source signal due to the effect of air absorption. This occurs at large distances outdoors, but also in rooms due the absorptive nature of the boundary surfaces and the large overall distances travelled by the indirect signals as they reflect around the room.

The perceived shift in frequency due to the movement of a source relative to the listener, i.e. the Doppler effect, also provides an indication of the relative motion of the source.

2.3.1 Summary of Spatial Hearing

The preceding section summarized the different localization mechanisms which are used by the auditory system to determine the direction and distance of a source signal. It is still unclear, however, as to how these multiple, different and potentially conflicting localization cues are resolved into a single distinct spatial impression. Most theories of auditory localization now propose that the perceived source location is the one that satisfies as many of the localization cues as possible. When multiple conflicting localization cues are present, a simple majority decision is used to determine the location of the sounding object. Most real sources produce a complex signal which will produce localization cues that support each other in the frequency ranges at which they dominate. In addition, the corresponding visual component of a real source will also support the direction suggested by the auditory senses. Broadband signals are able to satisfy more localization mechanisms than narrowband signals and are therefore, in general, easier to locate. In addition the presence of onset transients greatly increases localization accuracy. Narrowband signals such as sine tones, particularly in the 2kHz region, are difficult to localize even under ideal conditions as no localization mechanism is particularly effective in this region. Early reflections and reverberation can also have a significant effect on the perceived size of the source, which can result in a corresponding decrease in localization accuracy.

2.4 Spatial Hearing with Multiple Sources

In a complex auditory environment, sounds generated by numerous different sources at different spatial locations all arrive together at the ears. Yet, most listeners can readily segregate the audio scene into its multiple distinct components. This process is often referred to as the cocktail party effect, where a listener can consciously focus their attention on any one of a number of overlapping and concurrent conversations. Albert Bregman attempted to explain this phenomenon with the theory of auditory scene analysis (ASA) which was informed by Gestalt psychology and describes this process whereby an audio scene is segregated into multiple distinct streams [Bregman, 1990]. Auditory scene analysis describes the relationship between the different elements in an auditory scene in terms of the Gestalt principles of perceptual grouping, namely;

- *The principle of proximity*: elements which are positioned close together in space or time were probably generated by the same event
- *The principle of similarity*: sounds with a similar timbre or frequency probably belong to the same event
- *The principle of good continuation and completion*: sounds generated by the same event tend to be continuous and follow each other
- *The principle of common fate*: sounds with similar frequency, dynamic or rhythmic trajectories probably originated from the same event.

These grouping principles were described by Bregman as *primitive segregation*, as these perceptual processes are involuntary and innate rather than learned. More refined musical knowledge, and prior experience with different sounds is structured in units referred to as *schemas* and these are voluntarily employed when attention is being paid to a specific sound or piece of music. *Schemas* therefore operate over a longer time-scale than primitive segregation as it involves prior experiences and also expectations about the sounds to come. It would therefore appear that auditory perception depends upon two concurrent perceptual processes in two overlapping timescales. Primitive processes firstly perform a quick interpretation of the composite ear signals, dividing it into separate streams, while at the same time longer *schema* processes interpret each stream as it changes in terms of prior experience and expectation of its future state. Interestingly, a similar two stage process is thought to operate for the visual senses, as first uncovered in the 1959 paper "Receptive fields of single neurons in the cat's striate cortex," by Hubel and Wiesel [Hubel *et al*, 1959]. Further research in machine vision has suggested that when the visual input enters the brain from the eyes, it is immediately sent through two separate neurological pathways [Goodale *et al*, 1992]. The fast path quickly transmits a rough, blurred outline of the image to the frontal cortex while the second path performs a slower (the slow image arrives in the prefrontal cortex about 50ms after the fast image) analysis of the image, using prior experience and knowledge to fill out and refine the rather crude initial impression.

It has been shown that spatial auditory cues are not a dominant factor in determining the number of competing sound sources [Bregman, 1990]. However, other studies have shown that spatial hearing is highly important for the intelligibility of multiple, simultaneously presented speech signals [Shinn-Cunningham, 2003; Best, 2004] and that our ability to segregate an audio scene into multiple streams strongly influences our perception of fundamental musical constructs such as melody and rhythm [Bregman, 1990]. The concept of spatial masking, in which a listener’s ability to detect and understand the content of multiple signals is improved when the signals are spatially separated is highly important in spatial music. As shall be seen later in Chapter Seven, composers such as Charles Ives, Henry Brant and Karlheinz Stockhausen regularly used multiple, spatially separated musicians to reinforce the inherent polyphony in the musical work. Of course spatial masking can also be used in the opposite way, to deliberately undermine the individuality of each source in an attempt to create a single sound mass using multiple musicians positioned together on stage. Works such as *Atmospheres* by Gyorgy Ligeti or *Metastasis* by Iannis Xenakis are clear examples of orchestral works which deliberately utilise spatial masking in this way. These two different uses of spatial masking can be further expanded in purely electronic works, where the various spectral components of a signal can be heightened through the dynamic spatial separation of the various components. The electroacoustic composer Natasha Barrett illustrates this process with a number of audio examples in a paper on spatio-musical composition strategies [Barrett, 1990]. In the first example, eight continuous sounds with a significant degree of temporal and textural overlap are played simultaneously at static points in space while in the second example each sound is dynamically moved in space. Barrett suggests that only five distinct sources are perceivable when positioned at static spatial locations but that this number increases when each source is provided with individual and dynamic spatial trajectories.

2.4.1 The Limits of Auditory Perception

Although ASA clearly suggests that spatial cues are an important factor in auditory perception, Bregman does include certain qualifications such as the frequency dependence of the cocktail effect and the detrimental effect of reverberation and reflections. Harley developed the table shown in Figure 2.6 which summarizes the interaction between the various spatial cues and stream segregation and illustrates the conflicts which may potentially occur [Harley, 1998a].

1	Difference and separation	Different sounds from different locations belong to separate sound complexes.
2	Confluence of cues	Spatial difference increases the amount of segregation arising from other cues (spectral, durational, onset/offset).
3	Innate quality	The tendency to segregate sounds from different spatial locations belongs with primitive (innate) principles of auditory analysis.
4	The role of attention	'Cocktail party effect' (selectively focusing on sounds from one direction) suggests the importance of the conscious focus of attention; the effect is strongest around 1 kHz, worsening below 400 Hz.
5	Reasons for imprecise localisation	Sound localisation is imprecise in reverberant environments, in the presence of strong echoes, and because of the bending of sounds around obstacles.
6	Conflict of cues	If two sound events occur at the same point in space, their integration may depend on other cues (i.e. spectrum, envelope, common fate).
7	Spatial and visual cues	Spatial cues are weaker than visual ones. In the 'ventriloquism effect', sounds originating as far as 30 deg away from their apparent, visible source will be integrated with this source.
8	World-structure cues	There is a tendency in audition to expect sounds of similar spectral characteristics and close in time to be related.

Fig. 2.6 Spatial cues and stream segregation

The limits of auditory perception can also be exceeded by electronic processes such as, for example, in the speed of movement of a particular source. Blauert reported that a cyclical lateral movement must take at least 172ms to occur if the trajectory is to be accurately followed while 233ms is required for a front-rear movement. Sounds following a circular path around a listener that exceed these limits will at first appear to oscillate from left and right before stabilising at a central position at even faster speeds [Blauert, 1997]. This effect was deliberately employed by the composer Karlheinz Stockhausen in his eight-channel composition *Cosmic Pulses* [Sonoloco, 2007]. The technical implementation of this work was carried out by Joachim Haas and Gregorio Garcia Karman at the Experimental studio in Freiburg between December 2006 and April 2007. Karman describes the perceptual effect of the OKTEG (Oktophonic effect generator) system developed for this piece to implement very high speed rotational effects, as follows;

> *"Like in the Rotationsmühle –a device used in the spherical auditorium at the World's Fair in Osaka and later implemented as output stage of the Klangwandler - the OKTEG provides the performer with manual control of rotation velocity, and different routings are accomplished by means of matrix programs. The Rotationstisch, first used as a spatialization instrument in KONTAKTE, was later further developed for exploring the artefacts, which appeared at very high rotation speeds. Following this idea, the OKTEG provides with sample accurate trajectories and arbitrary high rotation speeds, assisting the exploration of a continuum linking space and timbre. When sound trajectories get close to the upper velocity range of 16 rot/sec in the composition of COSMIC PULSES, the perception of movement is gradually transformed into a diffuse and vibrating spatial quality. Higher rotation frequencies manifest themselves as audible modulation effects."* [Sonoloco, 2007]

2.4.2 Spatial Hearing and Virtual Sources

It has been shown that our ability to localize the position of a sounding object is determined by the interaction of:

- The audio signal produced by the source
- The body, head and ears of the listener
- The acoustic environment containing both the source and the listener

The preceding chapter summarized the effect of these parameters on our perception of the direction and location of a real sounding object. The composition of spatial music obviously involves the direct manipulation of the first parameter, however, the latter two parameters can only be dealt with indirectly, if at all. Some spatial music, particularly acoustic music for distributed musicians, can be understood relatively simply in terms of the localization mechanisms discussed in this chapter. However, most electronic spatial music does not confine itself to multiple independent signals produced by individual loudspeakers. Instead, different electronic processes are used to position and dynamically move virtual sources around the listening area using multiple coherent loudspeaker signals.

2.4.3 Spatial Audio Techniques

A great many techniques have been developed over the past century for the production of spatial audio. However, in general, the approach taken consists of either:

- The manipulation of level and/or time differences in pairs or multiple pairs of loudspeakers.
- The reconstruction of a sound field over a listening area using a loudspeaker array.
- The reconstruction of the ear signals using headphones or highly localized loudspeaker signals.

The first approach of manipulating either phase/time or more usually level differences between pairs of loudspeakers is often referred to as stereophony. The production of ITD and ILD cues through the manipulation of these factors can be achieved both acoustically through the use of different microphone arrays, and through electronic processing techniques such as amplitude panning. Stereophony originally referred to any method of reproducing a sound field using a number of loudspeakers, but is now generally used to refer specifically to techniques based on the manipulation of level and/or time differences in pairs or multiple pairs of loudspeakers, such as in two-channel stereo and 5.1 surround sound.

Ambisonics and Wavefield Synthesis are two techniques which attempt to reconstruct a sound field within a listening area using loudspeaker arrays. Ambisonics is a complete set of techniques for recording, manipulating and synthesizing artificial sound fields [Malham *et al*, 1995] which has been regularly used in spatial music and theatre for the past three decades. While never a commercial success, Ambisonics has proved enduringly popular for spatial music presentations for various reasons, such as its independence from a specific loudspeaker configuration and its elegant theoretical construction which is based on spherical harmonics.

Wavefield Synthesis (WFS) is a more recently developed technique which can be considered as an acoustical equivalent to holography, or holophony [Berkhout, 1998]. The technique uses large numbers of loudspeakers arranged in linear arrays and can theoretically recreate a sound field over a much larger listening area than is possible with other sound field reconstruction techniques such as Ambisonics.

The third approach uses HRTF data to either record or synthesize spatial auditory cues. This binaural approach is highly applicable for a single listener as it requires a strict separation of the two ear signals, such as when listening with headphones. However, it is much more difficult to extend this approach to large groups of listeners and so will not be covered in this thesis.

Apart from potentially WFS, these techniques can really only recreate the directional perceptual cues. The simulation of distance is often achieved using additional processes which will be discussed later.

3 Stereophony

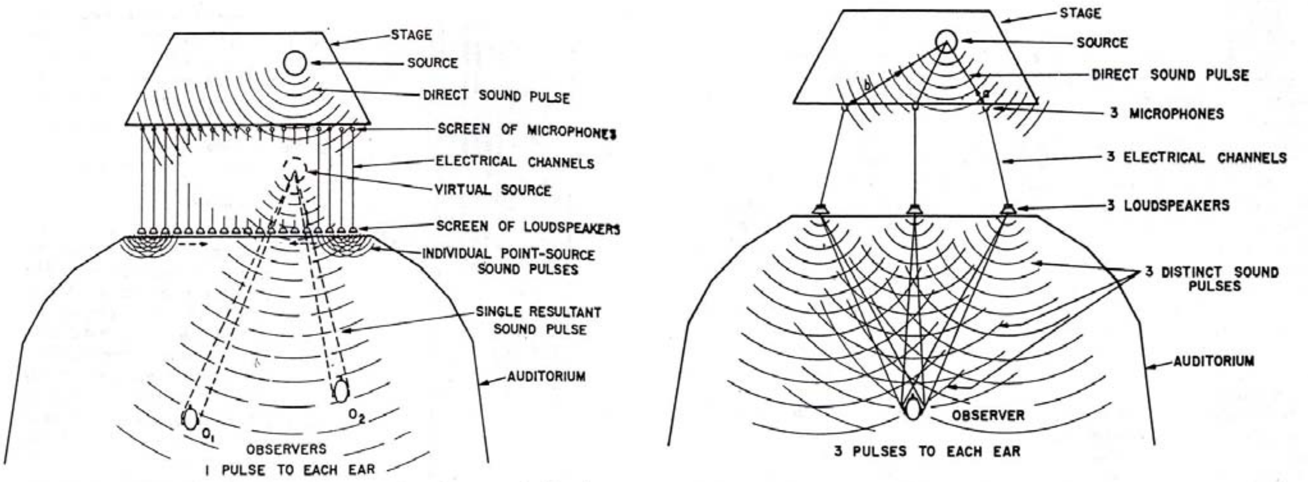

Fig. 3.1 Bell Labs stereophony, proposed (left) and implemented (right)

The earliest work on stereophony was carried out independently by both Bell Laboratories in the United States and Alan Blumlein at EMI in the UK in the early nineteen thirties. The approach adopted by Bell labs was based on the concept of an acoustic curtain [Steinberg *et al*, 1934], namely that a sound source recorded by a large number of equally spaced microphones could then be reproduced using a matching curtain of loudspeakers (Figure 3.1 left). In theory, the source wavefront is sampled by the microphone array and then reconstructed using the loudspeaker array. In practice, this approach had to use a reduced number of channels, so a system was developed using three matching spaced omni-directional microphones and three loudspeakers placed in a front-left, centre and front-right arrangement (Figure 3.1b). This approach was problematic however, as the reduction in channels distorted the wavefront and audible echoes sometimes occurred due to the phenomenon of spatial aliasing (see Section 3.2.2). Spaced microphone techniques such as this capture the different onset arrival times of high frequency transients, and so capture the ITD localization cues present in the original signal. However, this also makes it difficult to process the audio afterward as unpredictable time differences are fixed in the recording.

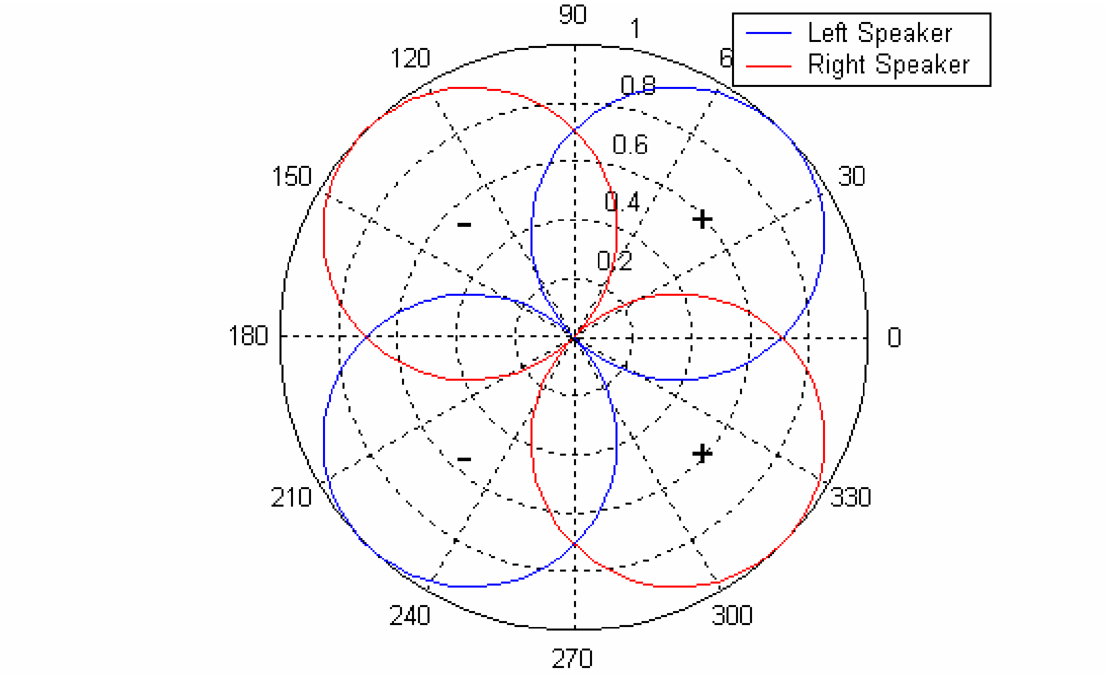

Fig. 3.2 Blumlein's coincident microphone arrangement

At the same time, Alan Blumlein was developing various alternative arrangements, such as the two coincident microphones with figure-of-8 directivity patterns shown in Figure 3.2 [Wiggens, 2004]. This coincident microphone arrangement records level differences which vary with the angular position of the source, but as the microphones are coincident, time differences are not captured. However, Blumlein realized that the resulting level differences would in fact result in an IPD at low frequencies due to the unavoidable cross-talk between the loudspeakers. To illustrate this, consider two sources radiating a low frequency signal with no time difference, but with a greater amplitude signal radiating from the left loudspeaker (Figure 3.3). The listener will receive at his left ear the louder signal from the left loudspeaker, combined with the quieter signal from the right loudspeaker, which is now delayed due to the greater distance travelled. The sum of these two wavefronts will be a phase-shifted and amplified version of the louder wavefront. A similar and inverse summing process occurs at the right ear and Blumlein realized that the resulting difference in phase between the two ear signals will produce an interaural time cue at low frequencies that is proportional to the amplitude difference between the loudspeaker signals. In turn, at higher frequency ranges, head-shadowing acts as a greater obstacle to the two wavefronts, so the amplitude differences produce an effective ILD cue. Stereophony, therefore, to some extent resembles natural hearing, as it produces IPD and ILD cues in the frequency ranges at which these localization cues are most effective. It therefore uses the unavoidable cross-talk between the loudspeakers as an advantage, as this cross-talk produces an IPD which is related to the original source direction. Critics of this approach of summing localization argue that level differences alone cannot produce the ITD cues necessary for correct localization of onset transients [Thiele, 1980]. However, subjective listening tests have shown that this is not the case and that transients can be clearly localized in Blumlein stereo recordings [Rumsey, 2001]. In addition, this approach allows for the post-processing of the stereo image by adjusting the combination of the two microphone signals. More recently alternative microphone arrangements such as ORTF or the Decca tree have been developed which represent a trade-off between the two approaches and reduce the conflicting ITD cues that arise for transient and steady-state signals with purely coincident techniques.

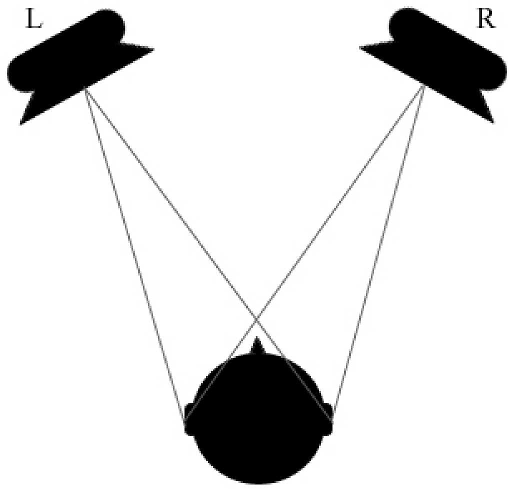

Fig. 3.3 Standard stereophonic arrangement

These microphone techniques can of course also be adapted to artificially position a monophonic recording in a stereo field. The introduction of time differences to a monophonic signal routed to two loudspeakers can be used to position or pan the signal between the loudspeakers. This approach, however, can introduce contradictory phantom image positions for the transient and steady-state parts of the signal [Martin *et al*, 1999a], as additional phase differences are introduced by the summing effect of the loudspeaker cross-talk. In addition, comb filtering can occur when both channels are summed to a monophonic signal and the perceived source position is highly dependent on the position and orientation of the listener [Rumsey, 2001]. Amplitude panning introduces level differences by simply weighting the signal routed to each loudspeaker and this technique is quite effective when used with a symmetrical pair of loudspeakers in front of a single, centrally positioned listener, with an optimal separation angle of $\pm 30^{o}$. Amplitude panning can be considered as a simplification of Blumlein's coincident microphone technique shown in Fig. 3.2. With this arrangement, a signal in the front left quadrant will arrive at the maximum of the blue microphone response characteristic and at the null point of the red microphone. Amplitude panning simplifies this idea so that a signal panned hard left will only be produced by the left-most loudspeaker, and vice versa, while a signal panned to the centre will be created as a phantom image by both loudspeakers. This has the result that a slight yet perceptible change in timbre occurs when a signal is panned from a loudspeaker position to a point in between.

The specific implementation of stereophony for two loudspeakers, i.e. two-channel stereo, is by far the most commonly used audio format in the world today. However, as this format only utilises a pair of front loudspeakers, it must necessarily reproduce both the direct source signal and reverberation from the front. One of the earliest formal extensions of this method to more than two channels is the Quadraphonic system, which is summarized in the next section.

3.1 Quadraphonic Sound

Quadraphonic sound was one of the earliest attempts to develop a standardized system for the presentation of spatial audio. Introduced in the early seventies, the system used four loudspeakers positioned equidistantly and symmetrically at the four corner points of the listening space. Quadraphonic sound was a commercial failure for various reasons such as a number of incompatible delivery formats and issues with multiplexing four discrete channels into two-channel media. However, the Quadraphonic format continued to be used for performances of spatial music for many years and in fact, four equally spaced loudspeakers were used in electroacoustic music long before the development of Quadrophonics as a commercial format. Karlheinz Stockhausen's landmark work Kontakte was written in 1960 for four symmetrical loudspeakers, which were however, unlike Quadraphonics, placed at the front, left, right and rear positions.

The standard Quadraphonic loudspeaker arrangement exhibits a number of deficiencies when creating phantom images, even for a single listener. The extended angular separation of ±45° between each loudspeaker pair degrades the stereo image significantly, as small changes in levels result in large changes in the perceived source position [Thiele *et al*, 1976]. In addition, lateral and, to a lesser extent, rear pairs of loudspeakers have been shown to be incapable of producing a reliable phantom image as the perceived direction abruptly shifts from one loudspeaker to the next (Figure 3.4 [Ratliff, 1974]). This result is perhaps unsurprising as it has been shown that stereophony depends upon the crosstalk between the loudspeaker signals at the two ears to reproduce the correct IPD. In this situation, both loudspeaker signals will arrive at the nearest ear at the same time and the stereophonic image will be distorted.

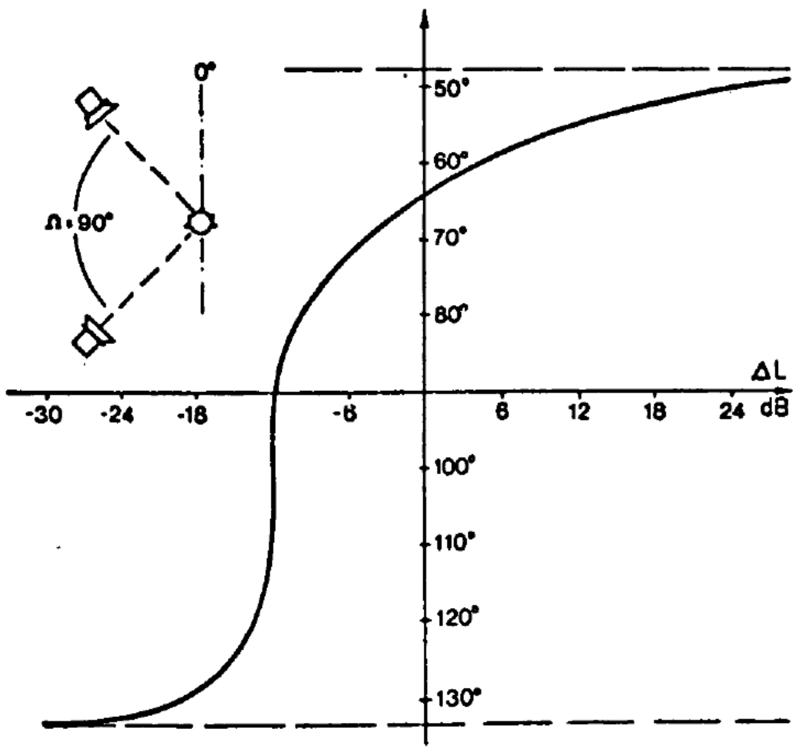

Fig. 3.4 Lateral phantom source direction versus ILD for a quadraphonic layout

The decision to adopt a four-channel format, such as Quadraphonics, was almost certainly influenced by the difficulties in storing more than four channels on analogue media and the prohibitive cost at that time of large numbers of amplifiers and loudspeakers. Today, the storage and playback of many channels of audio can be readily and inexpensively achieved with digital technology and the cost of amplifiers and loudspeakers has also decreased significantly. These factors, combined with the inherent perceptual defects of Quadraphonics, have resulted in the development of similar formats which utilise a greater number of loudspeakers.

3.2 Cinema Surround Sound and 5.1

5.1 surround sound is a spatial audio format which is commonly used in cinemas and also for music and films released for the domestic market on Digital Versatile Disk (DVD) media. The loudspeaker layout and channel configuration for 5.1 surround is specified in ITU-R BS.775 and shown in Figure 3.5. The frontal bias of this arrangement indicates its origins in cinema and explains its alternative name of 3-2 stereo. 5.1 is primarily intended to support a frontal visual image while maintaining backward compatibility with traditional two-channel stereo. Hence, the front left and right loudspeakers maintain an angular separation of +/- 60^o, even though the presence of the centre loudspeaker could enable the creation of a wider frontal image. Similarly, the rear/side channels are only intended to be used to produce reverberation and ambience in support of the primary frontal image. The final ".1" or low frequency effects (LFE) channel is used for sub-bass effects, although in many systems this loudspeaker is used to produce all the low frequencies in conjunction with five smaller loudspeakers.

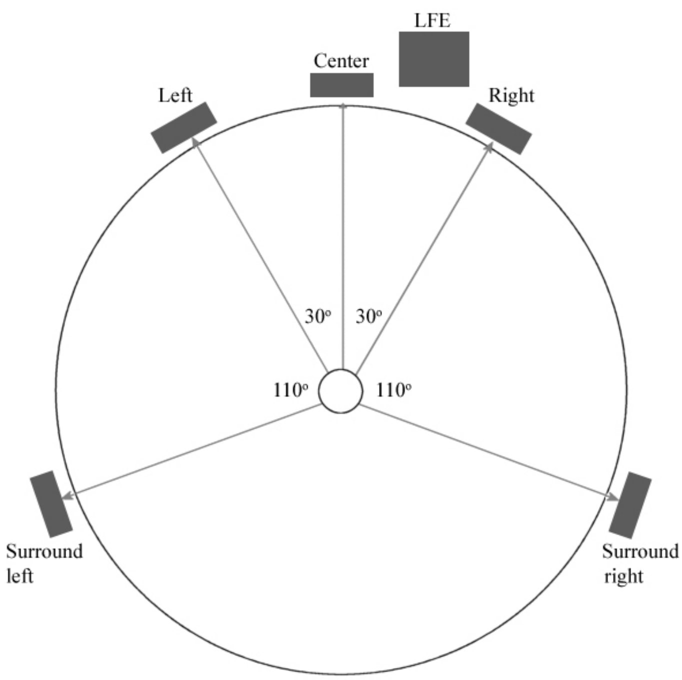

Fig. 3.5 ITU 5.1 loudspeaker arrangement

While 5.1 surround can be highly effective when used for its intended purpose as support for a frontal visual image, it is much less suitable for the presentation of spatial music to a distributed audience. The front three loudspeakers allow for accurate frontal images to be created in this direction but the problems with lateral and rear virtual images discussed in the previous section are magnified. Of course the commercial attractiveness of 5.1 cannot be understated. The means of production and delivery are by now well-established and it is unlikely that any new format will replace it in the near future. Some research has reported good results using 5.1 as a delivery format for playback with an alternative loudspeaker arrangement and encoding such as Ambisonics [Gerzon, 1992c; Bamford, 1995; Wiggins, 2004].

3.3 Multichannel Amplitude Panning Techniques

Numerous extensions of the basic stereo amplitude panning method have been developed which can be utilized for non-standardized loudspeaker arrangements such as a regular eight-channel array. Vector Base Amplitude Panning (VBAP) is a popular technique developed by Ville Pulkki [Pulkki, 1997]. This vector-based reformulation of the amplitude panning method can be used to extend the basic stereophonic principle to an arbitrary number of loudspeakers. Once the available loudspeaker layout is defined, virtual sources are positioned by simply specifying the source azimuth. If a virtual source is panned to the same direction as any of the loudspeakers, then only that loudspeaker will be used while sources panned to a point between two loudspeakers will be produced using this loudspeaker pair and the tangent panning law. VBAP also includes a spread function which ensures that the same number of loudspeakers is used to produce the image regardless of the specified position. This is highly advantageous for dynamically moving sources as it reduces the timbral shift that occurs as the signal moves from a position at a loudspeaker, to a position between a loudspeaker pair [Pulkki, 1999].

Spat is a real-time modular spatial-sound-processing system developed by IRCAM and Espaces Nouveaux for the Max MSP environment [Jot, 1999]. The system allows for the positioning and reverberation of audio sources in three dimensions using a high level control interface based on a number of perceptual parameters. The design of Spat is largely based on the spatial processing algorithms developed by Chowning and Moore in the seventies and eighties [Chowning, 1971;

Moore, 1983] which will be discussed later in this Chapter. The supplied output module can be configured for reproduction using standard two-channel stereo, discrete intensity panning over various multichannel loudspeaker configurations (up to a maximum of eight), Ambisonics B-format encoding, or binaural encoding for reproduction over headphones.

3.4 Theories of Stereophonic Localization

After the development of stereophony, a number of theorists attempted to develop a more complete mathematical framework to describe and predict the behaviour of stereophonic phantom imaging. Experimental work with subjective listening tests had raised questions about the effect of head movements in stereophonic listening [Bauer, 1962] while other tests had shown that reliable phantom images could not be created with lateral and rear loudspeaker pairs [Ratliff, 1974]. A mathematical description of the relationship between the ear signals and the perceived location was needed in order to resolve these issues. However, the way in which multiple coherent audio signals combine to produce a perception of direction is quite complex. The term *summing localization* is used to describe this mode of auditory localization which arises from multiple coherent loudspeaker signals.

Clark developed a mathematical framework which explained basic Blumlein stereophony in terms of the manipulation of level differences to produce IPD cues at low frequencies and ILD cues at mid to high frequencies [Clark et a, 1958l]. They also discovered, however, that the basic stereophonic method produces slightly different results at low and high frequencies. In a number of listening tests, the perceived direction reported was proportional to the predicted angle in both frequency ranges, yet the perceived direction was generally wider for high frequency signals. Clark recommended that this difference be corrected by adjusting the weighting of the difference and sum signals by a factor of 0.7 for frequencies above 700 Hz.

Bauer used phasor analysis to describe the stereophonic principle and formulated the stereophonic law of sines from which the standard stereo panning law is derived [Bauer, 1962]. Bauer also assumed that while head movements did indeed alter the perceived source direction they did not contribute directly to localization accuracy in stereophonic listening.

Makita's theory of stereophonic localization assumed that the perceived direction of a sound source corresponds to the direction of propagation of the wavefront produced by that source [Makita, 1962]. Essentially, it states that the angular direction of the wave-front produces the interaural cues which then in turn give rise to a perception of direction. Makita then went on to describe the wavefront as a function of the direction, velocity, and variations in velocity of the summed wave-fronts. A method of predicting the perceived direction of a stereophonic image was created and verified experimentally for a centrally positioned listener. A number of anomalies remained, however, as just like Clark, Makita found that the perceived direction was frequency dependent. In addition, while Makita proposed that head movements were indeed a factor in summing localization, experimental results were inconclusive.

3.5 Critics of Summing Localization

Summing localization is generally regarded as the best current description of the process whereby multiple coherent loudspeaker signals produce a single perceived source direction. However, this theory is not without its critics and various alternatives have been suggested. Gunther Thiele suggested that the localization of stereophonic images does not occur due to the physical superposition of the loudspeaker signals at the ears. Instead, the auditory system first detects the the multiple loudspeaker signals at different spatial locations, and then merges them together to a phantom source in a psychoacoustic process after their signal content was detected to be congruent [Thiele, 1980]. Thiele's Association Model therefore supposes that phantom sources and natural sources are perceived differently and rely on different processes. The ear signals which arise due to the superposition of multiple loudspeaker signals cannot therefore provide all the information about the properties of the phantom source. While Thiele's Association Model can account for certain aspects of stereophonic localization such as the timbral shift exhibited by phantom sources, it cannot predict the perceived source direction, and has proved difficult to verify experimentally. For these reasons, the theory of summing localization and not Thiele's Association Model is used throughout the rest of this thesis.

3.6 Meta-Theories of Localization

While these localization theories had shown that stereophony was reasonably reliable for two channels in front of a central listener, experience with Quadraphonics had suggested that this principle was not necessarily the best approach for three-dimensional audio. Michael Gerzon suggested an alternate approach which separates the process into two distinct stages of recording/encoding and playback/decoding. This division allowed for the design of a decoder that would be psychocoustically optimized, meaning it would be designed to satisfy as many of the existing localization theories and auditory cues as possible. Gerzon suggested that this approach, which forms the basis of the Ambisonics system, would provide the best performance that could be expected of any spatial audio reproduction system. His meta-theory of localization [Gerzon, 1992a] incorporated various different models of localization and is fundamental to the design of ambisonic decoders. He described a hierarchy of localization models and for each derives a vector whose direction θ gives the predicted direction of the sound, and whose magnitude R represents the stability of the localization. A real source would therefore have a vector magnitude of one with a positional angle described using spherical coordinates. The optimal decoder design would therefore ensure that the localization vectors for each model agreed for all frequencies, and that their magnitude was as large as possible, in every direction.

Gerzon's theory emphasized the importance of the two models based, respectively, on the velocity and energy flow of the sound field at the ears. The theories of Clark, Bauer and Makita are described as special cases of a more general description based on the velocity of the wavefront. This velocity model is primarily concerned with the IPD cues produced at low frequencies. Gerzon showed how a velocity vector equal to that produced by a real source can be produced by a multichannel system and that this will ensure that the perceived source direction will remain stable as the head is rotated. However, this can only be achieved at low frequencies, as at higher frequencies, the signal wavelength becomes comparable to the size of the human head and the effect of head-shadowing becomes more pronounced. The frequency range in which the velocity model is applied, i.e. the decoder cross-over frequency, is therefore related to the size of the effective listening area. Above a certain limit, the size of the listening area is smaller than the human head, and is therefore not useful. Choosing a frequency of approximately 700Hz would produce an effective listening area suitable for a single listener, while 400Hz [Gerzon, 1992a] would be suitable for a domestic listening situation with approximately six listeners.

Above this cross-over frequency, the decoder emphasizes the ILD localization cues which arise due to the directional behaviour of the energy field around the listener. It can be shown mathematically that it is only possible to recreate the energy field of a real sound source using a small number of loudspeakers, if the sound happens to be at the position of one of the loudspeakers. Therefore, at mid and high frequencies, not all of the ear's localization mechanisms can be satisfied in a practical reproduction system. The direction of the energy localization vector can, however, be adjusted so it matches the velocity localization vector ($\theta_E = \theta_V$) for all frequencies up to 4kHz. This is similar to the stereophonic approach recommended by Clark (see Section 3.4) [Clark *et al*, 1958]. In addition, Gerzon's design optimizes R_E in all directions, which necessarily compromises localization in the directions of the loudspeakers in favour of making the quality of the localization uniform in all directions [Benjamin *et al*, 2006]. This effectively eliminates the timbral change which occurs with amplitude panning as the signal moves from a position at a loudspeaker to one in between loudspeakers. This also means, however, that the localization of sources positioned at a loudspeaker will be less than optimal. In summary, therefore, Gerzon recommends that the following optimizations be implemented when designing an ambisonic decoder:

- The velocity and energy vector directions are the same up to 4kHz ($\theta_E = \theta_V$)
- At low frequencies, the magnitude of the velocity vector should be near unity for all directions ($r_V = 1$)
- At high frequencies, the energy vector magnitude should be maximised as much as possible and made consistent in all directions (maximum r_E)

The fundamental design theory of ambisonic decoders therefore concentrates on the velocity and energy models while other models are only used to further refine the design. It is clear therefore that the Ambisonics system was based on psychoacoustic principles, however, this system also built upon the work of Alan Blumlein with coincident microphone techniques and has been developed into a complete set of recording and production techniques.

4 Sound Field Reconstruction

Two of the most common techniques that attempt to reconstruct a sound field over a given area are Ambisonics and Wavefield Synthesis. While these techniques were initially derived from quite distinct theoretical fundamentals, recent research has shown that under certain conditions, both theories can be considered equivalent [Daniel *et al*, 1999; Nicol *et al*, 1999].

4.1 Ambisonics

Ambisonics was developed by Michael Gerzon, Peter Fellgett and others in the 1970s as an alternative spatialization technique for surround and periphonic loudspeaker arrangements [Gerzon, 1974a; Fellgett, 1975]. The basic approach taken by Ambisonics is described by one of its author as follows.

> *"For each possible position of a sound in space, for each possible direction and for each possible distance away from the listener, assign a particular way of storing the sound on the available channels. Different sound positions correspond to the stored sound having different relative phases and amplitudes on the various channels. To reproduce the sound, first decide on a layout of loudspeakers around the listener, and then choose what combinations of the recorded information channels, with what phases and amplitudes, are to be fed to each speaker. The apparatus that converts the information channels to speaker feed signals is called a "decoder", and must be designed to ensure the best subjective approximation to the effect of the original sound field* [Gerzon, 1974b]. "

This approach differs from Blumlein stereophony in a number of ways. Firstly, the initial encoding stage is removed from the eventual playback system, its sole aim being to capture as much information about the sound scene as possible using a certain number of channels. The decoding stage can now use the recorded spatial information to determine the appropriate loudspeaker signals that will recreate this spatial scene. Furthermore, as discussed in the last section, this decoding stage can be psychacoustically optimized so that as many localization mechanisms as possible are satisfied [Gerzon, 1992a]. The technical means by which this system was realised built on the work by Blumlein with coincident microphone techniques. This chapter will therefore begin with a summary of the Ambisonics system and the associated encoding and decoding process based on microphone directivity patterns. This will be followed with a more detailed description of the theoretical basis of the system and the application of psychoacoustic principles to the decoding process.

4.1.1 Ambisonics and Microphone Directivity Patterns

Consider the arrangement shown in Fig. 4.1. A point source S is producing a wavefront which travels from a frontal position to the centre listening position P. Four directional microphones are positioned coincidentally at the centre point P such that each microphone is pointing directly toward a loudspeaker positioned at the four corners. The basic stereophonic approach would be to record the wavefront with the four microphones and then replay the four signals from the four loudspeakers in order to reproduce the original wavefront at the centre listening position P.

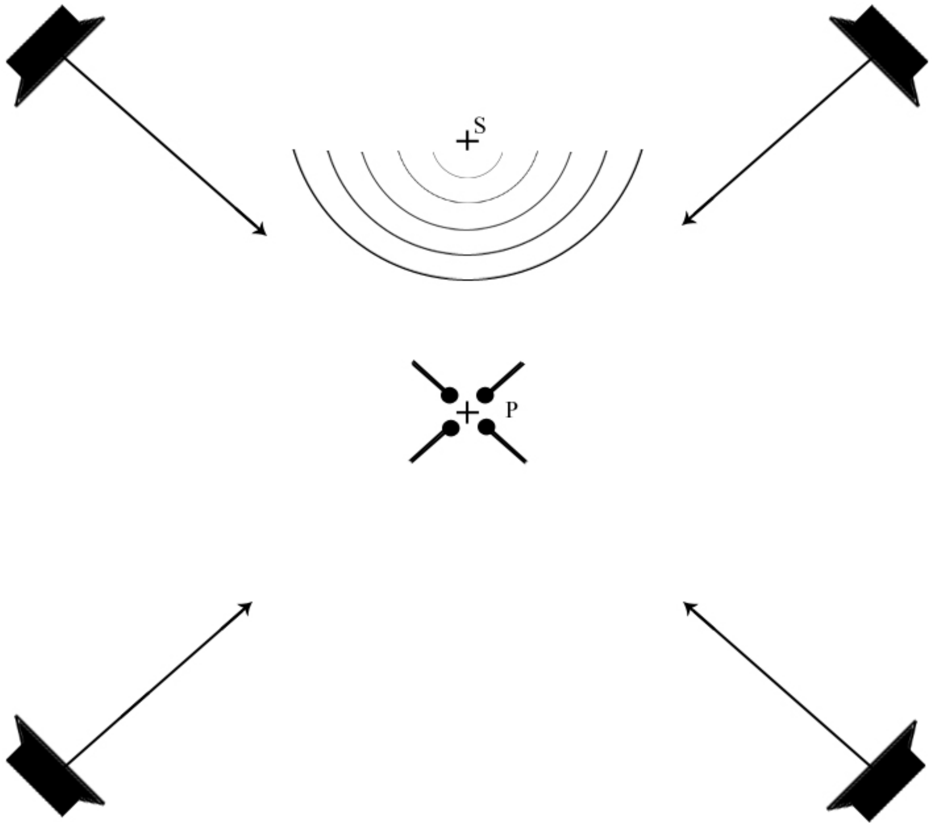

Fig. 4.1 Basic multichannel stereophony example

A unique characteristic of Blumlein's coincident figure-of-eight arrangement (shown in Fig. 4.2 [Wiggens, 2004]) is that the two response patterns can be combined to generate a new virtual figure-of-eight response pattern pointing in any direction in the same plane. The addition of a third vertically orientated figure-of-eight microphone extends this idea to three dimensions. Any figure-of-eight microphone can be combined with an omni-directional microphone to produce a response characteristic with increased directivity in a certain direction. So if an omni-directional microphone is added to this coincident arrangement of three figure-of-eight microphones, it will be possible to create almost any response pattern in any direction. If the four microphones shown in Figure 4.1 are replaced with this new arrangement, then four virtual response patterns corresponding to the original loudspeaker arrangement can be derived. However, if the number of loudspeakers is increased from four to eight, eight new virtual response patterns corresponding to this new loudspeaker arrangement can also be derived. In fact, with this particular arrangement of four microphones it is possible to create response patterns that correspond to whatever loudspeaker arrangement is available. In addition, the directivity of the response characteristic can be adjusted which is highly important in the decoding process.

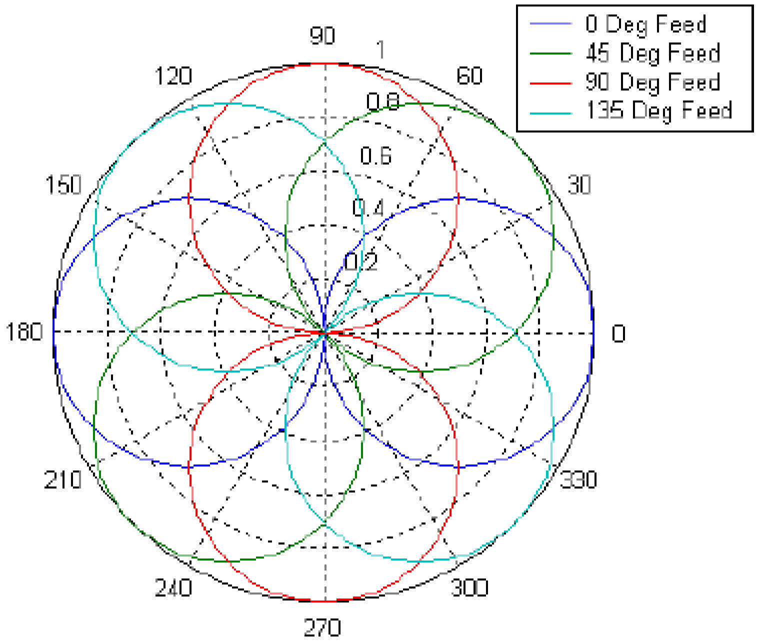

Fig. 4.2 Microphone responses derived from two figure-of-eight microphones

The process just outlined describes in broad terms the basic Ambisonics encoding method. This particular microphone arrangement is called a Soundfield microphone and is shown in Fig. 4.3. This apparatus contains the four described microphone capsules and produces four audio signals which correspond to the four microphone response patterns. The Ambisonics terminology for this set of four signals is referred to as A-Format. Obviously it is not possible to have four microphone capsules occupying the exact same point in space, however, Soundfield microphones are able to overcome this problem with electronic equalization so that the output produced is essentially coincident up to 10 kHz [Rumsey, 2001]. These corrected signals are labelled W, X, Y and Z in reference to the four microphone capsules and are generally collectively referred to as a B-format signal in Ambisonics terminology.

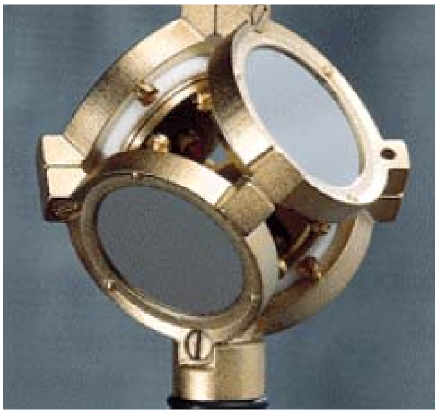

Fig. 4.3 The Soundfield microphone

The preceding description of the Ambisonics system illustrates how it developed from Blumlein's research into coincident microphone techniques. It describes how a relatively simple arrangement of four microphone capsules can be used to capture a significant amount of directional information about a sound field. The functionality of the Soundfield microphone readily lends itself to a description based on spherical harmonics. Using this mathematical description, any function of direction can be described as a function on the surface of a sphere, and expressed uniquely as a sum of spherical harmonics. The order of the theory is the order of spherical harmonics used by the model and a minimum of zero and first order components are required to represent a three dimensional field. Figure 4.4 graphically illustrates the zero and first order spherical harmonics [Wiggens, 2004], and their relationship with the Soundfield microphone response characteristics is clear. Using this mathematical description, it is also possible to encode an existing monophonic signal into a B-format signal and position it at any location on a unit sphere surrounding the central listening position. This technique forms the basis of ambisonic panning and the encoding equations can be readily applied in software. This mathematical description of the B-format encoded sound field also allows operations such as rotations and translations to be applied to the whole sound field. The tilt, tumble and dominance effects found in many ambisonic processors implement these particular effects [Malham, 1998].

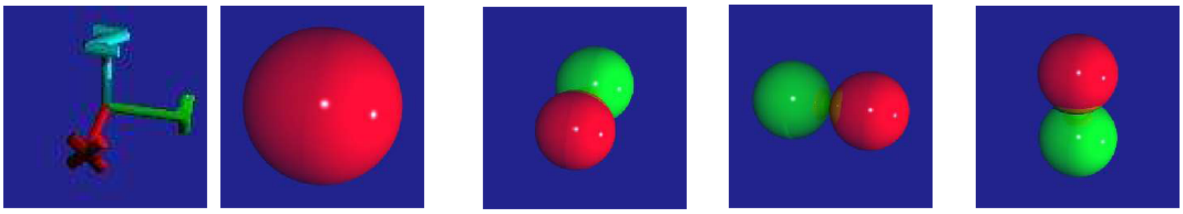

Fig. 4.4 Zero and first order spherical harmonics

4.1.2 Ambisonic Decoders

One of the unique advantages of the Ambisonics system is that the encoded signal can be decoded for a variety of loudspeaker arrangements from mono to stereo, horizontal arrays of loudspeakers or even periphonic (with height) arrays. It has been shown how the decoder design is psychoacoustically optimized in terms of the velocity and energy localization vectors described by Gerzon in his meta-theory of auditory localization [Gerzon, 1992a]. The decoding is accomplished by calculating the virtual response patterns needed for the available loudspeaker array and then performing two different decoding strategies at low and high frequencies. This is often implemented using shelf filters (consisting of 90^{o} all-pass filters) which smoothly alter the gain of W relative to X, Y and Z above and below the chosen crossover frequency. In more recent designs, a cross-over filter network is used to split the signal instead of the traditional shelf filters [Farina *et al*, 1998]. In practice, optimal decoders can be readily designed for regular polygon loudspeaker arrangements where the loudspeakers are placed in diametrically opposed pairs. Optimal decoders for non-regular loudspeaker arrangements are much harder to derive and are the subject of considerable research [Wiggens, 2004].

Gerzon defined the Ambisonics system in terms of the decoder design. However, a number of different decoders have since been suggested by Gerzon and others [Malham, 1995, Farina *et al*, 1998]. The original specification used shelf filters to apply two different sets of weightings to the W and XYZ components at low and high frequencies. This produces different directivity characteristics which in turn maximizes the velocity and energy vectors r_V and r_E at the frequencies in which these localization models dominate. As stated previously, the choice of cross-over frequency is related to the size of the effective listening area, with Gerzon recommending a figure of 700Hz for a single listener and 400Hz for a small group of listeners [Gerzon, 1992c]. As the size of the required listening area increases further, the appropriate cross-over frequency will continue to decrease. Intuitively, this suggests that a single-band decode would be more suitable for very large areas and indeed dual-band decoding using shelf filters is generally not recommended for large listening areas [Malham, 1992]. This modification becomes necessary when large numbers of listeners are seated away from the centre point, due to the way in which the velocity vector r_V is optimized in the original decoder specification. In order to ensure that $r_V = 1$ at the centre point, the loudspeaker directly opposite to the source direction must produce a signal with inverted phase relative to the loudspeaker signal at the source direction. The phase relationship between the multiple loudspeaker signals will therefore be distorted at off-centre listener positions. This is particularly true for listeners seated near to a loudspeaker as these anti-phase components can distort the perceived source direction toward the loudspeaker directly opposite the intended direction. At higher frequencies, it can be shown that ILD localization cues are produced by the temporally averaged directional behaviour of the energy field around the listener [Gerzon, 1992a]. The energy model is therefore not dependent on the phase relationship between the loudspeaker signals and should therefore be useful for non-central listeners. Various decoders suitable for large listening areas have been proposed. These implement a single frequency band decode which optimizes the energy vector R_E and reduces or completely eliminates anti-phase components. The effectiveness of these decoders will be examined in more detail in the next chapter.

4.1.3 Higher Order Ambisonics

The original Ambisonics system was based on the zero and first order decomposition of a sound field as this approach produced acceptable results with an economical number of channels. It has been shown how a function of direction may be described as a function on the surface of a sphere and expressed uniquely as a sum of spherical harmonics. If the order of spherical harmonics is increased then the accuracy of the encoded sound field also increases, as this effectively increases the spatial resolution of the system. The order of the system reflects the order of spherical harmonics used, so the traditional Ambisonics B-format signal represents a first order system, while a second order system would also include the additional spherical harmonics shown in Figure 4.5 [Wiggens, 2004]. Obviously it is extremely difficult to create higher order Soundfield microphones as there is a limit to how many capsules can be positioned at approximately the same point in space. This topic is currently the focus of considerable research and various higher order microphone arrangements are being examined (see Figure 4.6) [Bertet *et al*, 2009; Moreau *et al*, 2006]. However, there is no such problem when encoding monophonic recordings in software and it is in this regard that higher order Ambisonics has proved most useful. Various software implementations of high order ambisonic encoders and decoders are now available and this has important consequences for situations involving multiple listeners.

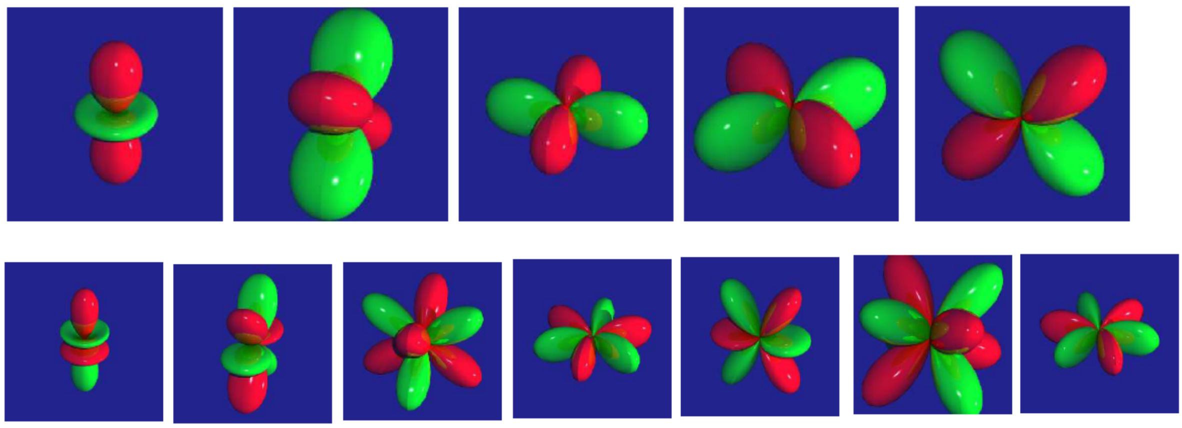

Fig. 4.5 Second (top) and third (bottom) order spherical harmonics

Fig. 4.6 First, second & third order microphones

4.1.4 Ambisonics in Practice

While the specification and design of first order B-format Ambisonics systems is well established and standardized, the same cannot be said for higher order systems. Although, Gerzon presented equations up to third order in his 1975 paper, these were specified using Cartesian co-ordinates rather than the polar system which later became standard in ambisonic systems. As a result, a defined terminology for higher order systems, and a spherical harmonic formulation for the higher order channels in a form which is consistent with B-format systems, has not yet been agreed upon [Malham, 1999].

W	=	$(1 / \sqrt{2})$
X	=	$\cos(A) * \cos(E)$
Y	=	$\cos(E) * \sin(A)$
Z	=	$\sin(E)$
R	=	$(3*\sin^2(E)-1) / 2$
S	=	$\cos(A) * \sin(2E)$
T	=	$\sin(A) * \sin(2E)$
U	=	$\cos^2 (E) * \cos(2A)$
V	=	$\cos^2 (E) * \sin(2A)$
K	=	$\sin(E) * (5*\sin^2(E)-3)/2$
L	=	$(\sqrt{135/256}) * (\cos(E)*\cos(A)*(5*\sin^2(E)-1))$
M	=	$(\sqrt{135/256}) * (\cos(E)*\sin(A)*(5*\sin^2(E)-1))$
N	=	$(\sqrt{27/4}) *(\cos(2*A)*\sin(E)*\cos^2(E))$
O	=	$(\sqrt{27/4}) *(\sin(E)*\sin(2*A)*\cos^2(E))$
P	=	$\cos(3*A)*\cos^3(E)$
Q	=	$\sin(3*A)*\cos^3(E)$
where A = source azimuth & E = source elevation		

Table 4.1 Furse-Malham set of encoding coefficients

While the various formulations of the ambisonic equations are mathematically equivalent, the same equations must obviously be used for encoding and decoding if the sound field is to be reconstructed correctly. The Furse-Malham (FMH) formulation of the ambisonic equations was developed to maintain compatibility with older B-format hardware and software while extending the system specification to third order spherical harmonics. The FMH set deviates from a strict mathematical description of spherical harmonics as it applies weightings to the channels so that all the harmonic coefficients have a maximum value of one [Malham, 1999]. In a mathematically strict formulation, as the order M of the harmonics is increased, the maximum value that each harmonic may attain increases. While this is not a problem for diffuse sound fields, point sources produced by panning could result in excessive signal levels in the higher order channels. The FMH set therefore applies weightings to each channel to prevent this occurring. In addition, the FMH set reduces the zeroth order W signal by 3dB (gain of 0.707). This gain factor was originally applied to produce equivalent signal levels in the W and XYZ components and was included in the FMH set to maintain compatibility with the original B-format specification used in existing ambisonic hardware such as the Soundfield microphone. The FMH set of coefficients for encoding a signal into third order Ambisonics is shown in Table 4.1.

Criteria	Furse Malham	N3D	N2D	Hybrid W	SN3D
Extensible to higher orders in a straightforward manner. Hence able to meet expanding users needs.	No	Yes	Yes	Yes	Yes
Consistent with 1st-order B-format recordings, equipment and software.	Yes	No	Yes	Yes	No
Consistent with higher-order horizontal-only FMH format recordings, equipment and software.	Yes	No	Yes	No	No
Consistent with the Ambisonics portion of MPEG-4	Yes	Yes	Yes	No	Yes
Consistent with FMH format recordings, equipment and software having periphonic order P greater than 1.	Yes	No	No	No	No
Components of the same degree have the same 3D gains. So that typical processing (such as rotation) can be coded reasonably cleanly.	No	Yes	Yes	Yes	Yes
The mathematical formulation is consistent across the degrees. No special cases.	No	Yes	Yes	No	Yes
The harmonics are orthonormal, which eases such things as beamforming.	No	Yes	No	No	No
The scheme benefits from direct support in general maths libraries.	No	Yes	No	Yes	Yes
The scheme has mathematical grounding in the 3D world.	No	Yes	No	No	Yes
Higher-degree components have headroom for when the sound field is not spherically isotropic. Means that 1st order monitoring can be used over higher orders (with more accuracy).	Yes	No	Yes	No	Yes
Greatest dynamic range when the major sources are in/near the horizontal plane	No	No	Yes	No	No
Greatest dynamic range when the major sources are distributed over the sphere	No	Yes	No	No	No

Table 4.2 Analysis of Ambisonics normalisation schemes

The deviation of the FMH set from a strict mathematical spherical harmonic formulation does, however, have some negative consequences. Firstly, while decoding equations can be readily obtained up to third order, this is much more difficult at higher orders. In addition, the calculation of matrices for tilt, tumble and dominance effects has proven to be extremely difficult, and, the alphabetical naming scheme is unpractical above third order. As it is much easier to design and implement Ambisonics systems based on more mathematically pure formulations of the basic equations, particularly at higher orders, a number of different formats have been proposed. Efforts are being made by the Ambisonics community to standardize these different formats and an analysis of the advantages and disadvantages of the prominent formats is shown in Table 4.2. The necessary weightings to convert between these different formats are also readily available [Malham, 1998; Ambisonics Standards, 2009].

It should be noted that the Ambisonics system assumes that all loudspeakers are at a great enough distance from the listener so that the signals arriving at the ears can be approximated by plane waves. However, this may not be the case when there is only a single listener within a relatively tightly spaced loudspeaker array. This effect is illustrated in Figure 4.7 [Nicol *et al*, 1999] which compares the reproduction of a plane wave using ideal and real loudspeakers. In this case nearfield compensation must be applied to the decoder to negate the low frequency boost which results due to the spherical radiation pattern of the nearby loudspeakers [Daniel, 2003].

The decoding equations for regular loudspeaker layouts are readily available up to second order [Furse, 2000] and numerous higher order software decoders are available in Steinberg's VST plugin format or as externals for packages like Max MSP and Audiomulch. The decoding equations can be formulated in a number of different ways, such as;

$$S = (\sqrt{2}/N) * [W + \sqrt{2}(\cos(a) * \cos(e) * X) + \sqrt{2}(\sin(a) * \cos(e) * Y) + \sqrt{2}(\sin(e) * Z)]$$
[Gerzon, 1992a]

$$S = (0.5) * [(2 - d)(\sqrt{2})(W) + d((\cos(a) * \cos(e) * X) + (\sin(a) * \cos(e) * Y) + (\sin(e) *))]$$
[Wiggens, 2004]

$$S = \text{zerobase} * (W) + \text{firstbase} * [(\cos(a) * \cos(e) * X) + (\sin(a) * \cos(e) * Y) + (\sin(e) * Z)]$$
[Schacher *et al*, 2006]

where S is the loudspeaker signal
W,X,Y and Z are the B-format signals
N is the number of loudspeakers
d is the directivity factor (0 = omni, 1 = cardioid, 1.5 = hypercardioid, 2= Fig-of-8)
zerobase, firstbase = reformulated directivity factor

These formulations of the first order decoding equations differ in a number of ways. The first difference relates to how the 3dB gain adjustment (factor of √2) is distributed between the zero and first order components. Similarly, the gain adjustments which are used to alter the directivity of the response characteristic can be applied in a number of ways, as shown in the second two equations. Interestingly, only the first equation includes a weighting factor due to the number of loudspeakers, so if the second two equations are used the perceived volume will increase as more loudspeakers are added. The first approach has some potential benefits as, if the volume is automatically reduced when more loudspeakers are added, as in the first equation, then, the signal to noise ratio (or the bit resolution in a digital implementation) will also be reduced.

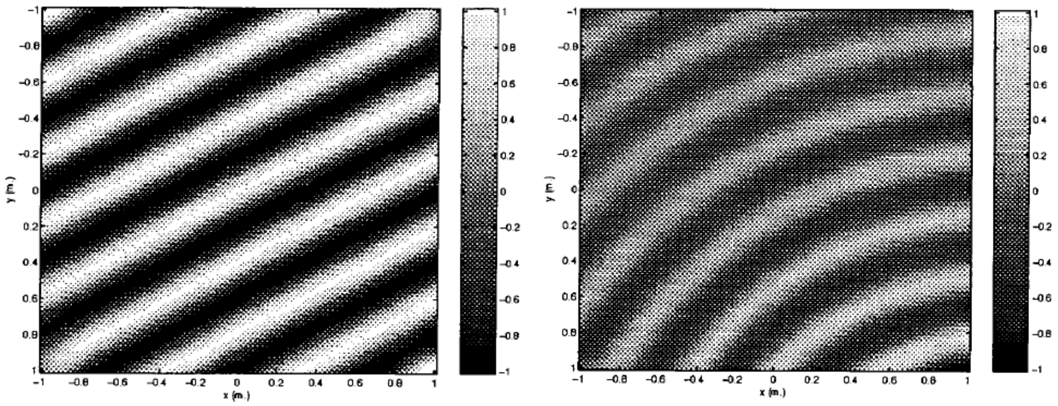

Fig. 4.7 Ambisonic plane wave - theoretical (left) and real (right) sources

Decoder Titles / Descriptions	Strict, Matched, Velocity, max-r_V, Furse-Malham, Systematic, Ideal	Energy, max-r_E, Rationalised Energy Decoder, Rationalised Square Decode	In-phase, Controlled opposites, Cardiod
Approx. optimum frequency ranges	< 700 Hz	500Hz - 5kHz	All
Response Characteristic	Hypercardioid, strong anti-phase components	Hypercardioid, reduced anti-phase components	Cardioid, no anti-phase components
Effective Listening Range	Single Listener	Increased listening area, still some anti-phase components	Large listening area but reduced localization accuracy.
Criteria (1st order)	$r_V = 1$ $r_E = 0.667$	$r_V = r_E = 0.707$	$r_V = 0.5$ $r_E = 0.667$
Criteria (2nd order)	$r_V = 1$ $r_E = 0.8$	$r_V = r_E = 0.866$	$r_V = 0.667$ $r_E = 0.8$
Criteria (3rd order)	$r_V = 1$ $r_E = 0.857$	$r_V = r_E = 0.924$	$r_V = 0.75$ $r_E = 0.857$
Crieria (all)	$\theta_V = \theta_E$	$\theta_V = \theta_E$	$\theta_V = \theta_E$

Table 4.3 Summary of ambisonic decoding schemes

The precise decoder weightings, which correspond to the directivity characteristic, can be derived mathematically in various different ways. In general there are four particularly common decoding schemes, which are summarized in Table 4.3. The original dual-band decoder proposed by Gerzon optimizes the velocity and energy vectors in different frequency ranges. Each of these two decoding schemes can of course also be applied across the entire frequency range. The final common decoding scheme is the in-phase decode proposed by David Malham for large areas. Figure 4.8 illustrates the directivity patterns of these different ambisonic decoding schemes [Daniel, 2000].

It should also be pointed out that there are a minimum number of speakers needed to successfully reproduce each ambisonic order, and that this number is always greater than the number of channels available for the decoder [Wiggens, 2004]. The number of loudspeaker required can be calculated from the following equation, where M is the order of the system, and N is the number of loudspeakers.

$N > 2M+1$ for horizontal arrays only

$M > (N+1)^2$ for periphonic arrays.

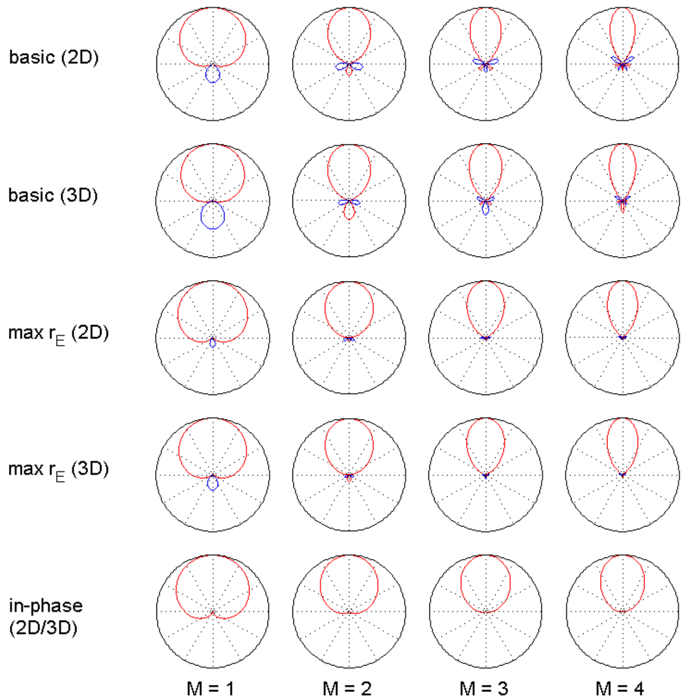

Fig. 4.8 Directivity patterns of various ambisonic decoding schemes

The preceding discussion illustrates the many different implementations of the Ambisonics system that have been developed since its conception in the 1970s. In many respects the Ambisonics system is defined in terms of the decoder, and the design of the decoder is perhaps one of the most crucial aspects of any Ambisonics system. Various objective and subjective assessments have been carried out in an attempt to evaluate the perceptual effect of these different decoding schemes, and these will be summarized in the next Chapter.

4.2 Wavefield Synthesis

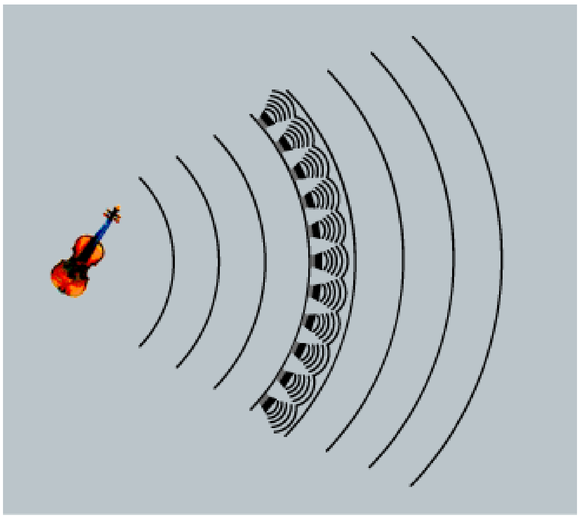

Fig. 4.9 Illustration of the Huygens principle

In much the same way that Ambisonics can be considered as an extension to the theories of Alan Blumlein, Wavefield Synthesis (WFS) can be considered as an extension of the acoustic curtain approach developed at Bell Labs [Steinberg *et al*, 1934]. The theoretical basis of WFS was developed by Berkout at TU Delft in the eighties. Berkout, whose background was in seismology, used Huygen's wave theory as the basis for a system of acoustical holography [Berkout, 1998]. Huygen's principle states that any wavefront can be regarded as a superposition of elementary spherical waves. Therefore, any wavefront can be synthesized by a number of secondary point sources, as illustrated in Figure 4.9. Berkout used Kirchoff's quantified version of this theory to create a generalised spatialization system using linear arrays of loudspeakers (see Figure 4.10). In effect, it allowed for the calculation of loudspeaker signals which would produce the same wavefront in the listening area as would have been produced by a real source in any position behind, or even in front of the array. While the particular integral equation involved is quite complex it essentially operates by adjusting the delay and gains of the signal produced by each loudspeaker so that they combine to produce the required wave. This equation includes a term which must be adapted to the actual directivity characteristics of the loudspeakers in the array [de Vries, 1995].

The WFS approach has a number of potential advantages over other spatialization techniques such as stereophony or Ambisonics. Firstly, as WFS effectively samples and reconstructs the desired wavefront, it can potentially reproduce the correct wavefront over an extended listening area. This has enormous potential for the presentation of spatial music and is quite different from the single listener orientated formats such as stereophony or Ambisonics. In addition, WFS can potentially position point sources both behind and in front of the array as illustrated in Figure 4.11

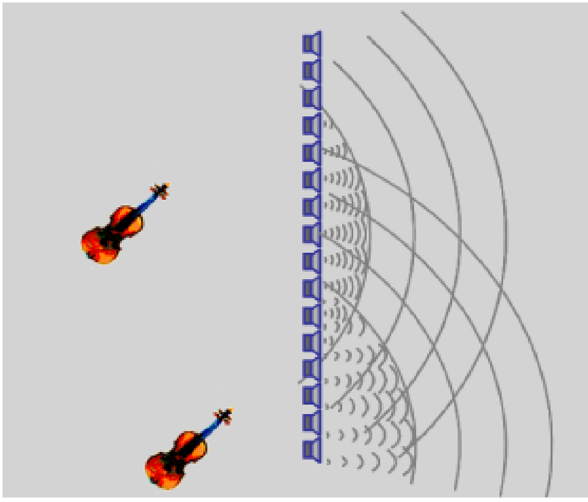

Fig. 4.10 The WFS method

Plane wave sources can also be achieved with WFS, as shown in the middle diagram in Figure 4.11. This type of reproduction mimics the effect of very distant sources and can potentially be used to extend the suitable listening area for standard two-channel stereo reproduction. Figure 4.12 illustrates how generating the two stereo signals as plane waves can potentially maintain the angular relationship between the two wavefronts across a greater area [Franco *et al*, 2004].

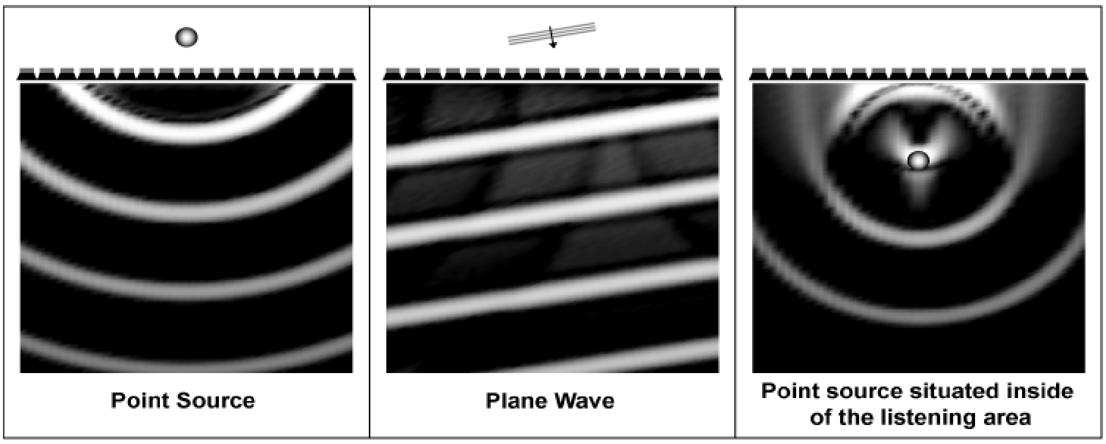

Fig. 4.11 WFS reproduction of different wavefronts

While WFS appears to be a promising technique there are a number of problems which have proved quiet difficult to overcome. Many of these issues arise from the simplifications which must be applied to the Kirchoff-Helmholtz integral equation so that it can be used with practical loudspeaker arrays. Firstly, the Kirchoff-Helmholtz integral describes the behaviour of spherical wavefronts in three dimensions, so the restriction to the horizontal plane introduces amplitude errors in the reconstructed wavefront. WFS can be implemented in three dimensions using a loudspeaker wall [Komiyama *et al*, 1991; Ono *et al*, 1997], however, this drastically increases the required processing power and number of loudspeakers. For this reason, typical WFS setups use linear arrays which are restricted to the horizontal plane and include additional processes in an attempt to reduce the errors introduced by this simplification.

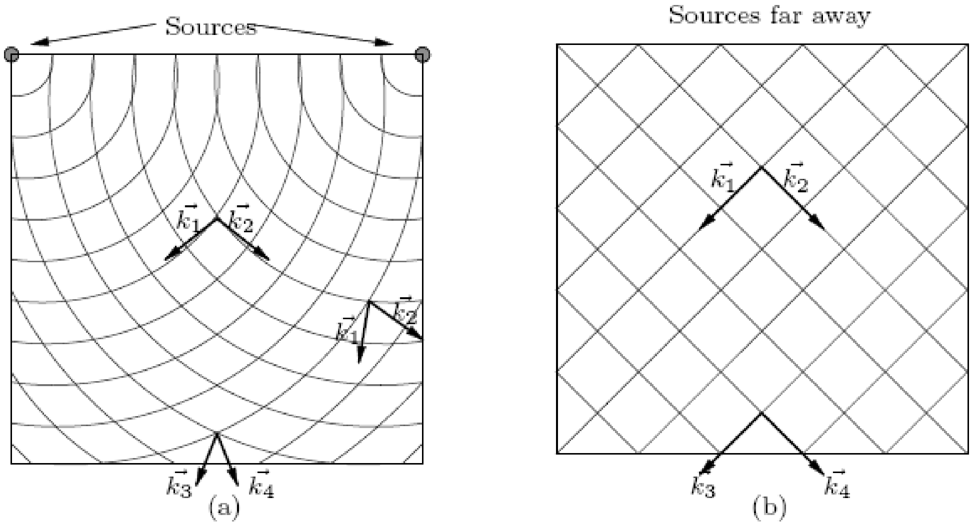

Fig. 4.12 WFS reproduction of two-channel stereo

In WFS theory, the new wavefront is constructed by summing an infinite number of secondary sources, whereas in reality only a finite number of loudspeakers are available. This simplification places boundaries on the effective listening area and can also potentially introduce echo signals due to diffraction effects at the edges of the array. These truncation effects can be clearly seen in Figure 4.13 which shows a source placed two metres behind the centre of the array and the two echo signals which arise from the edges of the array [Franco *et al*, 2004]. Truncation effects can be reduced by applying a tapering windowing function to the loudspeaker signals in order to reduce the weighting of the loudspeakers signals at the edges of the array

[Boone *et al*, 1995] However, this solution does reduce the effective listening area [Sonke, 2000].

Perhaps the most significant issue with WFS is the problem of spatial aliasing which arises due to the discrete nature of the loudspeaker array. Essentially, the inter-loudspeaker spacing places an upper limit on the frequencies which can be accurately reconstructed with the array. This effect is somewhat analogous to Nyquist's sampling theorem which states that the sampling rate must be twice the highest frequency to be reproduced. A WFS system effectively samples a virtual wavefront, and in this case it is the inter-loudspeaker spacing which places an upper limit on the frequencies which can be accurately sampled and reproduced.

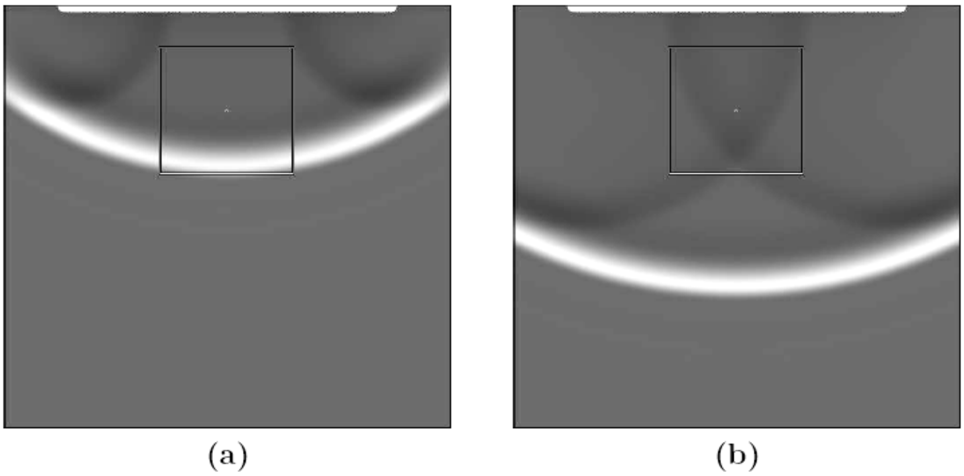

Fig. 4.13 Truncation effects (a) and (b) 4ms later

This frequency limit, the so-called aliasing frequency, is also dependent on the angular position of the source and listener, relative to the array. Figure 4.14 shows a distant source positioned at 30^{o} to the left behind the array for two frequencies above and below the aliasing frequency [Franco *et al*, 2004]. Above this limit, the original shape of the wavefront is not reproduced and the perceived direction and spectral coloration of the source will be highly dependent on the listening position.

The aliasing frequency can obviously be increased by using a greater number of loudspeakers placed closer together. However, in practical WFS systems this is often not possible. One solution is to randomize the time-offsets of all of the high frequency source content in an attempt to reduce the extent of the artefacts which occur due to spatial aliasing [Start, 1997]. This approach does, however, sacrifice the spatial accuracy of the sound field at high frequencies and has proven to be quiet difficult to implement. One promising solution to this problem is to use standard stereophonic phantom imaging above the spatial aliasing frequency, and WFS below. Figure 4.15 illustrates this approach of Optimized Phantom Source Imaging (OPSI) where the three blue loudspeakers are used to create phantom images which correspond to the direction of the virtual source generated by the WFS system [Wittek, 2002]. Experimental evaluation of this approach has been promising and suggests that it does not negatively impact on perceived source localization accuracy, even over large listening areas [Huber, 2002].

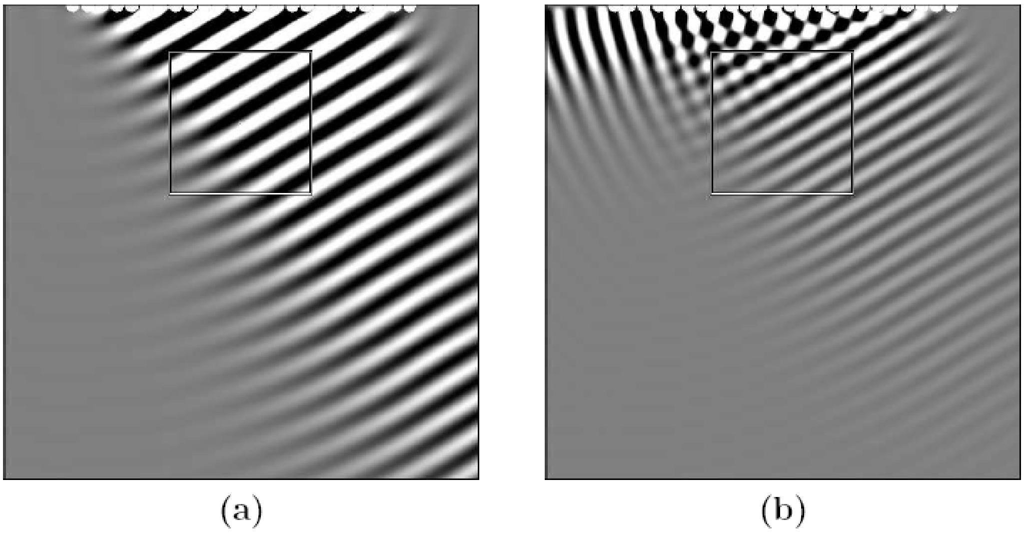

Fig. 4.14 A distant source (a) below the aliasing freq & (b) above

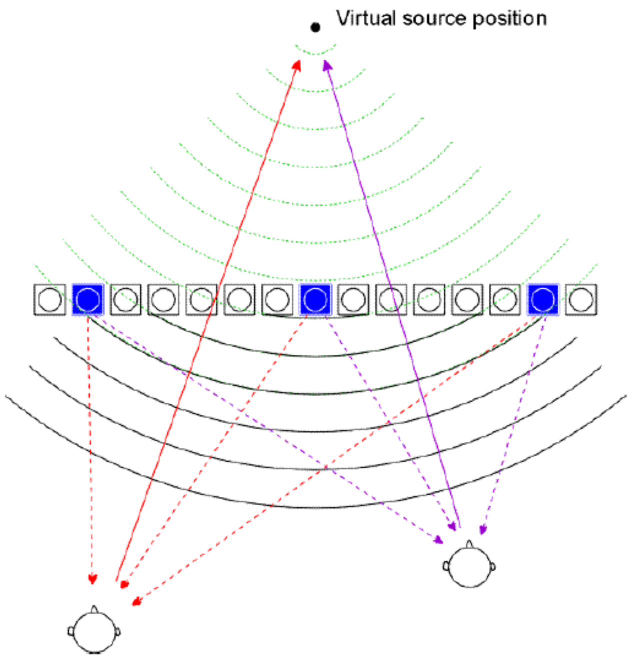

Fig. 4.15 Optimised Phantom Source Imaging WFS

The influence of acoustic reflections of the loudspeaker signals within the listening room can also potentially degrade the perceptual accuracy of the reproduced wavefronts to a significant degree. While traditional physical acoustic absorption techniques in the listening room can obviously be used, active cancellation using the reproduction loudspeaker array is also under investigation [Spors *et al*, 2003]. Some authors have suggested that the negative influence of the listening room acoustic is one of the most serious problems in practical WFS systems [Wittek, 2003]

Fig. 4.16 A WFS cinema system in Ilmenau

4.2.1 WFS and Spatial Music

While the WFS approach has numerous potential benefits, it is extremely demanding in terms of hardware and processing power. The relationship between the size and spacing of the array and the effective listening area means that large numbers of loudspeakers are required to create an effective listening area sufficient for a large audience. For example, in order to obtain a spatial aliasing frequency of 1kHz over a circular listening area with a diameter of two metres, a WFS system with 32 loudspeakers is required [Theile *et al,* 2003]. It is clear therefore that extending this system to cover a listening area suitable for a reasonably sized audience is logistically challenging, and for this reason, WFS has not to date been widely used for the performance of spatial music. However, the potential of WFS to create sources within the array and to work over an extended area, means that this technique will continue to be an intriguing prospect for composers of spatial music. WFS is also the subject of considerable research and a number of commercial WFS systems are now available. One system has been successfully operating in the Ilmenau cinema since 2003 (see Figure 4.16). In addition, the WONDER system, developed by Marije Baalman for the Linux environment, is a free open-source WFS program developed specifically for the production of spatial music [Baalman, 2006]. While WFS is certainly a highly promising technique, the large amount of hardware required for its implementation will most likely restrict its use to large dedicated venues for spatial music production, at least for the near future.

WFS is one of the few spatialization techniques which can theoretically recreate the correct wavefield over an extended listening area and its ability to simulate virtual sources within the listening area is certainly unique. Other techniques such as stereophony and Ambisonics can only reproduce sources at the distance of the loudspeakers, and additional processes must be used to simulate distance. These processes attempt to simulate the perceptual cues related to distance hearing presented in Section 2.2, and are discussed in the next Section.

5 The Simulation of Distance

As discussed previously in Section 2.3, the perception of distance is thought to largely depend on the following parameters;

- The relative amplitude of the source signal
- The spatio-temporal pattern of early reflections
- The ratio of the direct and reverberant signals
- The attenuation of high frequencies due to air absorption
- The Doppler effect
- The familiarity of the listener with the source signal

While recorded sounds may retain many of these perceptual cues, the same cannot be said for synthesized sounds. A method of simulating distance through the manipulation of these perceptual cues was first proposed by John Chowning in 1971. This system, although originally specified for a Quadraphonic system, can readily be adapted to other loudspeaker arrangements and directional panning techniques. When this system was first proposed, the influence of early reflections on distance perception was not yet fully understood. Chowning's system therefore used the direct to reverberant signal ratio as the primary distance cue, with some modifications.

5.1 The Simulation of Moving Sources

Chowning's model for the simulation of moving sources [Chowning, 1971] used amplitude panning to position a source at a certain azimuth within a Quadraphonic loudspeaker array. At the same time, the source signal is also routed to a multichannel reverberation unit. The perceived distance of the source is then controlled by simply adjusting the relative amplitude of the direct and reverberant signals. Chowning proposed that the amplitude of the direct signal should be set to 1/distance while the reverberant signal is set to 1/√distance. This configuration will result in the intensity of the direct sound reaching the listener falling off more sharply with distance than the reverberant sound, such as would occur with a real source. However, this approach has the result that at large enough distances, the reverberant signal will become greater than the direct signal, and the source direction will become ambiguous. To overcome this, Chowning proposed that the reverberation be divided into the global reverb signal which is reproduced from all loudspeakers and is proportional to 1/distance, and the local reverberation which is produced in the same direction as the direct signal and made proportional to 1-(1/distance). Therefore, as the distance increases, the reverberation becomes increasingly localized, compensating for the decreasing direct signal level. Chowning suggests that this is a fair approximation of a real acoustical situation as when the distance of a sound source increases, the distance to nearby reflecting surface decreases, thereby giving the reverberation some directional emphasis [Chowning, 1971]. More recent research has suggested that the directional aspect of the reverberant signal, particularly of the lateral early reflections, is used by the auditory system to estimate the egocentric distance to the source [Bronkhurst, 2002; Gerzon, 1992b]. The directional behaviour of the reverberant signal in Chowning's system is somewhat similar to this situation. However, as the precise spatio-temporal pattern of the reflected signal is not recreated, it is likely that this system will not provide an absolute sense of distance. The advantage of this system lies primarily in its efficiency and Chowning went on to develop a real-time digital implementation of the system using a standard Quadraphonic loudspeaker array. The inclusion of the Doppler effect and other secondary cues such as high frequency air absorption, when combined with independent control of the direct and reverberant signals resulted in an effective simulation of movement which was highly sophisticated for its time and indeed this approach is still widely used today, particularly for real-time applications. Chowning illustrated this new technique, along with a number of FM synthesis algorithms (also developed by Chowning) in the landmark work, *Turenas*, which was completed in 1972.

A significant increase in computational processing power was required if multiple discrete reflections, as well as the diffuse global reverberation, were to be accurately simulated. Indeed, it was more than a decade after Chowning's 1971 paper before such a spatialization system was proposed.

5.2 A General Model for Spatial Processing of Sounds

F. R. Moore developed Chowning’s algorithm into a more general conceptual spatial processing model which was implemented in the Cmusic sound synthesis language [Moore, 1983]. This approach draws upon previous work on artificial reverberation techniques [Schroeder, 1962; Moorer, 1979] and psychoacoustics

[Blauert, 1997; Gardner, 1962]. Moore's model represents the performance and illusory acoustic spaces as two rooms, namely;

- The listening area or inner room whose boundary is marked by the surrounding loudspeaker array, and
- The illusory acoustic space or outer room from which sounds emanate.

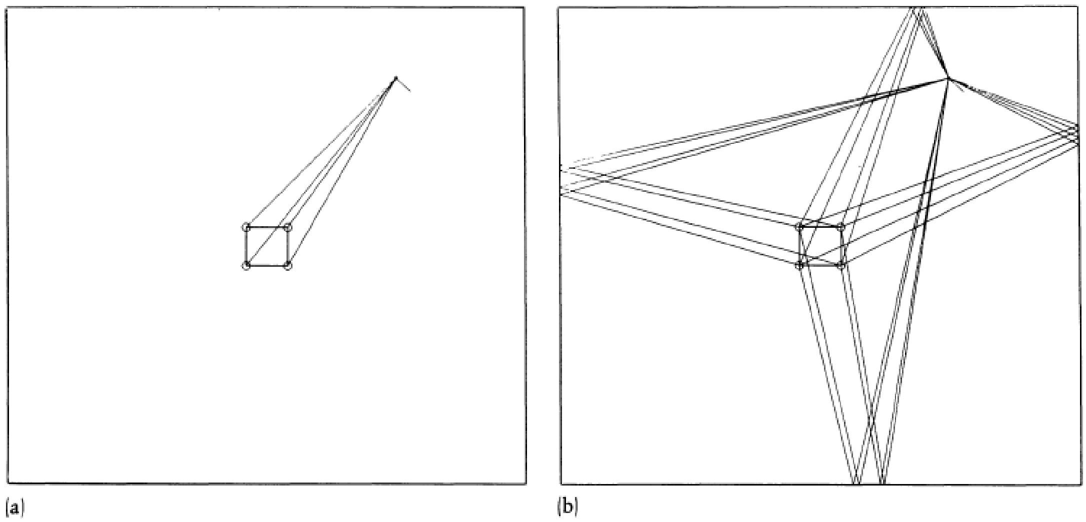

Fig. 5.1 Moore's spatial model (a) direct signal (b) reflected signal paths

In this model, the loudspeakers function as acoustic windows which transmit information from the illusory space outside the listening area perimeter. Increasing the number of loudspeakers therefore increases the spatial resolution of the illusory space. The paths of both direct and reflected signals are calculated using this two room model with the assumption that the boundaries of the inner room/listening area are completely absorptive. This has the effect that any paths which cut through the inner room are completely attenuated. Figure 5.1 (a) shows the direct signal path from a front-right source to a quadraphonic array. In this instance the back-left loudspeaker would be completely attenuated due to the absorptive characteristics of the inner room boundary. Figure 5.1 (b) illustrates sixteen potential reflected paths, one from the source to each wall and to each loudspeaker [Moore, 1983]. Each of these paths are evaluated in terms of the following parameters;

- Attenuation due to distance travelled
- Frequency dependent attenuation due to air absorption
- Frequency dependent attenuation due to reflection and absorption by the reflecting surface

- Absorption due to collision with the outer walls of the inner room (these are modelled as being completely absorptive)
- Time delay due to the finite speed of sound transmission

The various attenuation factors due to reflections and absorption are derived from measurements of real acoustic spaces and surfaces. A reverberation algorithm based on the overall size and shape of the outer room and the reflective properties of its walls is used to generate the diffuse global reverberation. Multiple delay lines are used to model the transmission times of the direct and reflected signals from the source to the listener. A change in source distance will therefore result in a change in the amount of delay, producing a pitch shift which is very similar to the Doppler Effect.

Real-time implementations of Moore's model were later developed [Ballan *et al*, 1994] including a version for the Max/MSP environment [Yadegari *et al*]. The simplifications introduced in order to produce a real-time configuration included the down-sampling of source paths, interpolating delay times, and improvements in the inner room ray intersection detection algorithm [Yadegari *et al*, 2002]. The real-time modular spatial-sound-processing system Spat (discussed previously in Section 3.1.3) is also based on the algorithms developed by Chowning and Moore [Jot, 1999]. More recent adaptions of this model for the Csound processing language have also been developed [Hofmann, 2008] and will be discussed later in Chapter Twelve

5.2.1 The Implications of Moore's Model

The spatial audio model developed by Moore differs from other spatialization techniques in that it emphasizes the simulation of distance rather than direction. Systems like stereophony and Ambisonics attempt to reproduce the psychoacoustic cues predominantly associated with directional hearing to the best possible extent, in every direction. Moore argued that this approach cannot be applied in a concert situation as different perceptual cues would be required at each listening position in order to produce the same perceived source direction for each listener. Moore's model therefore focussed on the physical characteristics of the real or imaginary space or spaces to be simulated, while ignoring the psychoacoustic parameters which vary for different listener positions. The contributions of the real listening environment are simply discounted and the composition is specified solely in terms of the virtual space. Moore argues that although each listener will hear something different, all listeners will in general perceive different perspectives of the same illusion. He states;

> *"Depending on proximity to the loudspeakers, each listener will hear these sounds from a slightly different perspective. Information is presented at each loudspeaker about all sound sources in the virtual outer room. The differences in perception among listeners in the inner room are analogous to perspective distortion.* [Moore, 1983]*"*

In effect, Moore assumes that each listener will localize the virtual source to a different direction, but that the relative source distance and the broad overall trajectory will still be apparent to each listener.

5.3 Ambisonics Distance Encoding

As discussed in Chapter Three, modifications must be made to an Ambisonics system when the loudspeakers are very near the listener. In this case, nearfield compensation is applied to the decoder to negate the low frequency boost which results due to the spherical, rather than the assumed planar, radiation pattern of the nearby loudspeakers. This is somewhat similar to the proximity effect which occurs with velocity microphones. Daniel proposed a new encoding format called Near-Field Compensated Higher Order Ambisoncs (NFC-HOA) which compensates for this bass boost in the encoding stage [Daniel, 2003]. Daniel also suggested that this effect could be used for the simulation of distance. Sources at the loudspeaker distance are simply uncompensated, as they will reproduce the distance related wavefront curvature anyway. Sources beyond the array will have reduced low frequency content to compensate for the nearfield distortion, while sources inside the array can in theory be simulated by increasing the proximity bass boost. As this filtering is applied in the encoding stage, the size of the loudspeaker array now has to be known in advance. This is a significant disadvantage as the clear separation of the encoding and decoding stages is one of the principal benefits of Ambisonics. However, it may be possible to apply additional filtering to compensate for any differences between the array size assumed during encoding and the actual array specified at the decoding stage [Daniel, 2003]. It has also been suggested that focussed ambisonic sources (similar to WFS) can also be created inside the array, but this approach has yet to be experimentally verified [Ahrens *et al*, 2009].

5.4 Evaluating of Spatialization Techniques

The preceding two chapters outlined the various perceptual mechanisms involved in spatial hearing and also the different ways spatialization techniques attempt to reproduce these auditory cues. A spatial location can be defined in terms of the angular direction and distance from the listener to the source, and different electronic processes have been developed to simulate these two parameters. WFS is the one exception, as this system can theoretically recreate the exact wavefield, and hence both the directional and distance cues. The effectiveness of this, or any approach to spatial audio can only really be evaluated through empirical research. This is particularly true for situations involving multiple listeners as many of these spatialization systems were primarily designed for a single or small group of listeners. The following chapter will therefore summarize the results of a number of experiments on the perceptual performance of various spatialization techniques, using both objective and subjective criteria. It is hoped that the analysis of these results will provide some insight into the capabilities and limitations of these techniques, which will in turn provide some indication of the perceptual validity of various compositional strategies in a performance setting.

6 The Assessment of Spatial Audio

Spatial audio production techniques have been evaluated using both subjective listening tests, and objective tests based on acoustic measurements and/or data simulations. Many of these tests focus on angular localization accuracy, particularly in the horizontal plane, and use this as the primary measure of the effectiveness of the spatialization technique(s) under examination. The source angle can be measured using subjective data, with listening tests for example, or with objective data generated from recordings or data simulations. The angular position of a static source in a fixed location is perhaps one of the easiest parameters to assess in spatial audio. However, it by no means follows that this is the only important factor. Naturalness is another highly important aspect of any sound reproduction system and this highly subjective parameter is often used to determine the listener's impression of the realism or "trueness to nature" of the auditory scene [Rumsey, 2001]. Rumsey suggests that "out of phase" effects, such as when the phase of one loudspeaker or headphone is inverted, can contribute to negative judgements of naturalness as this effect never occurs under natural circumstances [Rumsey, 2001].

The perceived distance, and in the case of dynamically moving sources, the smoothness of the trajectory, are both important parameters in spatial music composition. Is it preferable that every listener perceives the same trajectory, even if the movement crudely jumps from loudspeaker to loudspeaker, or is it better if every listener hears a natural and smooth movement, even if the precise trajectory varies from listener to listener? These broad questions are difficult to assess rigorously in an unbiased subjective examination and the wide range of parameters to be controlled means that data modelling or simulations can only provide limited results. For this reason, many researchers focus either on specific criteria, such as angular location, distance or spatial trajectory, or alternatively on broader, more subjective attributes such as envelopment or naturalness, which are often assessed with free-response or scaled judgements between different reproduction techniques. Concert hall acoustic studies in particular focus on the sense of spaciousness or envelopment which arises due to the diffuse, ambient sound and is desirable in this context (see Chapter Two). Localization studies are instead primarily used to examine the perceptual mechanisms related to directional hearing, or the performance of electronic spatialization techniques and so tend to focus on measurements of the size and location of the source signal. However, while this distinction serves a reasonable purpose in this case, it is not clear that such a distinction can be made in the area under discussion in this thesis. A complex sound scene will typically contain multiple sound objects which may overlap or even merge depending upon the context. This may result in a diffuse source which leads to a sense of spaciousness without a clearly perceived direction. In addition localization measurements only record the perceived source direction and do not indicate whether the source is sharply defined and focussed or widely dispersed around a particular direction. Blauert's measure of "locatedness" which was discussed earlier in this thesis will be an important measure in this regard [Blauert, 1997].

A common approach in this area is to restrict the discussion to the ideal situation of a single, centrally positioned listener in an anechoic or near-anechoic room. While this approach may provide useful information about the relative strengths of different techniques it does not address their performance under less than ideal conditions. For performances of spatial music, it is critical to also assess the influence of off-centre listener positions and the effect of the acoustic reflections in the playback room.

6.1 Localization Accuracy under Ideal Conditions

Numerous studies have been carried out to determine the localization accuracy of the human auditory system [Blauert, 1997; MacPherson, 2002]. It has been shown that the region of most precise spatial hearing lies in the forward direction with frontal hearing having an accuracy of between 4.4^0 and 10^0 for different source signals. Localization ability decreases as the source azimuth moves to the sides, with the localization blur at $\pm 90^0$ being between three to ten times its value for the forward direction. For sources to the rear of the listener, localization blur improves somewhat but is still approximately twice that for frontal sources [Blauert, 1997]. The assessment of dynamically moving sources is often measured in terms of the minimum audible movement angle (MAMA), which is the smallest difference in source direction perceptible by the listener. Studies have found that the MAMA is independent of the source trajectory but that its resolution is a U-shaped function of source velocity which degrades as the velocity increases or decreases beyond the optimum range [Saberi *et al*, 1990].

6.2 Assessing Stereophony

As discussed in earlier in Chapter Three, the basic stereophonic principle cannot reliably produce phantom images to the sides or rear of a listener. To produce a predictable phantom image, the listener must be facing a pair of equidistant loudspeakers separated by not more than 60^0. For a lateral loudspeaker pair, small changes in signal level result in large changes in perceived direction and localization is highly unstable (see Figure 3.4). The basic Quadraphonic system, which extends the loudspeaker angle to 90^0, cannot therefore produce reliable phantom images, even for a single listener. Theile recommended a six channel system to overcome these problems as this arrangement uses an additional loudspeaker at +/-90^0, and hence maintains the ideal loudspeaker separation angle of 60^0 [Theile *et al*, 1976]. The results of listening tests carried out under anechoic conditions with this arrangement and a single, central listener are shown in Figure 6.1. Clearly this arrangement improves localization accuracy for lateral sources. However, distortions still occur for phantom images positioned at either side of the loudspeakers at $\pm 90^0$.

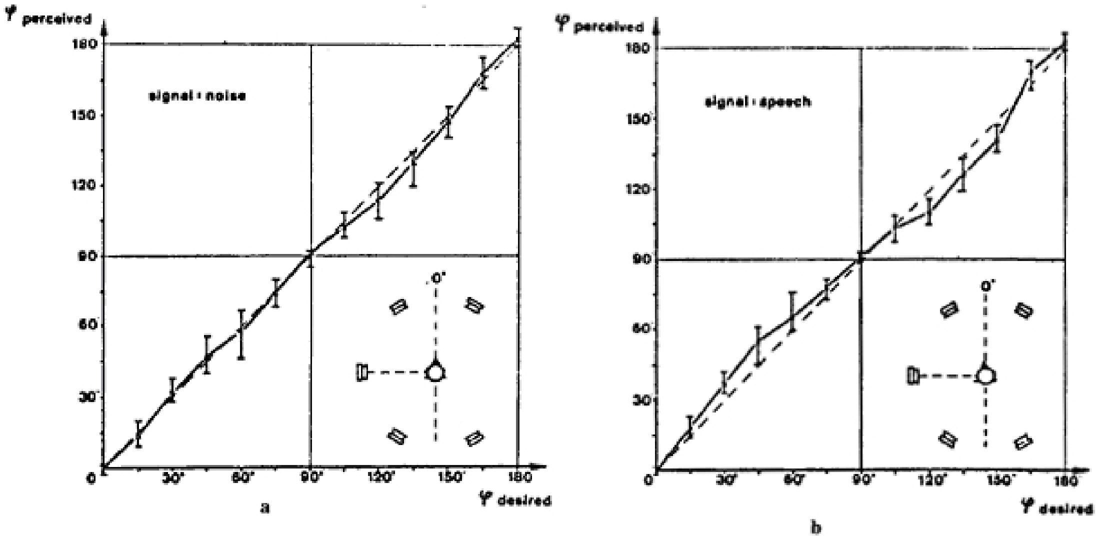

Fig. 6.1 Direction of phantom source versus ILD reported by Theile

Ville Pulkkie carried out a number of experiments to assess the localization accuracy of phantom images created using the VBAP system [Pullkie *et al*, 2001]. A number of different loudspeaker arrangements were examined in a large anechoic chamber using both listening tests and a binaural auditory model which calculated localization cues from the signals arriving at the ear canals. The results indicated that the panning direction accurately matched the perceived direction when the centre of the loudspeaker pair is near the median plane, but degrades as the loudspeaker pair is displaced laterally. In general, a bias toward the median plane was reported for sources produced by laterally biased loudspeaker pairs. Similar results were obtained for a variety of different source signals [Pulkki, 2002]. The results of tests using vertically orientated loudspeaker pairs suggest that the perception of the elevation of amplitude panned phantom images varies widely from person to person. Only if a virtual source is positioned at the same elevation as a loudspeaker can it be reliably localized in that direction.

Grohn also carried out listening tests to assess the localization of both static and dynamic sources in an immersive virtual environment [Gröhn, 2002]. The VBAP system was used with fourteen loudspeakers which were placed in a non-symmetrical arrangement due to the presence of various graphical displays. A variety of source signals were presented while subjects were asked to indicate the position and movement of the virtual source using a tracked baton pointer method. As expected, localization was best when the source was at a loudspeaker position, and worst when panned between loudspeakers. The trajectory of dynamic sources therefore tended to bend toward the loudspeaker positions and similar results were reported for all test signals. In an additional test, it was found that a second distracting stimulus decreased localization accuracy in line with other similar experiments [Blauert, 1997], but only if the distracting signal was at least 10-15dB less than the target signal.

6.2.1 Discussion

The results of the tests discussed in the previous Section clearly suggest that a minimum of six loudspeakers is required to ensure reasonably accurate localization for a single, central listener. In addition, the quality of lateral sources in quadraphonic systems decreases significantly as the size of the array is increased to accommodate a distributed audience (see Section 10.2). Other tests suggest, perhaps unsurprisingly, that localization accuracy increases as more and more loudspeakers are added [Ballas *et al*, 2001].

The performance of stereophonic systems in terms of localization is highly dependent on the position of the phantom image. If the source is panned to a loudspeaker position, then only that loudspeaker will be used, ensuring that localization is as accurate as possible. However, when the source is panned between a loudspeaker pair, localization accuracy degrades significantly, especially in lateral directions. The use of a greater number of loudspeakers would reduce the extent of the degradation in localization accuracy but this solution will not eliminate the resulting panning artefacts which arise for dynamically moving sources. In this case, as the number of contributing loudspeakers is constantly changing, a small yet clearly perceptible timbral shift occurs, which tends to emphasize the loudspeaker positions and distort the perceived trajectory. This effect is almost certainly due to the increased ASW exhibited by stereophonic phantom images compared to a source produced by a single loudspeaker. This occurs due to the slight difference in source direction which is perceived in stereophonic phantom images at low and high frequencies (see Chapter 3, Section 3.4).

6.3 Assessing Ambisonics

J. Daniel carried out a number of experiments to investigate different first and second order ambisonic decoding schemes [Daniel *et al*, 1998]. The decoders were assessed using both HRTF measurements and binaural simulations, and listening tests with four, five and six channel loudspeaker arrays. Under ideal conditions with a single, centrally positioned listener a dual-band decoding scheme was found to be preferable in terms of localization accuracy for the first order system. Similar results were achieved with the second order system using a single band decode which optimized r_V. The results indicated that decoders which perform well in non-ideal conditions (such as an off-centre listener position) are not the optimal low frequency solutions (r_V =1) designed for ideal conditions, but rather the max r_E and in-phase decoders. The tests also indicated that a homogeneous sound field can only be created if there are sufficient loudspeakers for the particular ambisonic order, as discussed in Section 4.4.

Benjamin conducted a number of listening tests to examine the differences between different ambisonic decoders and loudspeaker layouts [Benjamin *et al*, 2006]. First order Ambisonics material consisting of Soundfield recordings, and various encoded test signals were played back using square, rectangular and hexagonal arrays in a medium-sized, acoustically treated listening room. Interestingly, the initial tests were carried out in an ordinary untreated room but were abandoned as extremely poor localization was achieved. Four different decoding schemes were examined, a full-band velocity decode, a full-band energy decode, a full-band in-phase decode, and a dual-band energy/velocity decode based on Gerzon's original design and with a transition frequency of 400Hz. Subjects were free to move around within the array and were asked to listen for attributes such as directional accuracy, perspective, timbral changes, artefacts and loudspeaker proximity effects. Overall, the majority of test subjects preferred the hexagonal array with dual-band decoding. The square layout was least preferred due to poor lateral imaging and spectral changes for different source directions. The rectangular layout was found to work well for frontal sources with rear ambience. In terms of the decoding scheme, the velocity and in-phase decoders were least preferred for opposing reasons. The velocity decoder produced uncomfortable in-head imaging and comb-filtering, probably due to the high frequency anti-phase components which would be present in a full-band velocity decode. The in-phase decoder on the other hand was judged to be much too diffuse and reverberant, although comb filtering and artefacts due to listener movement were eliminated. The full-band energy decoder was judged to provide a balance between these two extremes and was found to work well at off-centre listener positions. However, the shelf filter decoder produced more defined sources as it appeared to pull the various spectral components of the signal to the same perceived direction. An interesting general finding was that the loudspeaker layout is significantly more important than the choice of decoder. The results of the initial failed test would suggest that the acoustics of the listening room are also highly important.

Guastavino conducted a number of listening tests which compared 1D, 2D and 3D ambisonic presentations [Guastavino *et al*, 2004]. In an acoustically treated room containing six symmetrically arranged loudspeakers in a hexagonal formation with two sets of three loudspeakers arranged above and below. The twenty-seven expert listeners subjects were first asked to rate various ambient recordings made with a Soundfield microphone decoded using a full-band in-phase decoding scheme. The test results show a strong preference for the 2D, hexagonal layout in terms of naturalness, source distance, envelopment and timbral coloration. The 3D schemes were described as sounding further away, indistinct and less enveloping while the 1D scheme was found to be the most stable with listener movement. In a second experiment, a more directive decoding scheme (similar to a max r_E scheme) was used as this provided a better balance between localization accuracy and sensitivity to listener position [Guastavino *et al*, 2004]. Similar results were achieved in both tests and an analysis of the results suggests that the preferred layout, at least in terms of naturalness, is dependent on the source material (see Figure 6.2 [Guastavino *et al*, 2004]). The 3D layout appeared to be preferred for indoor environments, while the 2D layout was preferred for outdoor scenes and the 1D scheme for frontal music scenes.

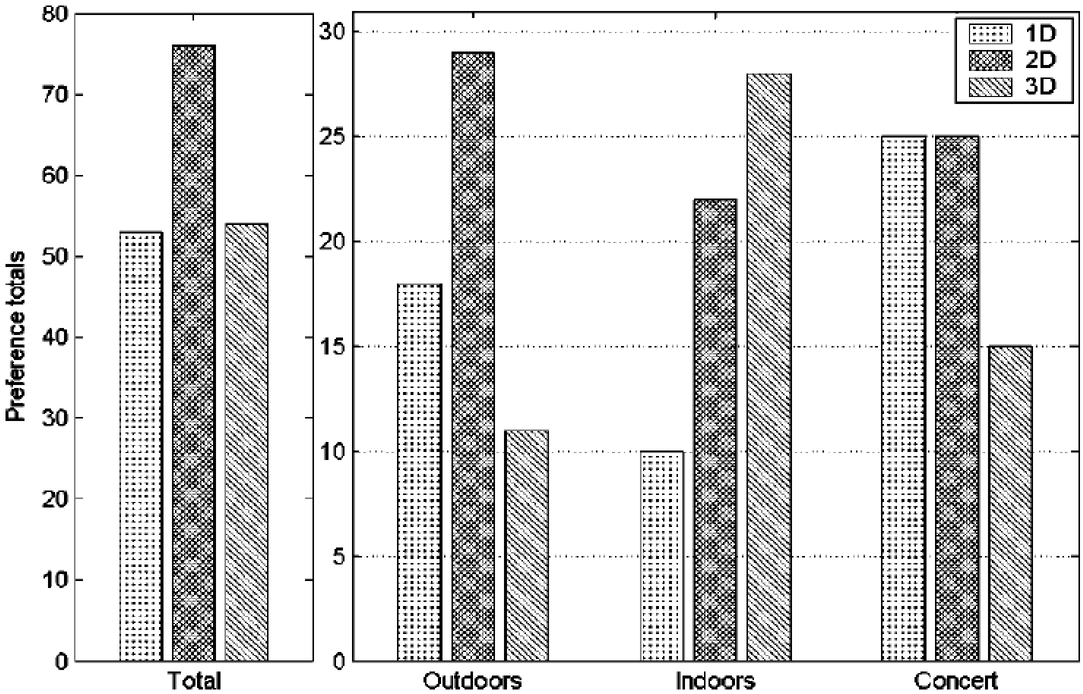

Fig. 6.2 Naturalness responses reported by Guastavino

Kratschmer conducted informal listening tests with a number of ambisonic decoding schemes and a forty-eight loudspeaker array [Kratschmer *et al*, 2009]. The results suggest optimal performance in terms of localization accuracy is achieved when the number of loudspeakers is matched to the Ambisonics order, using the formula given earlier in this section. The results of this test and others [Bertet *et al*, 2009] suggest therefore that the performance of an Ambisonics system decreases significantly when the number of loudspeakers greatly exceeds the minimum number required for that particular order.

David Malham published an informal paper outlined his experience working with large area Ambisonic systems for theatre and music performances [Malham, 1992]. He notes that positioning an audience within a periphonic array can be problematic due to acoustic screening of the lower loudspeakers by other audience members. He also suggests that the decoding scheme and loudspeaker layout often needs to be manually adjusted to compensate for the effect of the room acoustic.

6.3.1 Discussion

The results of these tests confirm Gerzon's original proposal in that a dual-band decoder which optimizes the velocity and energy vectors is preferred when there is a single listener. However, when off-centre listener positions are taken into account, decoders which optimize r_V are least preferred due to the significant anti-phase components which are required to maximize the velocity component. The in-phase decoding scheme eliminates these anti-phase components entirely and so is very stable across a wide listening area, but is also very diffuse. The max-r_E decoder represents a good compromise between these two extremes, particularly at higher orders. As with stereophony, it appears that Ambisonics requires a minimum of six loudspeakers for optimum performance.

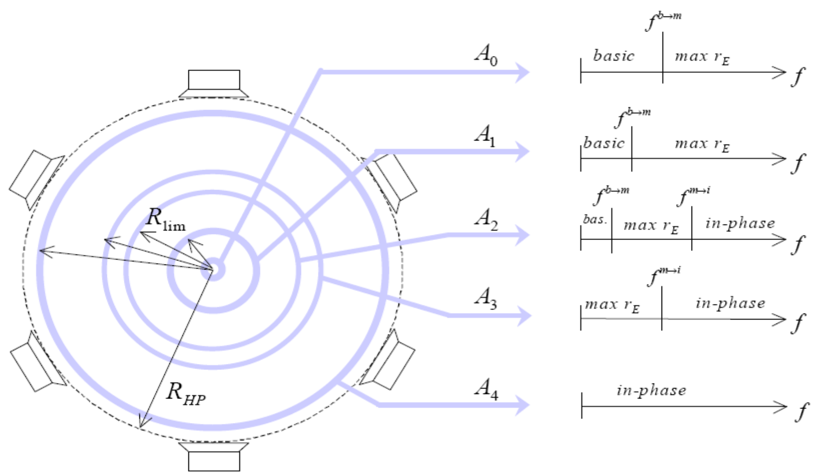

Fig. 6.3 Decoder criteria related to the size of the listening area

Daniel proposes that for a given order and distance from the centre of the array, the max-r_E decoding scheme is most suitable [Daniel, 2000]. If the listening area extends beyond this distance, or to the loudspeaker periphery, then the in-phase scheme is preferred (see Figure 6.3 [Daniel, 2000]). He proposes a tri-band decoding scheme which applies the basic, max-r_E and in-phase decoders in three consecutive frequency bands, based upon the size of the listening area.

6.4 Comparing Stereophony and Ambisonics

Bamford used various analytical techniques to assess the ability of stereo, 5.1 and first and second order Ambisonics to recreate a plane wave [Bamford, 1995]. In general, it was found that the second order Ambisonics system produced the best results and was more consistent for different source angles. The results indicate that the upper frequency at which a plane wave is accurately reconstructed at the centre of the array, is increased when as the order of the Ambisonics system is raised. The performance of both Ambisonics systems deteriorated at higher frequencies but this effect was reduced with the higher order system.

Martin investigated the focus or source width of phantom images produced using amplitude panning, first and second order Ambisonics and some non-standard amplitude panning functions [Martin *et al*, 199b]. Listening tests were carried out in an acoustically treated, medium sized room with a small number of test subjects and a symmetrical eight channel array which was also used to produce interaural cross-correlation measurements. Overall, pair-wise amplitude panning was found to produce the most focussed source as it used a single loudspeaker when the source is positioned at a loudspeaker position, and a maximum of two loudspeakers in other directions. On the other hand, this produced a noticeable change in timbre as the source direction moves across a loudspeaker position. The second order Ambisonics system and a polarity restricted cosine panning function produced the most consistent images across all angles. The polarity restricted cosine function is essentially amplitude panning with a consistent number of loudspeakers, similar to the spread function in Ville Pulkki's VBAP. Martin recommended this algorithm as it produced the most consistent results and appeared less sensitive to listener position than the second order Ambisonics system. It should be noted that very little detail is given in terms of the precise ambisonic decoding scheme adopted. However, Martin mentions that anti-phase components were a disturbing factor in the listening tests with Ambisonics, which suggests that a basic decoding scheme was used which optimizes the velocity component r_V. It is perhaps not surprising, therefore, that the restricted polarity cosine function was preferred, as this method contains no disturbing anti-phase components.

Dickins analysed the performance of first and second order Ambisonics and a custom non-negative least squares (NNLS) panning algorithm by measuring the directional energy vector magnitude r_E [Dickins *et al*, 1999]. Two listening locations have been considered, one in the sweet spot at the centre of the array, and another towards the rear of the array. The tests were carried out in an acoustically treated listening room. Martin suggests that in general a compromise must be made between optimizing the directionality of the source, and minimising panning artefacts as the source moves. The NNLS algorithm is therefore similar to Martin's polarity restricted cosine function and Ville Pulkki's VBAP in that it allows a trade-off between maximum directivity at loudspeaker positions and a more diffuse panning which is homogeneous in all directions. The NNLS algorithm was preferred to the second order Ambisonics system as it functioned well at off-centre listening positions and could be extended to non-symmetrical loudspeaker arrays. However, as with the tests carried out by Martin, little detail is given regarding the ambisonic decoding scheme used. A strict decoding scheme which optimized r_V would be expected to function poorly away from the sweet spot, while a decode which optimized r_E would be much more similar to the NNLS algorithm and would provide a better comparison.

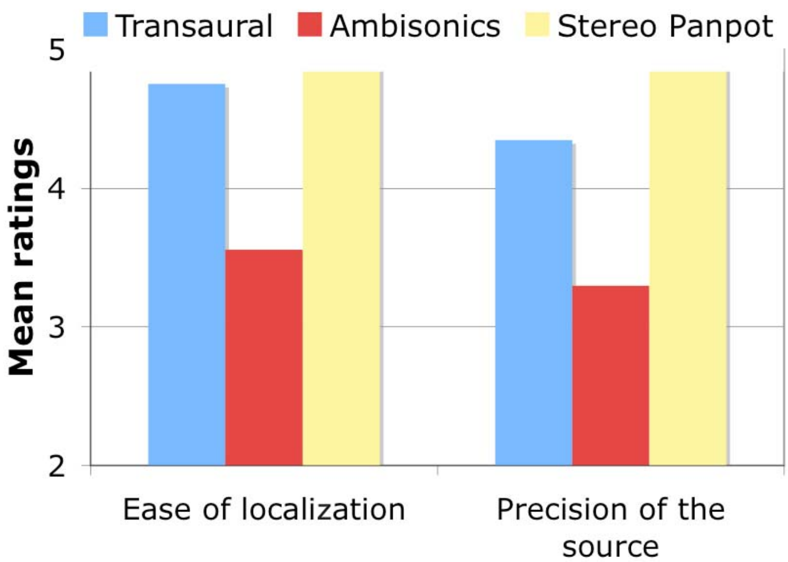

Fig. 6.4 Mean ratings as reported by Guastavino

Guastavino conducted a number of listening tests which compared two-channel stereo, transaural and b-format Ambisonics using six symmetrically arranged loudspeakers within an anechoic chamber [Guastavino *et al*, 2007]. A free verbalization task and a multiple comparison task was conducted with eleven experienced listeners. The subjects were first asked to rate recordings made with a Soundfield microphone, an ORTF stereo pair and binaural microphone in terms of various subjective measures such as envelopment, readability (meaning how well different sources in the scene can be distinguished) and naturalness. and an overall rating. The Soundfield recording was decoded using a single band in-phase decoding scheme. In a second experiment, subjects were asked to rate various monophonic signals positioned at various angles using amplitude panning, a Soundfield recording and a customised Transaural system. The results of the two experiments, shown in Figure 6.4 [Guastavino *et al*, 2007], indicate a strong contrast between Ambisonics and the other two techniques in that stereophony and transaural provide precise localization and a good readability but a lack of immersion and envelopment while Ambisonics provides a good sense of immersion and envelopment but poor localization accuracy and readability of the scene. In a similar study conducted by Capra (see Figure 6.5) slightly better results were achieved for an Ambisonics system implemented using a dual-band shelf filter decoder [Capra *et al*, 2007].

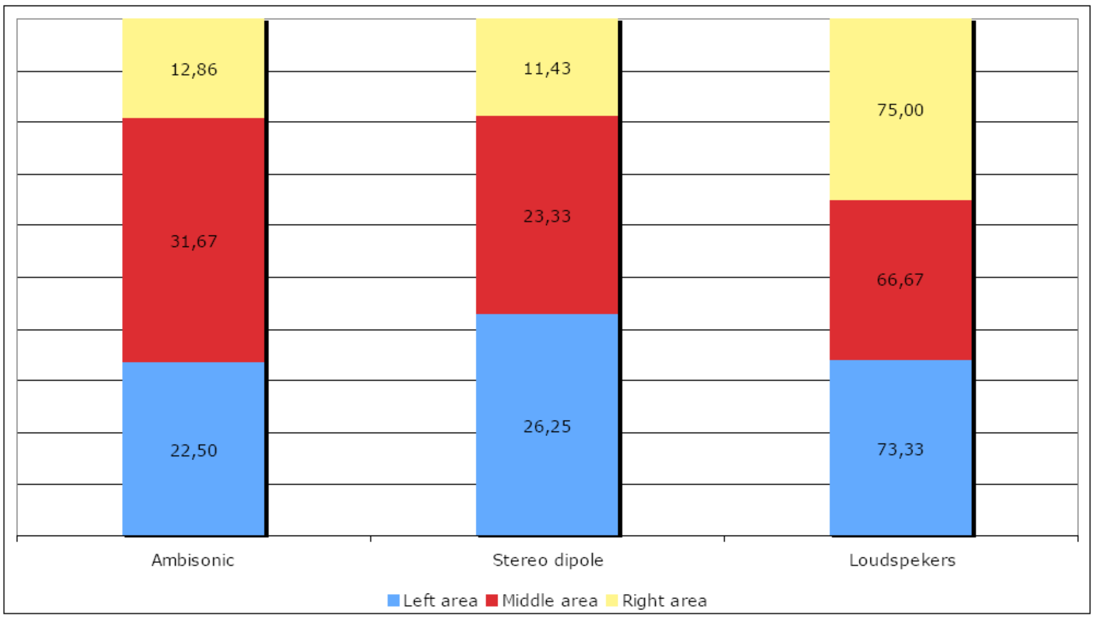

Fig. 6.5 Localization accuracy results as reported by Capra

Pulkki carried out a number of listening tests in order to assess the validity of a binaural auditory model [Pulkki *et al*, 2005]. The experiment was conducted in an anechoic chamber using a symmetrical eight-channel array, first and second order

Ambisonics, amplitude panning (VBAP) and a spaced microphone technique. Although the primary aim of the experiment was to verify the binaural model, some interesting results were obtained. Lateral sources produced with the first order Ambisonics system were found to be consistently biased toward the median plane and exhibited a strong frequency dependence. The second order system exhibited almost no bias, reduced frequency dependence and increased localization accuracy. The best average localization accuracy was obtained with amplitude panning, although the results for second order Ambisonics were equally good for frontal sources.

Jot compared a number of different amplitude panning and ambisonic techniques in terms of a number of objective localization criteria [Jot *et al,* 1999]. The comparisons were based on a single listener at the centre of a hexagonal loudspeaker array. Noticeable improvements are obtained when raising Ambisonics from first-order to second order, with an average performance comparable to that of the six-channel pair-wise panning but more uniform across all azimuths. While amplitude panning produced more stable and accurately localized images at loudspeaker positions, it was less uniform and tended to reveal the loudspeaker locations. No technique produced particularly accurate localization cues at high frequencies.

6.4.1 Discussion

The results of the tests presented in the preceding Section indicate that Ambisonics is consistently preferred to amplitude panning for dynamically moving sources as it produces a more uniform phantom image and hence disguises the loudspeaker position. However, amplitude panning was also consistently preferred for static sources as this method uses fewer loudspeakers and so reduces the localization blur. This would seem to support Martin's view that in general a compromise must be made between optimizing the directionality of the source, and minimising panning artefacts as the source moves [Martin *et al*, 1999a]. The results which indicated that Ambisonics produced a more diffuse enveloping sound field but less tightly focussed sources is arguably another interpretation of the same fundamental difference between the two spatialization techniques.

A number of alternative amplitude panning techniques were presented which attempt to reduce the timbral changes produced when a source is dynamically panned to different positions [Pulkki, 2005; Martin, 1999a, Dickins *et al*, 1999]. These techniques are similar in that they can control the number of contributing loudspeakers independently of the source azimuth. In this way, they can ensure that a minimum number of loudspeakers is always used, which then smoothes the perceived trajectory. This is clearly very similar to the Ambisonics approach of optimizing the energy vector r_E for all directions, at the cost of reducing the maximum localization accuracy that could be achieved at the loudspeaker positions. Pernaux point out that amplitude panning algorithms like VBAP can be considered as analogous to a local ambisonic velocity decode whereby only the loudspeakers closest to the source direction are used, and the requirement to optimize the velocity component ($r_V = 1$) is dropped [Pernaux *et al*, 1998]. They go on to develop a dual-band Vector Based Panning algorithm (VBP) which uses VBAP at low frequencies and a Vector Based Intensity panning algorithm (VBIP) at high frequencies, which like Ambisonics, ensures that $\theta_E = \theta_V$. The significant advantage of this system over Ambisonics is that appropriate decoding factors can be more readily obtained for non-regular loudspeaker arrays.

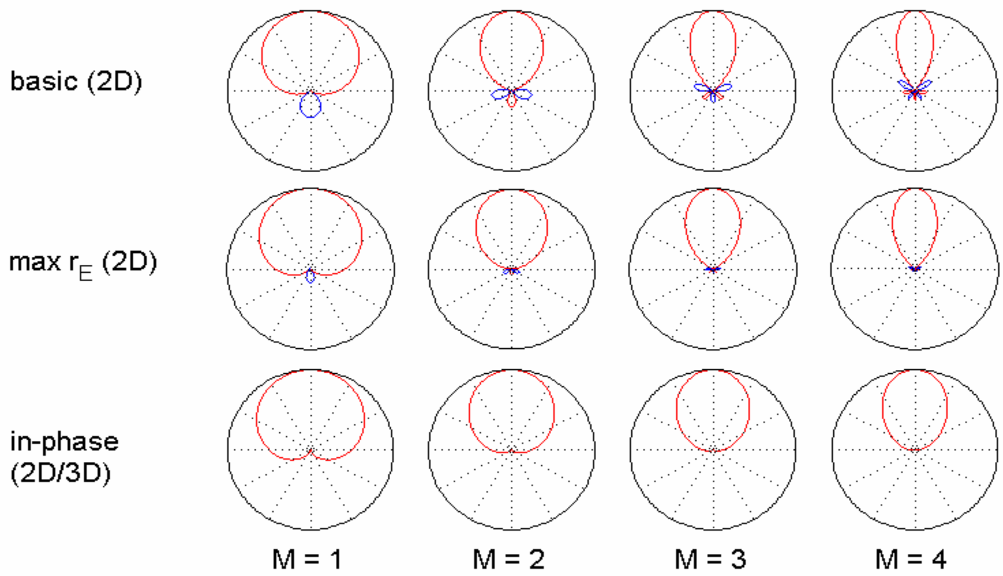

Fig. 6.6 Ambisonic decoder directivity patterns where M is the system order

These advanced amplitude panning schemes are highly reminiscent of Ambisonics, and particularly higher order Ambisonics systems which are optimal for larger listening areas. Higher order systems increase the directivity of the response characteristic (see Figure 6.6 [Daniel, 2000]) which in turn reduces the number of contributing loudspeakers. When this is combined with a decoder which reduces anti-phase components Ambisonics becomes highly similar to these advanced amplitude panning techniques such as VBP. However, some tests have indicated that Ambisonics was still preferred to VBP for moving sources [Pernaux *et al*, 1998].

6.5 The Influence of the Listening Room

The vast majority of the tests discussed so far have taken place in acoustically treated listening rooms or anechoic chambers, thereby eliminating the potentially detrimental influence of the room acoustic on the recreated sound field. This is particularly relevant in the context of spatial music as many performance spaces contain significant lateral reflections and reverberation, and will also contain numerous listeners seated away from the centre of the loudspeaker array. It is known that reverberation and reflections can reduce localization accuracy with real sources [Hartmann, 1983; Hartmann, 1985] and it is reasonable to assume that virtual source localization will be similarly affected. Furlong examined the influence of the listening room and loudspeakers on sound field reconstruction using a coupled room acoustic simulator which allowed user control over such parameters as listening room geometry and reflection coefficients, listener position, and loudspeaker number, position, orientation, and directivity [Furlong *et al*, 1992]. The simulations suggested that placing the loudspeakers too close to either the listeners or the room boundaries will have a significant effect on the perceptual performance of multichannel systems like Ambisonics. In order to verify these simulated results, a series of listening tests was carried out by the author in a small concert hall using various spatialization techniques such as VBAP, Delta Stereo and first and second order Ambisonics.

6.5.1 Spatialization Assessment in a Small Concert Hall

In these listening tests various test signals were presented at eight discrete locations for multiple listeners in an attempt to ascertain the relative level of localization blur [Bates *et al*, 2007b]. The tests were carried out in a small concert hall (shown in Figure 6.7) with a reverberation time (RT60) of 0.9secs at 1kHz, using a loudspeaker array consisting of sixteen Genelec 1029A loudspeakers arranged around a nine listener audience area. In a control test, monophonic signals were played back using the eight black loudspeakers shown in Figure 6.8 while the other ‘dummy’ loudspeakers were used to increase the choice of angle for the listeners [Bates *et al*, 2007b].

Fig. 6.7 Listening tests loudspeaker configuration

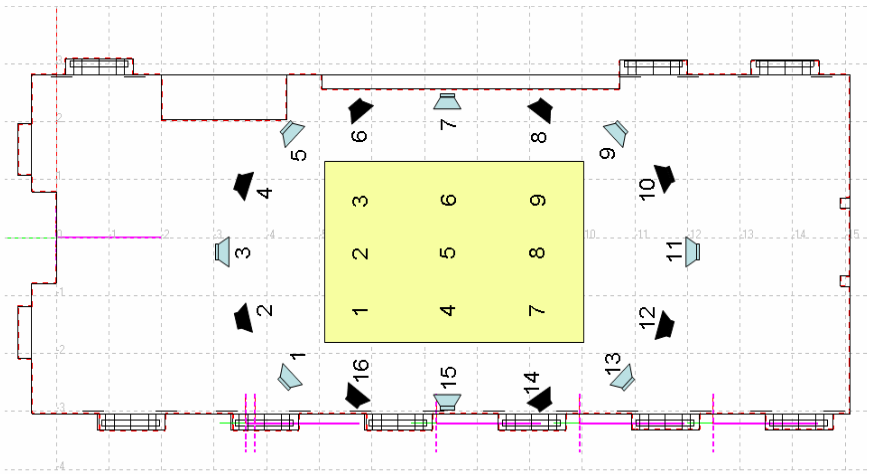

Fig. 6.8 Geometry of loudspeaker array and audience area

Although many spatialization techniques are optimized for regular equidistant loudspeaker arrays, this can be difficult to implement in concert halls which are often rectangular in shape. When the loudspeaker array is squashed to maximize the available seating area in a rectangular shaped room (such as the one used for this test), delay and gain adjustments must be made to displaced loudspeakers. In this test, the appropriate delay was applied to each of the two lateral loudspeakers when encoding the test signals. The gain adjustments were applied to these two loudspeakers by calibrating each loudspeaker in the array to 70dBA at the centre listening position. This approach is preferable to using the inverse square law when operating in a reverberant acoustic environment, due to the superposition of the direct and reverberant sound affecting the total SPL.

In order to assess the effect of various stimuli, users were presented with one second unfiltered recordings of male speech, female speech, Gaussian white noise and a music sample containing fast transients. The results indicate that for most combinations of listening and source position, the influence of the room is not enough to cause a listener to incorrectly localize a monophonic source away from the desired location when using an asymmetrical 8-speaker array. However, for extreme cases such as a front-corner listening position with a source positioned to the rear, the degree of accuracy becomes heavily dependent on the nature of the source signal.

In a second test, first and second order Ambisonics, VBAP and Delta Stereophony were assessed using the same loudspeaker configuration and a forced-choice, speaker identification method. The first order Ambisonics was implemented using IRCAM's Spat software and a traditional, dual-band decoding scheme. The second order system was implemented using the ICST externals for Max/MSP and a max-r_E decoding scheme devised by David Malham specifically for horizontal eight channel arrays [Schacher *et al*, 2006]. Delta Stereophony (DSS) is a sound reinforcement system which has been used successfully in large auditoria. The main objective of DSS is to reinforce the original direct sound while also ensuring accurate sound source localization. This can be achieved if the listener at any place in the room receives the first wavefront from the direction of the sound event being reinforced, rather than from any of the other loudspeaker positions [Fels, 1996].

The results of the listening tests were verified using calculated ITDs inferred from high resolution binaural measurements recorded in the test environment. An equivalent acoustic model was also implemented to investigate specific aspects of the effect of the room acoustic. The results indicate that neither amplitude panning or Ambisonics can create consistently localized virtual sources for a distributed audience in a reverberant environment. Source localization at non-central listener positions is consistently biased toward the nearest contributing loudspeaker in the array, irrespective of the spatialization technique or source stimulus. Both B-format and Second Order Ambisonics exhibited a consistent localization bias towards the nearest loudspeaker and due to the greater number of contributing loudspeakers in these systems, this resulted in significant distortions in the perceived source angles. This can be clearly seen in the results for a source position at loudspeaker 14 which are shown in Figure 6.8.

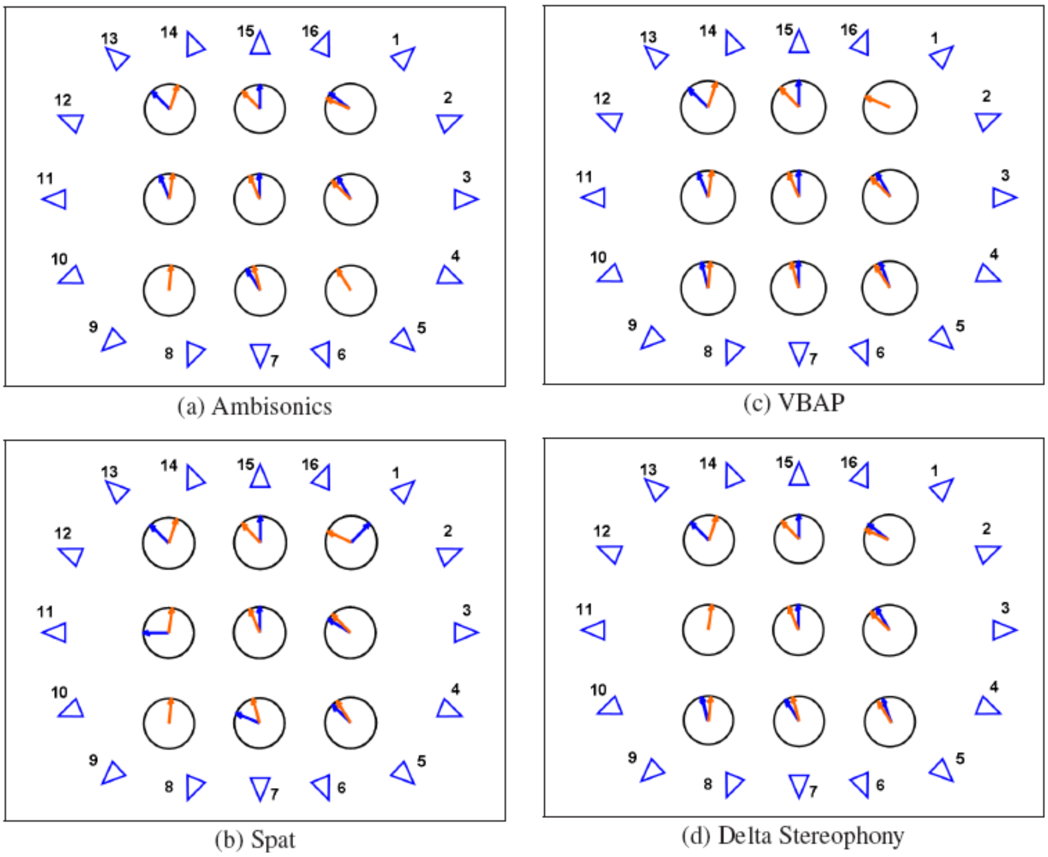

Fig. 6.9 Reported (blue) and actual (orange) direction for a source at speaker 14

An equivalent acoustically modelled environment was implemented using the EASE developer package in order to look at the influence of room acoustical response. The constructed model closely represents the absorption and reverberation characteristics of the real hall. Impulse response measurements were taken using maximum length sequence (MLS) noise at particular points in the hall to verify the accuracy of the model. The measured pressure levels and arrival times of the direct sound and early reflections were found to be comparable to the simulations in EASE. All systems showed small areas where the SPL drops significantly (more than 10dB).

This can be attributed to the superposition of loudspeakers presenting the source material whilst significantly exciting the room modes. Furthermore, each system exhibits large areas where the level of the reverberation is greater than that of the direct sound, which could also adversely affect localization. In all cases, the localization accuracy increases with distance from the source which clearly shows the importance of maintaining sufficient distance between the listeners and the loudspeaker array.

Each system was also analyzed in terms of its subjective *hit rate*, calculated by correlating the ideal localization histogram with the observed results. This measure, shown in Figure 6.9 expresses the percentage localization accuracy over all source positions and stimuli and provides a measure of the overall performance of each spatialization technique. It is clear that the localization performance of each system does not achieve that of monophonic, point sources. Overall, the intensity panning systems perform better for front and back sources, with VBAP providing 12.7% higher localization accuracy over DSS for rear sources. Second Order Ambisonics performs consistently better than B-format at all source positions and its performance is comparable to VBAP for lateral rear sources.

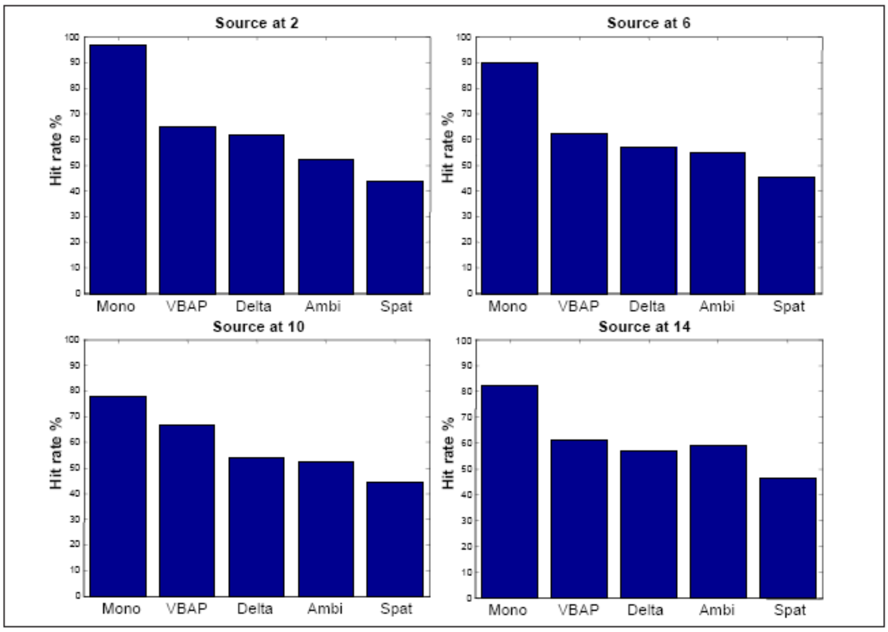

Fig. 6.10 Overall subjective localization performance

The results of this test suggest that accurate localization of phantom sources is difficult to achieve for multiple listeners in a reverberant environment. While acoustic reflections and reverberation in the listening room probably contribute to the overall reduction in localization accuracy, particularly at certain listener locations, the results shown in Figure 6.10 would seem to indicate that this is not the primary negative influence on the performance of the various systems. At locations away from the centre of the array, the temporal relationship between the contributing sound waves is distorted and so the precedence effect effectively dominates source localization. The proximity of the loudspeaker array to certain listeners results in phantom images which are consistently localized to the nearest contributing loudspeaker. The best results are therefore achieved with VBAP, as this system only ever uses a maximum of two loudspeakers, while the worst results were achieved with b-format as this system uses nearly all of the loudspeakers to produce a phantom image. The increased directivity of the second order Ambisoncs, max r_E decoding scheme performs better than the first order system particularly at off-centre listener positions.

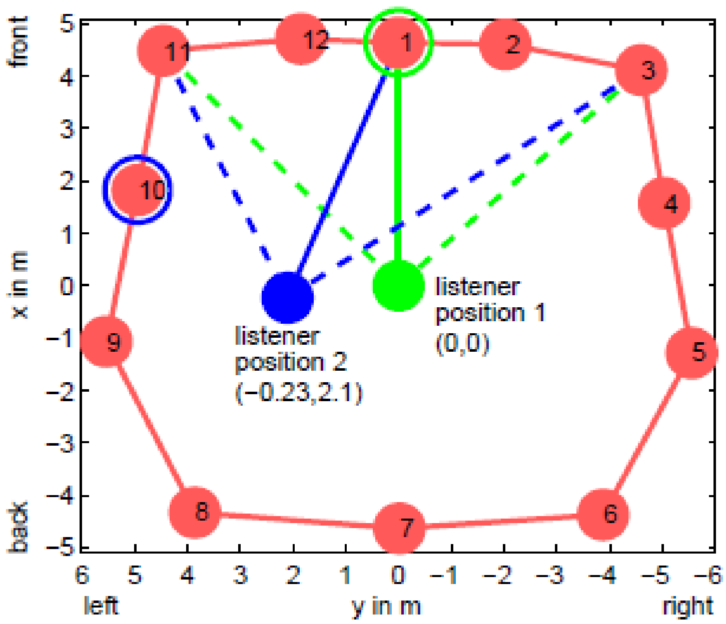

Fig. 6.11 Ambisonics system evaluated by Frank

Frank *et al* investigated the subjective localization accuracy of a virtual source created using the twelve channel, nearly circular 2D Ambisonics system shown in Figure 6.11 [Frank *et al*, 2008]. The room used for the test is a medium sized 12m x 10m room with some acoustic treatment. First, third and fifth order Ambisonics decoders were examined using basic, max-r_E, and in-phase decoding schemes. As shown, test subjects were placed at two different listening positions. The results indicated that localization improves as the order increases, independently of the listener position. As expected, localization at the central listening position is more accurate than at the off-centre position. The max r_E decoder was found to provide the best overall performance while the in-phase decoder was least preferred. Localization was biased toward the left side, presumably due to the closer proximity of the listener to the loudspeakers at the off-centre listener position. This proximity bias was significantly reduced as the order increased. Interestingly, the delay adjustments applied to compensate for the non-equidistant loudspeaker layout were found to be detrimental to the perceived performance of the system. The authors suggest that this unexpected result occurred due to pronounced phase distortions outside the sweet spot. However, the exact reason why the uncompensated system performed better is still uncertain.

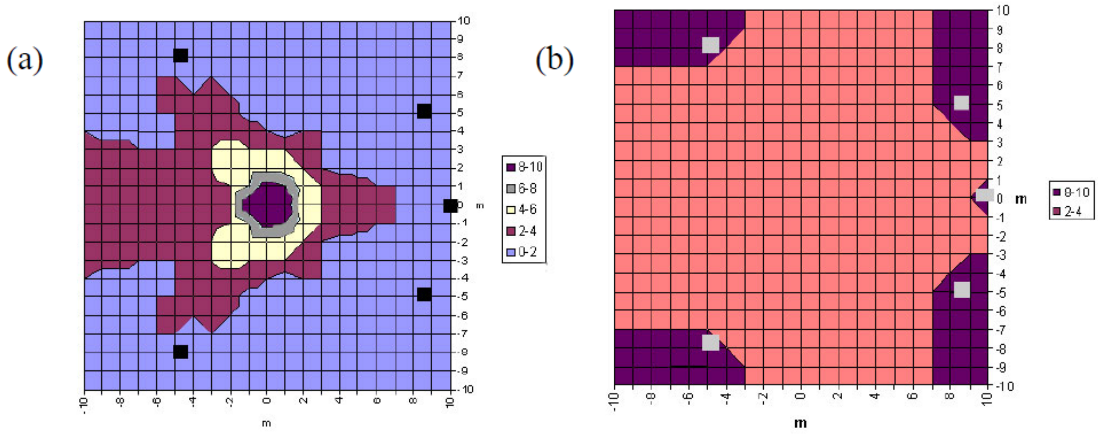

Fig. 6.12 Quality maps for 5ms (left) and 50ms (right) time difference

A large number of experiments have been carried out to examine the influence of the precedence effect [Wallach *et al*, 1949; Litovsky *et al*, 1999]. It has been found that a number of correlated sound waves arriving in close succession will be fused together and perceived as a single sound at a single location. The duration over which fusion will take place is highly dependent on the nature of the source signal, but studies have found an approximate lower limit of 5ms for transient sounds and an upper limit of 50ms for broadband sounds [Litovsky *et al*, 1999]. Moore used these limits to evaluate a number of listener positions with the ITU 5.1 layout in terms of the precedence effect. By checking the time difference between each loudspeaker pair, a mark out of ten was given for each position in the listening area allowing performance to be quantified at different positions in the reproduction area [Moore *et al*, 2007]. The results for a reproduction area of $20m^2$ are shown in Figure 6.12.

They indicate that transient signals will only be correctly localized in a relatively small central listening area due to the large travel distances involved in such a large array. A large effective listening area is shown for a 50ms time difference and the author suggests that the effective listening area could be extended using a form of transient suppression applied to the output signals [Moore *et al*, 2007].

It has been suggested that altering the directivity characteristic of the loudspeakers can help to reduce the detrimental effect of the room acoustic on stereophonic source localization. The results of a series of listening tests carried out by Harris suggested that diffuse acoustic radiators such as Distributed-Mode Loudspeakers (DML) reduce the degradation caused by room acoustics on stereophonic localization [Harris *et al*, 1998]. Zacharov examined the effect of loudspeakers with different directivity characteristics in terms of source localization, envelopment and naturalness [Zacharov, 1998]. In a number of listening tests more directional loudspeakers were preferred for all parameters and for all listening positions. Zacharov suggests that the increased directivity of these loudspeakers reduces the excitation of the listening room modes and hence reduces the negative impact of the room acoustic on stereophonic reproduction. However, further tests are required to see if a similar approach could improve the performance of Ambisonics under similar conditions.

It is clear that unless a certain minimum distance is maintained between the listeners and the loudspeakers, the precedence effect will ensure that localization will collapse to the nearest contributing loudspeaker. While it is clear that the reverberation and early reflections in the listening room also influence the perception of virtual sources, the extent and exact nature of this influence is more difficult to define as this will entirely depend on the exact dimensions and layout of the particular space.

6.6 Artificial Reverberation and Spatial Dissonance

The preceding chapter illustrated how the simulation of distance is primarily achieved through the addition of artificial reverberation and early reflections. It has been shown that the acoustics of the listening room affect auditory perception and so it must be expected that the addition of artificial reverberation will also have some influence on the perceived source. Begault investigated the effect of synthetic reverberation on three dimensional audio samples synthesized using HRTF data [Begault, 1992]. He found that the error between perceived and actual source angle increased with the addition of reverberation and suggested that this was due to the effect of early reflections, in line with earlier studies [Hartman, 1985]. The directionality of early reflections was also investigated but the results found no consistent relationship between the angle of early reflections and perceived source angle. A consistent relationship between reverberation and source distance was found and the addition of artificial reverberation dramatically increased the effective externalization of the synthesized sound scene. Although the results of this study using binaural techniques cannot be directly compared to multi-channel loudspeaker systems, it does seem to agree with the results of other studies and suggests that there is a certain trade-off between the simulation of distance cues and the production of a reliable source direction.

Martin investigated the influence of the angular reflection pattern of artificial early reflections and reverberation [Martin *et al*, 2001]. The reflections of any wave can be classified into two groups depending on the precise angle of reflection. If the reflective surface is large and flat relative to the wavelength of the reflected sound, then specular reflections will be produced according to a predictable angle as shown in Figure 6.13 (left). If the surface is irregular, then the incident wave is scattered and multiple diffuse reflections are produced instead of a mirror image reflection, as shown in Figure 6.13 (right).

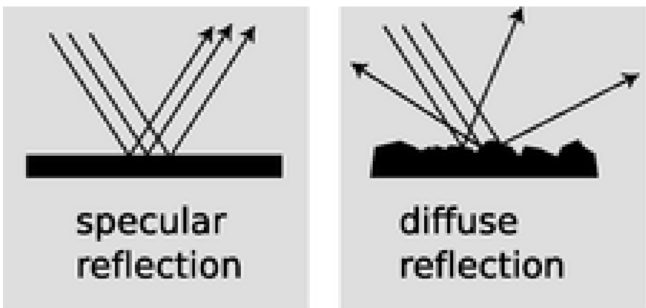

Fig. 6.13 Specular and diffuse reflections

The results of an initial test indicated that there was an easily recognizable difference between signals processed using a specular and diffused reflection model, whether a reverberant tail was included or not. In general, a mix of specular and diffused reflection models was found to be preferable to a typical perfectly specular reflection model. Martin also suggests that the inclusion of a height component in a synthetic room model improved the attractiveness and realism of the resulting sound, even when reproduced using a two-dimensional loudspeaker configuration [Martin *et al*, 2001].

It is still unclear how the addition of artificial reverberation affects auditory perception at non-central listener positions. Lund developed a 5.1 surround sound spatialization system based on amplitude panning with suitable spatially distributed early reflections and diffuse reverberation [Lund, 2000]. He suggests that the localization of virtual sources will be made more robust and accurate if the desired source location is supported by accurately modelled early reflections and reverberation. The results of a listening test with a very small number of participants (five listeners) seem to support this view as the system produced more consistently localized phantom sources and greatly increased the effective listening area [Lund, 2000]. However, further tests with a greater number of subjects are required to fully verify these results.

Michelsen found that the addition of artificial reverberation and early reflections produced a clear and unchanged sense of distance in both an anechoic and reverberent listening room [Michelsen *et al*, 1997]. This suggests that the listening room acoustic is not significantly detrimental to the artificial simulation of distance cues. However, the perceptual effect of combining artificial reverberation with the natural reverberation of the listening room itself is unknown. The composer Denis Smalley suggested that this would produce a negative cognitive effect, which he refers to as spatial dissonance, as the auditory system would be presented with localization cues that suggest two, potentially conflicting, acoustic spaces. However, very few listening tests have been undertaken to ascertain how detrimental this effect actually is. In the author's experience, the layering of two synthetic reverberation effects does not create two clearly perceivable acoustic spaces but instead results in one, slightly indistinct space. Clearly, the simulation of distance is highly dependent on the use of artificial reverberation and it is hard to see how else this effect could be implemented other than the physical placement of loudspeakers at different distances. As discussed in Chapter Three, wavefield synthesis differs from stereophony and Ambisonics in that it can theoretically recreate virtual sources at different distances by synthesizing the correct wavefront over an extended area. The evaluation of this system and the veracity of these claims will be examined in the next section.

6.7 Evaluating Wavefield Synthesis

It is well known that the distance between adjacent loudspeakers in a WFS system is directly related to the spatial aliasing frequency [Berkout, 1998]. This frequency is one of defining characteristics of any WFS system and represents the upper frequency limit for accurate reconstruction of the synthesized wavefront. Clearly, determining a suitable lower limit for the spatial aliasing frequency is an important area for investigation as this directly influences the necessary size of the loudspeaker array, and hence the listening area. In addition, while various techniques exist to reduce high frequency distortion in WFS virtual sources, it is important to assess the perceptual effects of these techniques. For example, while certain techniques reduce the timbral coloration introduced by this system, this can also influence the directional localization.

The ability to position virtual sources behind and in front of the loudspeaker array is an often lauded feature of WFS and this inherent ability to reproduce distance is unique. However, much of these claims are based on simulations, such as the example shown in Figure 3. 11 and subjective listening tests are required to verify these claims.

Finally, the influence of the listening room acoustic has had a significant and detrimental effect on other spatialization systems and it is expected that WFS will be no different.

6.7.1 WFS – Localization Accuracy and Spatial Aliasing

One of the primary goals of any spatialization system is to ensure that each listener can clearly and accurately perceive the direction of the virtual source, in any direction. Most practical systems eliminate the height component and so simply require a number of loudspeakers to be placed around the audience. WFS differs from other systems in that it is based on linear arrays of loudspeakers rather than a small number of distributed loudspeakers. A WFS system which can position a virtual source in any direction in the horizontal plane requires therefore a very large number of loudspeakers placed in a linear array surrounding the audience. This was demonstrated by Verheijen who found that accurate synthesis is not possible for sources which do not lie in a straight line from the listener to the source, through the array [Verheijen, 1998]. The only area in which a virtual source can be correctly synthesized is between two lines from the listener to the edges of the array, as shown in Figure 6.14. This area is further reduced if tapering windows are applied to the array to reduce truncation effects (see Section 3.2.2).

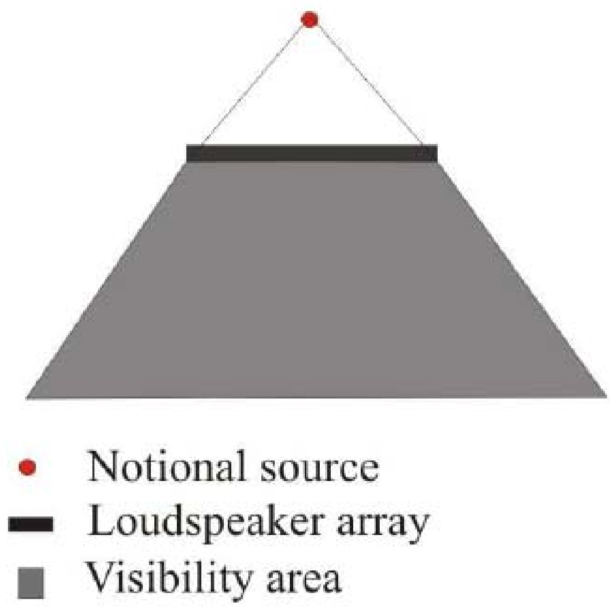

Fig. 6.14 Visibility of a virtual source in a WFS system

WFS can only accurately synthesize a sound field up to the spatial aliasing frequency, so clearly the high frequency content will be distorted in some way. Start reported that the imperfect reconstruction of the sound field above the spatial aliasing frequency gives rise to an increase in the apparent source width (ASW) due to the uncertain directionality of the high frequency content [Start, 1997]. Wittek points out that this must be taken into account when evaluating the localization accuracy of any WFS system [Wittek, 2003]. For example, a measure of the standard deviation in the reported directional data will not indicate whether listeners perceive a tightly focussed source within the range of directions reported, or a broad diffuse source distributed between the range of reported angles. As discussed earlier in Chapter Two, locatedness is a measure of the degree to which an auditory event can be said to be clearly in a particular location [Blauert] and this parameter is sometimes used in the assessment of WFS systems to determine the focus or apparent width of the virtual source.

Some of the earliest perceptual experiments with WFS systems were carried out by Vogel [Vogel, 1993]. In an experiment with an array of twelve loudspeakers each separated by 45cm, he found that correct directional localization was maintained despite the very low spatial aliasing frequency of 380Hz of this system. However, Wittek points out that as this system can only correctly synthesize frequencies below 380Hz, it cannot be assumed that WFS is responsible for the correct localization [Wittek, 2003]. For a WFS virtual source positioned behind the array, the loudspeaker nearest to a direct line from the listener to the source will be producing the earliest and often the loudest signal. The precedence effect would therefore provide a localization cue at all frequencies, which, in this case, coincides with the source position specified by the WFS system. The mean directional error reported in Vogel's test is no lower than what would be expected if the precedence effect was dominating localization and so these results do not indicate that localization accuracy is improved by this particular WFS system. This situation does not occur with focussed sources in front of the array, as in this case the first wavefront does not arrive from the same direction as the virtual source [Wittek, 2003].

Since these early tests further experiments have been carried out with loudspeaker arrays of greater size and resolution [Vogel, 1993; Huber, 2002]. The results of these tests demonstrated that localization accuracy was greater than the actual physical resolution of the loudspeaker array, no doubt due to the increased resolution of the array and the associated increase in the spatial aliasing frequency. The results of both these tests clearly indicate the importance of the spatial aliasing frequency in terms of the performance of the WFS system.

Start compared the minimal audible angle (MAA) of real sources and virtual sources produced using a WFS system [Start, 1997]. He found no difference between the MAA of a real source and the WFS source for a spatial aliasing frequency of 1.5kHz, for both broadband and low-pass-filtered noise signals. When the spatial aliasing frequency was reduced to 750Hz however, the MAA increased somewhat. Start suggested that this result implied that a spatial aliasing frequency of 1.5kHz would ensure that the dominant low frequency localization cues are satisfied and so the source will be accurately localized.

Huber conducted listening tests in an anechoic chamber to compare real sources, two-channel stereo, WFS with loudspeaker spacings of 4cm and 12cm, and an augmented WFS system based on Optimized Phantom Source Imaging (OPSI) [Huber, 2002]. OPSI uses amplitude panning to position the portion of the signal which lies above the sampling aliasing frequency [Wittek, 2002], thereby reducing the artefacts which occur due to the incorrect reconstruction of the high frequency signals in WFS systems. Figure 6.15 shows the scaled judgements of locatedness (which Huber refers to as localization quality) for each of the five systems. Clearly none are able to match the real source in terms of locatedness, however, a significant improvement is apparent when the loudspeaker spacing is reduced to 4cm (which results in a aliasing frequency of 3kHz). The worst results were achieved with stereo while the hybrid OPSI method was found to produce approximately the same results as the normal WFS system. Interestingly, the standard deviation in localization accuracy shown in Figure 6.16 does not indicate any differences between the real and WFS sources, which indicates the importance of assessing the perceptual sense of locatedness as well as the perceived direction.

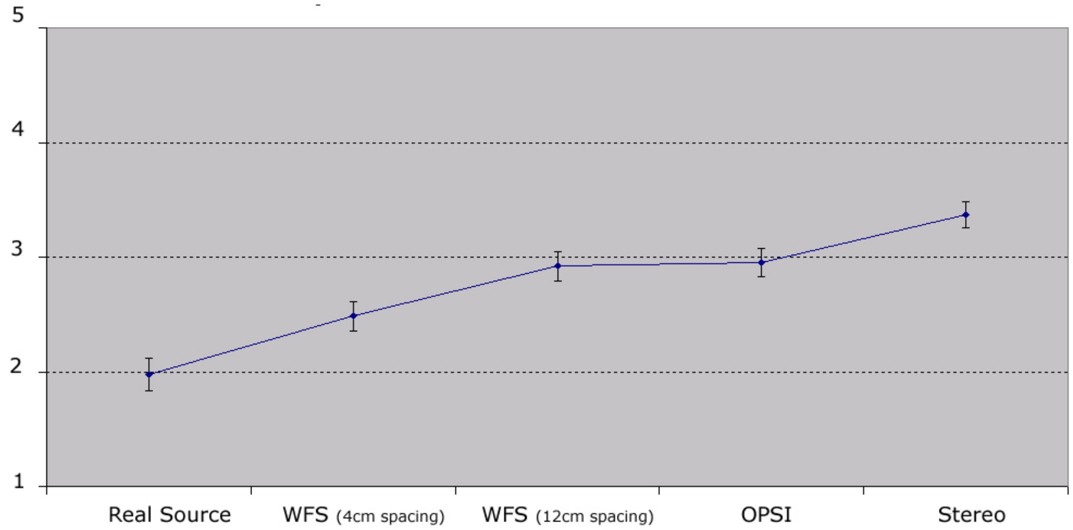

Fig. 6.15 Subjective, scaled (1-5) judgments of locatedness reported by Huber

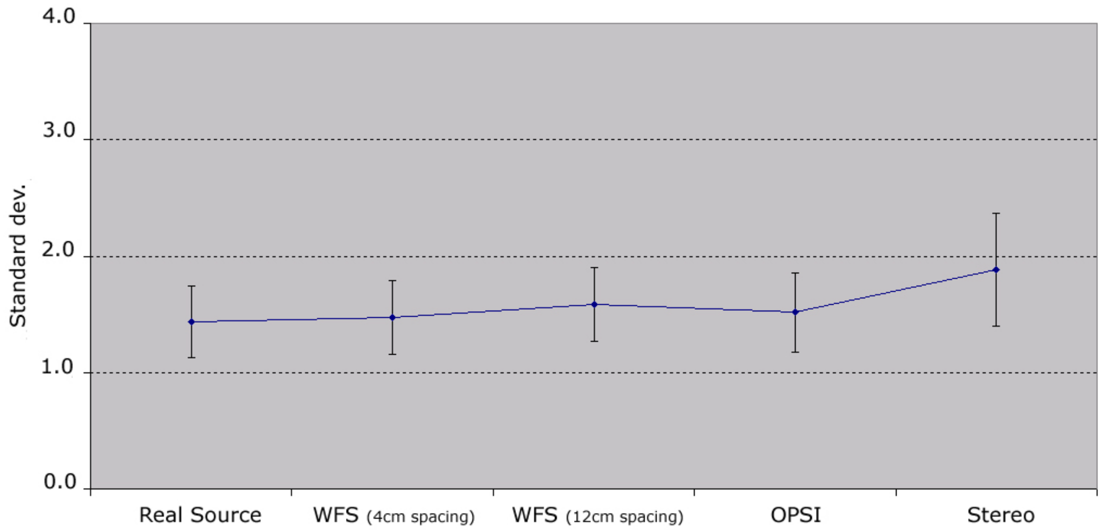

Fig. 6.16 Standard deviation in horizontal localization reported by Huber

A similar comparison test was carried out by Wittek on the localisation performance of various WFS systems compared to a real source, amplitude panning and a hybrid OPSI system (see Figure 6.17) [Wittek, 2007]. For the WFS systems, a virtual source was positioned one metre behind the array using loudspeaker arrays with a spatial aliasing frequency of 2.5kHz (labelled WFS12) and 7.5kHz (labelled WFS4) respectively. The OPSI signals were produced using the WFS12 system and three loudspeakers at spacings of 76cm to produce the phantom source with a crossover frequency of 2kHz.

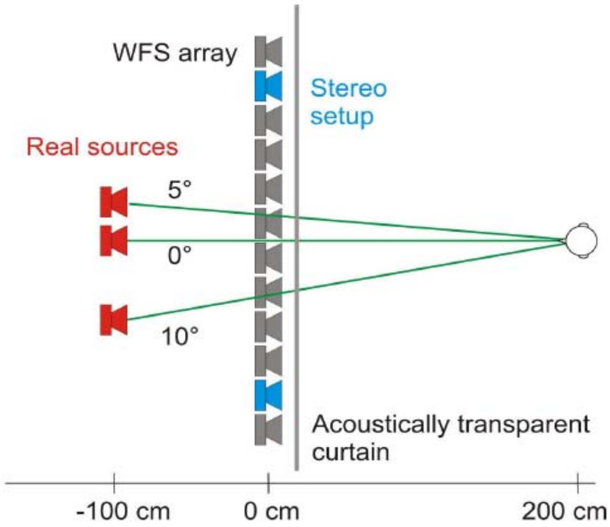

Fig. 6.17 Test setup for Wittek's listening tests

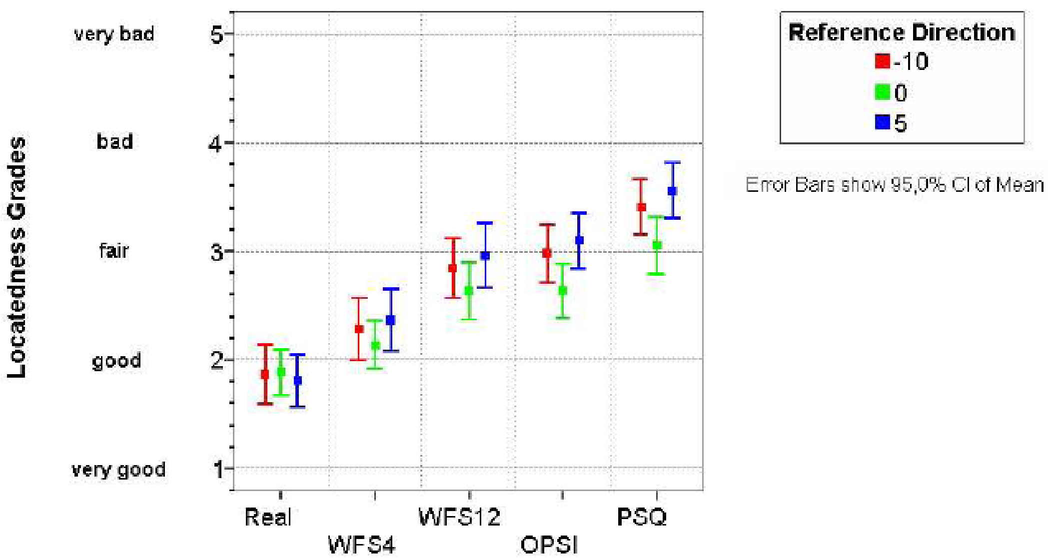

Fig. 6.18 Subjective assessment of locatedness reported by Wittek

The results were similar to Huber's in that none of the systems matched the performance of a real source in terms of locatedness, but better results were reported with the WFS system than with amplitude panning (see Figure 6.18). As with Huber, better results were achieved when the spatial aliasing frequency was increased from 2.5kHz to 7.5kHz. Similarly, these differences are not evident when only the standard deviations of the measured auditory event directions are considered. No degradation in localization quality was found using the hybrid OPSI method.

Sanson examined localization inaccuracies in the synthesis of virtual sound sources using WFS at high frequencies [Sanson *et al*, 2008]. Objective and perceptual analyses were carried out through a binaural simulation of the WFS array at the ears of the listener using individual head related transfer functions (HRTFs). The array could be configured for loudspeaker spacing of 15cm, resulting in an aliasing frequency around 1500Hz, or a loudspeaker spacing of 30cm results in an aliasing frequency around 700Hz. Two listener positions were evaluated, one central and one laterally displaced to the right by 1m. The results of the test indicated that localization accuracy was dependent on the listening position, the source position and the frequency content of the source signal. Localization accuracy decreased as the listener position moved laterally away from the centre point, As the source cut-off frequency was increased, localization at the off-centre position degraded, but not at the central listening position. The authors suggest that this is due to the unequal distribution of high frequency content at either ear when the listener is positioned at a non-central location. This would provide a conflicting localization cue relative to the low frequency content which is accurately reproduced by the WFS system. Clearly, technical solutions to the distorted high frequency content in WFS systems must address localization for off-centre listener positions.

6.7.2 WFS in Real Rooms

Start carried out a number of experiments with a WFS system in various different rooms such as an anechoic chamber, a medium sized auditoritum and a large concert hall (see Figure 6.19) [Start, 1997]. The test signals consisted of speech, broadband noise and low and high pass filtered noise (below and above the spatial aliasing frequency). The author made the following observations based upon the results of a number of listening tests in each room:

– The averaged RMS error for the low-pass-filtered noise signal is almost identical for the real and virtual sources in the anechoic room but the results for a real source are somewhat better in the auditorium and concert hall.
– The best results were achieved with the speech signal
– The averaged RMS error for the high-pass-filtered noise signal is much larger for synthesized sources.
– The localization performance of the WFS system is worse in the concert hall while real sources are localized to the same degree of accuracy as the other rooms.

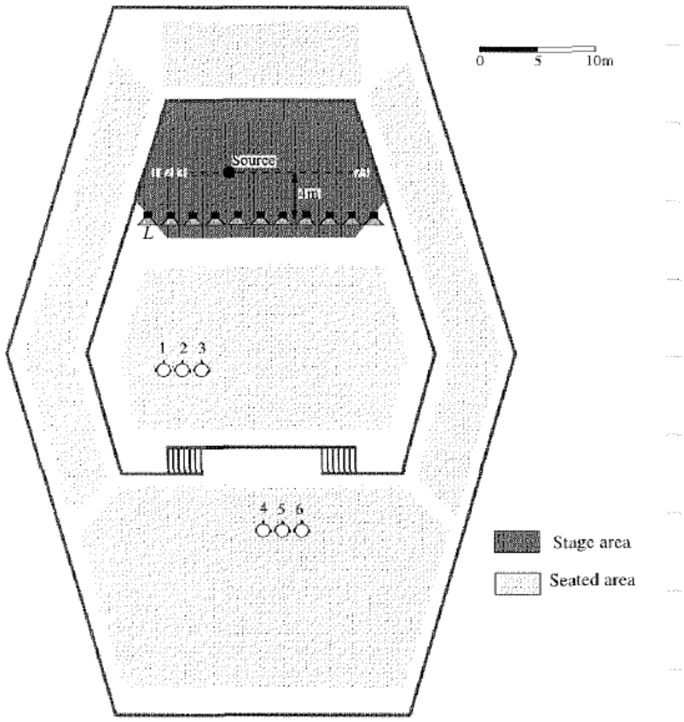

Fig. 6.19 Concert hall layout used by Start

Start suggests that the reduction in localization accuracy of the WFS system in the concert hall is primarily due to the lower spatial aliasing frequency (750Hz) of this system. The standard deviation of the averaged perceived direction for each room is shown in Figure 6.20. In the anechoic room and the auditorium a clear difference can be seen in the results for the low and high pass filtered noise. In each case, the low frequency signal (<1.4kHz and <1.2kHz, respectively) whether real or synthesized, is localized more accurately than the high-pass-filtered signal. The localization accuracy with synthesized sources is particularly worse with the high-pass-filtered signal. However, this is not the case in the concert hall as here the worst results are achieved with the low-pass-filtered signal (now <750Hz) for both real and synthesized sources. These results suggest that the WFS system is not working correctly in this particular room. Start suggests that this is solely due to the decrease in the spatial aliasing frequency. However, the drastic reduction in performance for a low-pass-filtered noise signal results suggest that the performance of the WFS system is also significantly affected by the reproduction room acoustic.

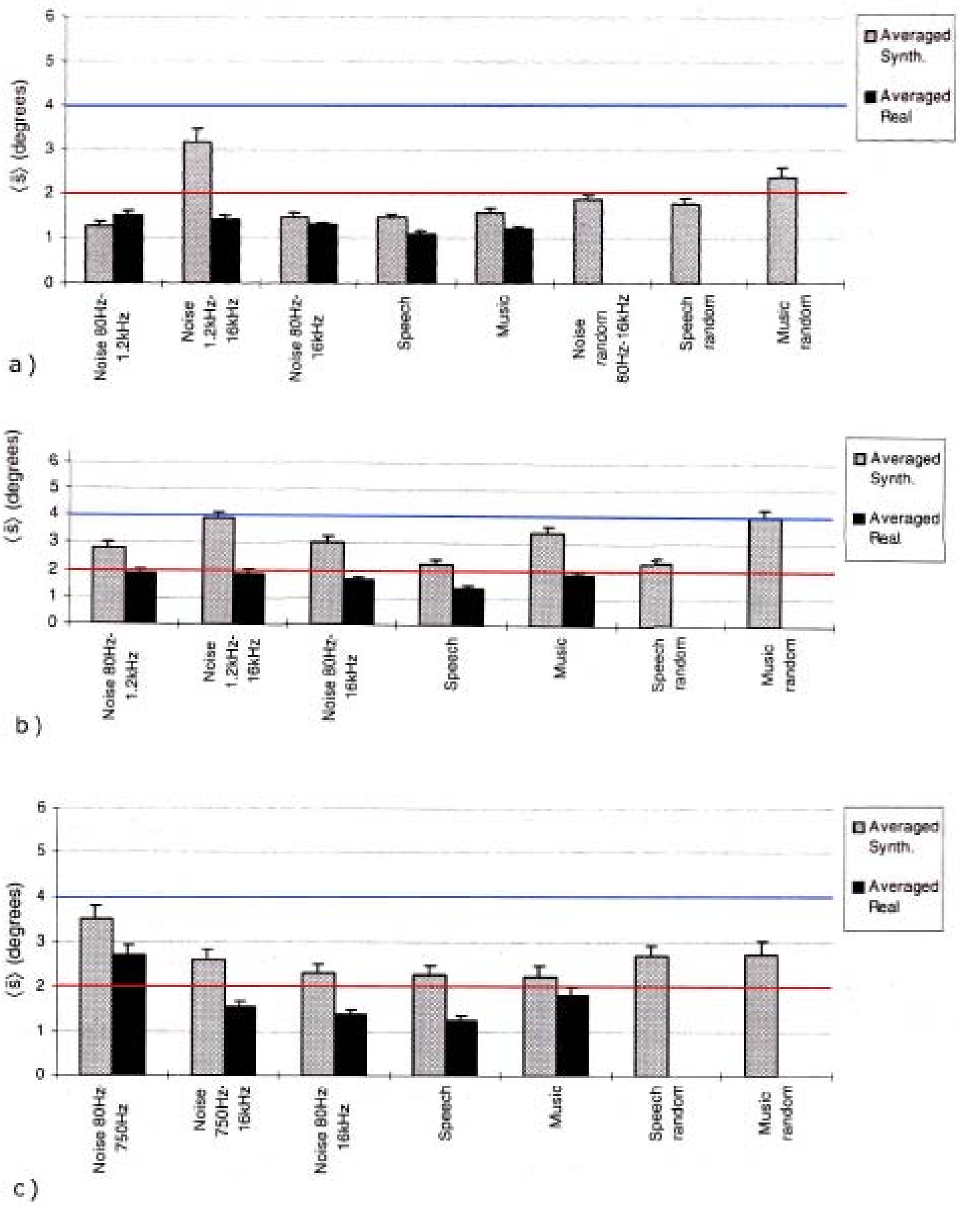

Fig. 6.20 Perceived WFS virtual source directions reported by Start in

a) Anechoic Chamber

b) Auditorium, Delft University of Technology

c) Concert hall 'De Doelen', Rotterdam [Start]

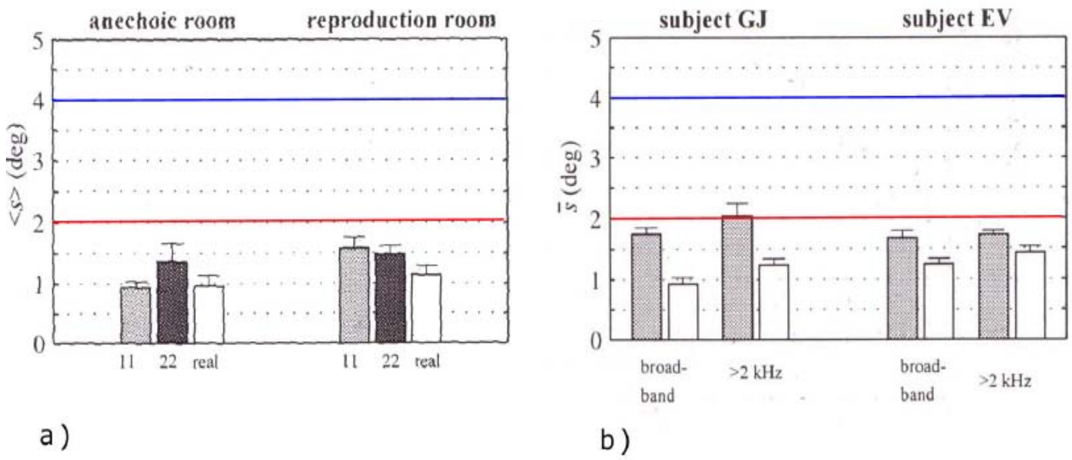

Fig. 6.21 Virtual Source (a) behind the array & (b) in front of the array
gray bars: virtual sources (loudspeaker spacing = 11cm)
black bars: virtual sources (loudspeaker spacing = 22cm)
white bars: real sources [Verheijen]

A similar experiment was carried out by Verheijen using two different arrays with spatial aliasing frequencies of 0.75kHz and 1.5kHz respectively, and virtual sources behind and in front of the array [Verheijin, 1998]. The mean standard deviation results, shown in Figure 6.21(left), indicate an improvement with the increased array resolution in the anechoic room but not in the reproduction room. This again seems to suggest that the influence of the reproduction room acoustic is as significant a factor as the spatial aliasing frequency. The results for two test subjects and a virtual source in front of the array are also shown in Figure 6.21(right). This test was carried out in the anechoic room and a comparison of the results with a source behind the array suggests that localisation accuracy decreases when the source is positioned in front of the array.

Marentakis examined localization accuracy with WFS in a variable-acoustics concert hall using the Minimum Audible Angle (MAA) as a measure of localization performance [Marentakis *et al*, 2008]. The MAA was estimated for different listener positions, listener orientations and varying acoustical conditions. The large hall could be configured to adjust the reverberation time from 0.4 to 4sec and in this test was set to an absorptive setting with a relatively low reverberation time, and a reflective setting which increased the energy of side reflections and the overall reverberation time. A 48-channel WFS system was placed in the hall at 7m from the rear wall and 4m height, as shown in Figure 6.22. A double logarithmic loudspeaker spacing was utilized so that central loudspeaker spacing (24cm) was greater than lateral loudspeakers (13cm). This unusual arrangement results in an aliasing frequency which varies with the source angle and listener position, as shown in Figure 6.23.

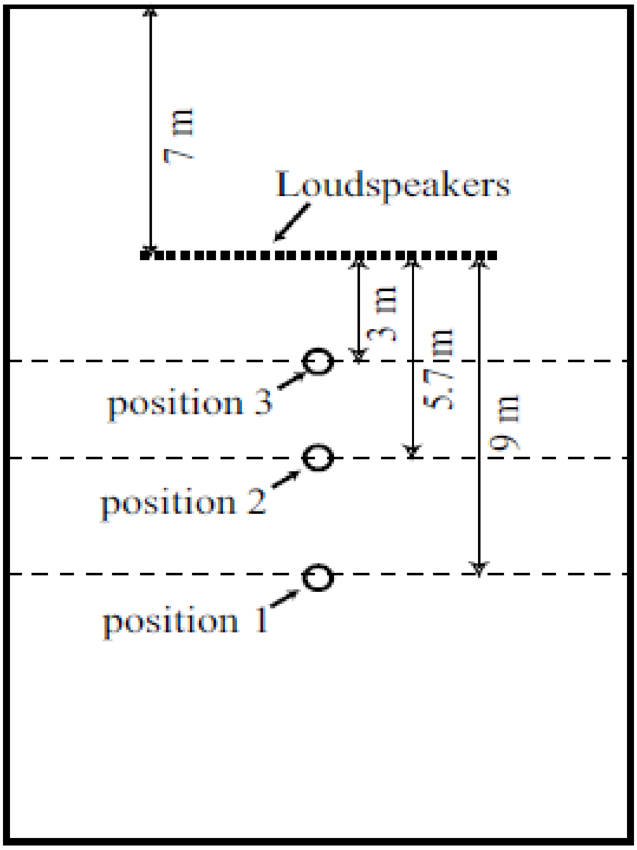

Fig. 6.22 Test setup for Marentakis's listening tests

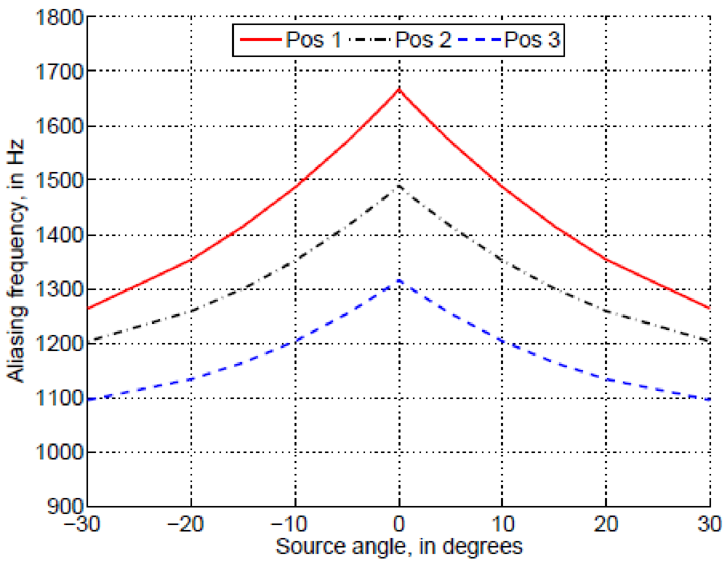

Fig. 6.23 Aliasing frequency relative to listener position reported by Marentakis

The results illustrated the general and expected trend that localization ability decreases as the listener orientation changes from frontal to lateral orientation. This is particularly evident for the listener position closest to the loudspeaker array, which suggests that there is a limit to how close to the loudspeakers listeners can be placed without a significant decrease in localization accuracy. The best results were obtained for the central listener position while a slight reduction in accuracy was reported for the listener position furthest from the array. As expected localization accuracy generally decreased in the more reflective room but, in general, good localization was achieved. However, Marentakis points out that signals with strong transient characteristics, such as the enveloped white noise signals used in this test, are localized independent of the room reverberation [Hartmann, 1983].

6.7.3 WFS Distance Effects

WFS differs from other spatialization techniques as it can theoretically reproduce the wavefront curvature associated with a virtual source in front of, or behind the loudspeaker array. As WFS recreates the sound field over an extended area, the amplitude and location of a virtual source should therefore change in a realistic fashion with listener movement. It has been suggested that this idea of motion parallax, where the perspective changes naturally as the listener moves, can provide an indication of the distance of a WFS virtual source.

Boone assessed the intelligibility of speech by comparing noise and speech signals played back using a single loudspeaker, and two WFS virtual sources which differed only in terms of their distance [Boone *et al*, 2003]. The results showed an improvement in intelligibility when the speech and noise source were separated using WFS which would seem to indicate that there is at least some perceptual difference between two WFS virtual sources differing only in distance. However, Wittek points out that this perceptual difference could occur due to differences in the way a single loudspeaker and an array of loudspeakers interact with the acoustic of the reproduction room [Wittek, 2003].

Nogues conducted an experiment which provided better evidence that the movement of the listener through a WFS sound field does indeed provide some indication of distance [Noguès *et al*, 2003]. They asked the subjects to control the distance of a WFS virtual source so that the source distance matched the distance of two other simultaneously reproduced virtual reference sources (see Figure. 6.24). The subjects were asked to move around in the listening area throughout each test. In the first experiment, the subjects could only manipulate the WFS source distance while other parameters such as the direct to reverberant energy ratio and signal level were kept constant. The results showed that the subjects were indeed able to position the middle guitar in between the other two solely using the perspective cue of the WFS sound field. This result suggests that the perception of distance due to the virtual source position in a WFS system can be perceived independently of the subjective distance impression.

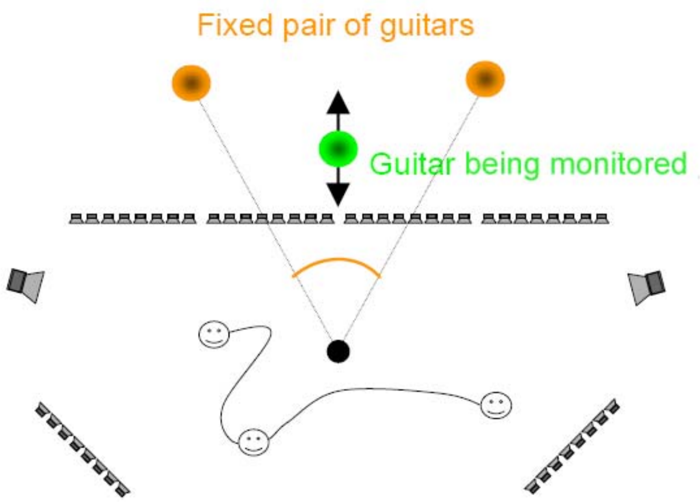

Fig. 6.24 Technical setup for listening tests by Nogues

In a second experiment, other distance cues such as the direct to reverberant ratio were included. A number of sources were synthesized at different distances and the subjects were asked to adjust the direct to reverberant ratio of each source so that the perceived distance matched that of a pair of reference sources. The results showed that the perception of distance was primarily influenced by the direct to reverberant ratio, and that the wavefront curvature is a weak localization cue which is easily overridden by other cues.

Kerber implemented listening tests to compare the perceived distance of real and virtual focussed dry sources. The results for a real source shown in Figure 6.25 support the results of other tests in that only small distances (< 1m) are perceived accurately and that large distances are consistently underestimated. Focussed WFS sources do not seem to be able to produce the same distance perception as real sources. Wittek suggests that these results indicate that at a fixed listening position, the curvature of the wavefront of a dry WFS virtual source does not support distance perception. However, in spite of not being a crucial cue, a correct wavefront curvature (and thus a consistency between curvature and actual distance) may support the perception of distance, particularly if the listener can move [Wittek, 2003]

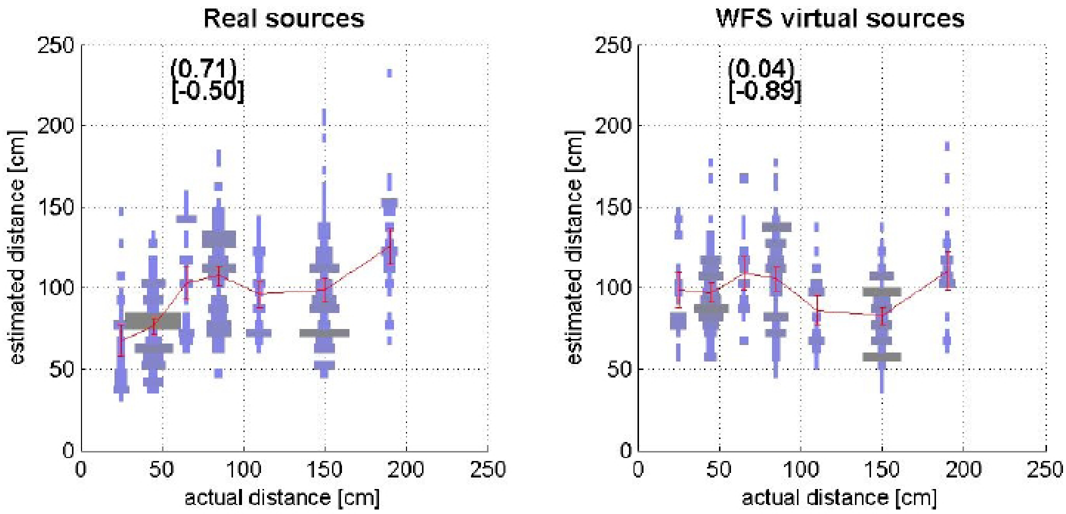

Fig. 6.25 Distance of real (left) and virtual (right) sources reported by Kerber

Usher similarly found that in the absence of any indirect sound, when a source is positioned beyond a certain distance using a WFS system, the curvature of the wavefront does not seem to be used to determine the distance of the virtual source, but rather the timbre of the perceived source dominates [Usher *et al*, 2004]. It would seem, therefore, that in order to accurately produce WFS virtual sources at different distances, some form of artificial reverberation is needed to provide additional distance cues. However, Wittek points out that disturbing reflections caused by the WFS array itself may in fact also hinder the perception of the distance of virtual sources in front of or behind the array. It is not the case that a dry WFS virtual source will automatically produce a natural reflection pattern in the reproduction room [Wittek, 2003] and this is illustrated in Figure 6.26 which shows a WFS system with a virtual source (blue dot) positioned in front of the array. The correct reflections that would arise if a real source was at this position are indicated by the green dots, while the actual reflections that arise are shown as orange dots. Clearly both the timing and direction of the actual reflections do not correspond to the desired source position, but rather to the distance of the loudspeaker array itself. The virtual source distance will therefore most likely be perceived to be the distance of the loudspeaker array rather than the specified source distance. This again indicates the significant influence of the reproduction room acoustic on the performance of WFS systems. Various listening room compensation schemes have been proposed which can actively cancel early reflections in the horizontal plane [Corteel *et al*, 2003; Spors *et al*, 2003]. The results of simulations suggest that these techniques could help to reduce the detrimental effect of early reflections in the listening room over a large area.

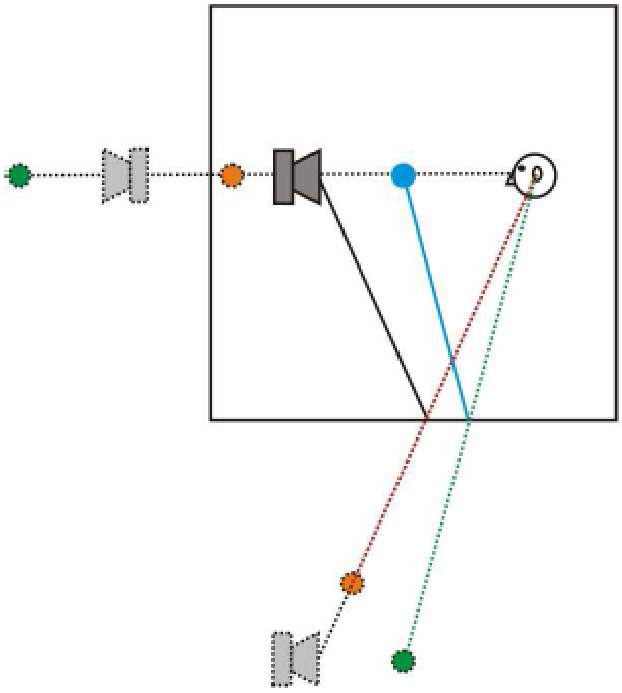

Fig. 6.26 WFS reproduction room reflections

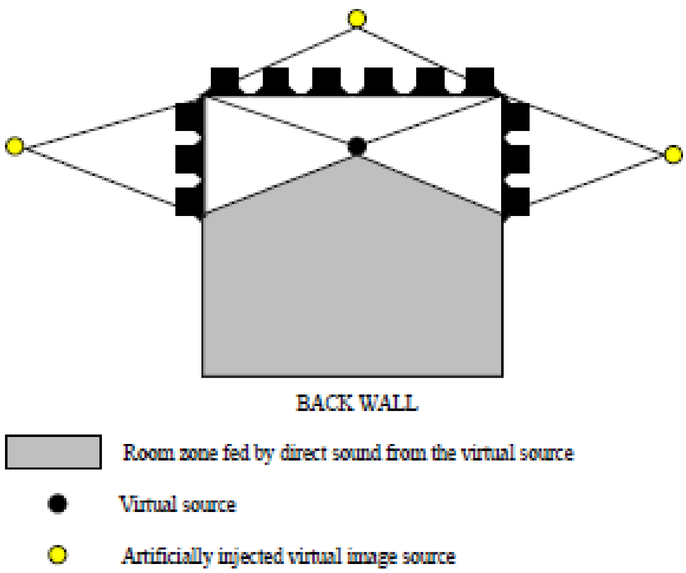

Fig. 6.27 WFS virtual source and simulated room acoustic

Or course, WFS can also simulate different source distances through the addition of early reflections and reverberation. Figure 6.27 illustrates a scheme proposed by Caulkins to artificially inject early reflections which are otherwise absent in the WFS reproduction of focussed sources [Caulkins *et al*, 2003]. As with other spatialization techniques, more listening tests are required to fully determine the perceptual effect of combined and potentially conflicting virtual and real acoustic reflections.

6.7.4 WFS Spectral Colouration

Perceptible changes in the timbre and spectral content of the source can occur in WFS systems because of distortions in the high frequency content caused by spatial aliasing and diffraction. Start found that the colouration due to spatial aliasing changes rapidly with listener movement, and is very distinct for broadband signals [Start, 1997]. In addition, the dynamic movement of a virtual source also introduced spectral changes, but this effect was much less noticeable with speech signals. Interestingly the tests found that source colouration was significantly reduced in the reverberant reproduction room compared to the anechoic room. Start recommended a high-frequency optimization scheme to reduce these colouration artefacts. However, this approach does have a negative influence on localization accuracy.

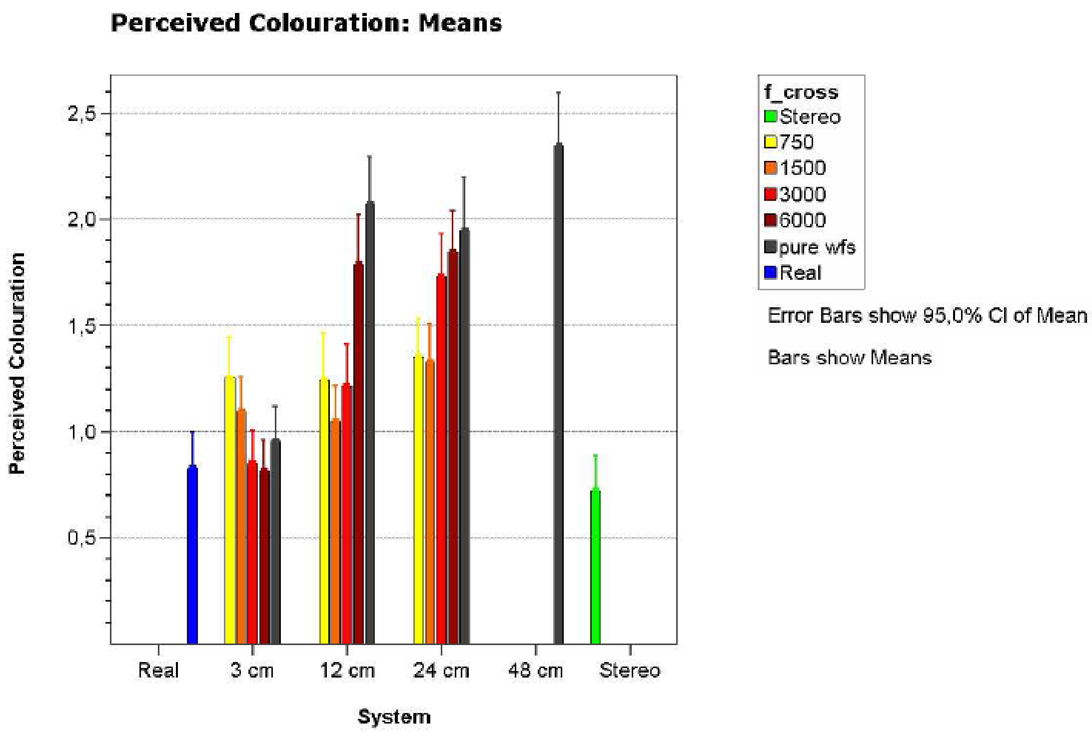

Fig. 6.28 Perceived colouration for various WFS and OPSI systems

Wittek carried out a series of listening tests to determine the relative level of colouration of real, WFS, WFS+OPSI and stereophonic sources in an acoustically treated listening room [Wittek, 2007]. The results, shown in Figure 6.28, indicate that the amount of signal colouration increases with spatial aliasing, and that the OPSI method can significantly reduce the perceived colouration in comparison to the WFS system. The author suggests that the non-zero result for the real reference source was due to the non-individualised HRTF used in the experiment. Interestingly, the lowest level of colouration was reported with the standard stereo system.

6.7.5 WFS – Conclusions and Discussion

The results of the various tests discussed in the previous chapters illustrate the demanding nature of the WFS system. Clearly, the spatial aliasing frequency is a crucial parameter, and the results of Start, Huber, Wittek illustrate the degradation in localization accuracy and timbre that occurs when this factor is too low. Consequently, a very large number of loudspeakers are required if a WFS system is to be successfully implemented for a large group of listeners. Start presented a WFS system with an inter-loudspeaker spacing of 0.25m and a spatial aliasing frequency of 1360Hz, which, for now will be taken as a minimum system specification. A circle with a 2m radius has a circumference of 12.56m which, based on our chosen system, equates to approximately 50 loudspeakers. Assuming a single seated listener occupies an area of approximately $0.5m^2$ (1m x 0.5m) and that all listeners must be seated at least 0.75m from the array, it can be shown that approximately ten listeners will be accommodated by this system. Clearly, implementing a full 360^0 WFS system for a larger audience will be a significant logistical challenge. Techniques to increase the perceived spatial aliasing frequency without increasing the number and resolution of the loudspeaker array are currently being investigated [Corteel *et al*, 2008], although it appears this can only be achieved over a limited listening area.

Much work remains to be done in terms of the evaluation of WFS, as it appears that the perceptual effects suggested by simulations may not be as significant in practice. In terms of localization accuracy it seems that, providing the spatial aliasing frequency is high enough, WFS can indeed create well-localized virtual sources, at least under anechoic conditions. However, it is still unclear whether the precedence effect is also contributing to localization, particularly in reverberant conditions. This view seems to be supported by the results of tests which found that localization accuracy is decreased for focussed sources in front of the array [Verheijen, 1998] or for non-central listener positions [Marentakis *et al*, 2008]. The quality of localization or locatedness is another important factor, as measures such as the standard deviation in source angle may suggest good localization was achieved when in fact significant source broadening occurred. Further subjective tests are required to fully determine the localization accuracy of WFS systems for focussed sources in front of the array and for non-central listener positions.

It has been suggested that WFS can be used to enlarge the effective listening area of other stereophonic techniques, and this appears to be true for certain applications. Various tests [Corteel *et al*, 2004] have found that WFS does significantly increase the listening area for two-channel stereo as sources are less likely to collapse to the nearest loudspeaker when listeners are displaced laterally from the sweet spot. WFS has similarly been successfully used in a domestic situation as a flexible and robust method for 5.1 reproduction as it is less sensitive to listener position and non-standard loudspeaker positions [Corteel *et al*, 2004]. Other studies have found that plane wave reproduction with a WFS system can be used to diffuse the rear channels in a 5.1 reproduction set-up, again increasing the effective listening area [Boone *et al*, 1999]. The use of WFS as a flexible reproduction format for domestic cinema and audio applications would therefore seem to be one of the most promising applications of this technique. The CARROUSO Project (Creating, assessing and rendering in real-time of high quality audio-visual environments in MPEG-4 context'') has attempted to merge WFS with the flexible MPEG-4 standard for the transfer of both recorded and synthesized sound fields between different reproduction systems. Much of this work has concentrated on developing practical implementations of WFS suitable for the domestic market, such as flat-panel, distributed mode loudspeakers [Farina *et al*, 2000].

The reproduction of virtual sources at different distances is one of the most widely lauded features of the WFS method. However, it appears that when the listener position is fixed, the correct wavefront curvature produced by a WFS virtual source does not provide any perceptible distance information [Kerber *et al*, 2004]. In addition, reflections from the array and reproduction room walls will tend to pull the perceived source distance to the distance of the loudspeakers [Wittek, 2003]. Although it has been suggested that this weak cue might help to support other more dominant distance cues [Nogués *et al*, 2003], there is little evidence to support this claim. The one significant exception is when the listener can move, and in this instance the ability of WFS to reproduce a sense of changing perspective or motion parallax has been shown to support the estimation of distance.

6.8 Analysis of Results and Recommendations

The experiments discussed in the preceding sections illustrate the wide range of interrelated factors which influence the perceived performance of spatialization techniques such as stereophony, Ambisonics and WFS. Virtual sources positioned using multichannel stereophonic techniques have been found to be unstable if the source is not positioned at a loudspeaker. This is true at all listener positions for lateral and rear sources but also for front sources if the listener is displaced laterally closer to one of the loudspeakers. However, although source localization often collapses to the nearest loudspeaker, this will always be to the loudspeaker pair about the desired source position. Therefore, increasing the number of loudspeakers will also increase the overall localization accuracy and six loudspeakers seems to be the minimum number of channels required for reasonable accuracy in all directions, at least for a single listener. The differences between virtual sources positioned at, or between loudspeakers also has an effect on dynamically moving sources and a number of studies have found that trajectories created with stereophony tend to therefore highlight the positions of the loudspeakers.

It is clear that for Ambisonics, the number and arrangement of loudspeakers in the array is particularly important. A regular hexagonal array was found to produce much better results than either square or rectangular layouts, particularly for lateral sources, and so can be taken as a minimum specification, as with stereophony. Likewise, the precedence effect influences localization accuracy in much the same way as with stereophony. The results of a number of tests clearly indicate that different decoder designs are optimal depending on whether playback is for a single listener or a group of listeners. A full-band max r_E (or in-phase, if the audience is very near the array) has been shown to be the optimal solution for larger listening areas as these decoders reduce (or eliminate, in the case of in-phase) the anti-phase components which aid localization at the centre point but significantly distort it at other listener positions. The results of a number of studies also seem to confirm that localization accuracy is improved when the order of the Ambisonics system is increased [Pullki *et al*, 2005; Jot *et al*, 1999; Daniel, 2000]. This result was largely expected as an increase in order represents an increase in the spatial resolution of the spherical harmonic representation of the reconstructed sound field. The directional information represented by the spherical harmonics is therefore more accurately represented and, consequently localization accuracy improves. A number of studies have also clearly demonstrated that the size of the effective listening area also increases with the order of the system [Bates *et al*, 2007b; Frank *et al*, 2008]. A number of theoretical studies indicate that Ambisonics can only perfectly reconstruct a sound field in a very small area at the centre of the array, and this is only possible up to a certain frequency [Poletti, 1996; Daniel *et al*, 1998; Bamford, 1995]. Daniel suggests that the reconstructed sound field becomes increasingly and linearly distorted away from the centre point. Consequently, if the system order (and hence the reconstruction frequency limit at the centre point) is increased, this will also increase the accuracy of the reconstructed sound field at other, off-centre positions [Daniel *et al*, 1998].

Ambisonics is generally preferred to stereophony for dynamically moving sources as it produces smooth trajectories that do not highlight the positions of the loudspeakers. Consequently, Ambisonics will never produce a source using a single loudspeaker and so unlike stereophony, cannot produce the most tightly focussed virtual image possible. Increasing the order the system reduces this effect as the increase in directivity reduces the number of loudspeakers which are active at any one time. There is some evidence that a similar trade-off is apparent with recorded sounds as in one test, Soundfield microphone recordings were found to be more spacious and enveloping than stereophonic recordings, but also less accurate in terms of directional localization.

The simulation of distance appears to be largely dependent on the addition of artificial reverberation. Other processes such as the Doppler effect, air absorption, or wavefront curvature (as in WFS) do not seem to be able produce a reliable perception of distance on their own, but are important as secondary distance cues. While a straightforward ratio between the direct and diffuse reverberant signals can produce a relative sense of distance, the results of a number of tests indicate that early reflections are more important than the diffuse reverberant signal in this regard. Some tests have found that a mixture of directionally accurate specular and diffuse early reflections is preferable to purely specular reflections [Martin *et al*, 2001]. The addition of artificial reverberation clearly supports distance modelling but its effect on directional localization is unclear. Some have suggested that these additional indirect signals reduce localization accuracy especially at off-centre listening positions [Begault, 1992], while others have suggested the exact opposite and argue that the increased realism of such a sound scene benefits localization [Lund, 2000].

WFS is in many respects very different from stereophony or Ambisonics and is certainly much more demanding in technical terms. While the results of various listening tests seem to suggest that well-localized virtual sources can be created with WFS this is very much dependent on the spatial aliasing frequency, and hence on the size of the array. In addition, questions remain as to how focussed these virtual sources are, and the contribution of the precedence effect in localization with WFS. While it appears that WFS can be used to increase the effective listening area of other spatialization techniques such as two-channel stereo or 5.1 surround sound, extending this to a full, large scale system is difficult, if for no other reason than the many, many loudspeakers which would be required to surround a large audience. In addition, the results of listening tests do not seem to support the claim that WFS can position sources behind or in front of the array through the reproduction of the correct wavefront curvature. The one notable exception to this is when the listener can move through the listening area, otherwise WFS systems must use artificial reflections and reverberation to simulate different source distances, in much the same way as other spatialization techniques. These results indicate that WFS is perhaps not, at the moment at least, the most suitable system for the presentation of spatial music as the perceptible benefits do not seem to justify the vastly increased technical requirements.

6.8.1 Discussion

The preceding discussion illustrates the difficulties in the presentation of spatial audio to multiple listeners. The influence of the precedence effect is particularly noticeable for off-centre listeners and it appears that a high degree of directional localization accuracy can only really be achieved for every listener if a single loudspeaker is used. Spatialization techniques such as pair-wise amplitude panning, and to a lesser extent, higher order Ambisonics, produce the next best results, as the number of contributing loudspeakers is restricted and are situated in the same approximate direction as the source. As lower order Ambisonics systems generally utilize every loudspeaker to produce the virtual image, these are particularly susceptible to localization distortion due to the precedence effect. WFS, while appropriate for certain applications, does not seem to be viable as of yet for presentations of spatial music due to the technical and logistical restraints and the limited benefits.

It is also clear that Ambisonics is consistently preferred to stereophony for moving sources as it disguises the positions of the loudspeakers which results in a smoother trajectory. A number of non-standard amplitude panning techniques have been developed which attempt to overcome this problem through increasing the number of loudspeakers which are used at any one time. While these techniques certainly appear to improve matters for dynamically moving sources, it is not clear if they provide any advantage over a high order Ambisonics system other than the fact that they can readily collapse the virtual image to a single loudspeaker. Max Re and in-phase decoding schemes, without shelf filtering, appear to be the optimal decoding schemes for larger listening areas.

The results presented in this Chapter suggest that a minimum of six loudspeakers is required to produce optimal results for a single listener with either stereophony or Ambisonics. An eight channel system would therefore seem to represent an acceptable minimum layout for larger number of listeners as it contains eight discrete spatial locations to which sources will be localized with a good degree of accuracy, it is sufficient for third order Ambisonics, and is reasonably achievable in terms of hardware. It has been the experience of the author that quadraphonic systems produce extremely poorly localized lateral virtual sources when extended for larger numbers of listeners and movements from front to back instead abruptly switch between each position (see Section 10.2). While an additional pair of lateral loudspeakers alleviates this issue somewhat, the wide angle between lateral loudspeaker pairs is still problematic. An eight channel system contains a pair of lateral loudspeakers and this provides a more useful degree of discrimination in lateral positions and movements, while still being reasonably efficient and economical. For these reasons, the author has adopted an eight channel loudspeaker array as a standard system for performances of spatial music.

6.8.2 Implications

The results presented in the preceding section suggest that it is very difficult to produce spatial locations and trajectories which are unambiguously perceived by every listener, in the same way. Even in the case of point sources which are clearly localized, each listener will be orientated differently with regards to the loudspeaker array, and so will have a different perspective on the spatial layout. As noted earlier, directional localization accuracy is the main topic under investigation in many of these tests, but this is not necessarily the only way in which space can be utilized in a musical composition. The results presented earlier suggest that this may in fact be a necessity. However, it is just as important to know if these other uses of space are clearly perceptible to an audience, and if so, which spatialization technique can achieve this most effectively, if at all? Clearly, Ambisonics is the preferred spatialization technique for dynamically moving sources. However, it is also clear that the precise trajectory perceived by each listener will be strongly influenced by their position within the array.

If a recorded sound is to be used in spatial music composition, the ambisonic Soundfield microphone represents the most flexible recording option if an enveloping sound field is required. However, if a more directional diffusion is required, then monophonic or stereophonic microphone techniques are perhaps more applicable as although multi-channel microphone techniques can be very effective, they are tied to a specific reproduction layout.

While many composers continue to utilize various multi-channel techniques, others have adopted an entirely different approach based upon a single two-channel stereo source and a large, disparate collection of spatially distributed pairs of loudspeakers, i.e. a loudspeaker orchestra. This aesthetic represents a very different approach to the multi-channel techniques discussed in the preceding chapters. However, the art of diffusion is admirably focussed on the perception of the audience and the real technical problems which arise in these kinds of performances, something which is often lacking in multi-channel tape compositions.

The second half of this thesis will focus on spatial music composition via the analysis of a number of different composers and aesthetics, and some original compositions by the author. Different approaches to the use of space as a musical parameter will be assessed in terms of the technical and perceptual research presented in the preceding chapters. Inevitably, greater emphasis will be placed on music from the twentieth century as many significant aspects of spatial music are dependent on technical developments from this era, however, spatial music is not solely a twentieth century phenomenon. The spatial distribution of performers has been used for centuries in European religious choral music, and this antiphonal style is itself derived from the even more ancient call-and-response form. The next chapter in this thesis will examine this early form of spatial music and investigate the development of acoustic spatial music in the first half of the twentieth century, prior to the development of recording and amplification technology and electronic spatialization techniques.

7 Acoustic Spatial Music

Spatial music is often closely associated with technological developments in the twentieth century, yet the use of space as a musical parameter is much older. Call-and-response patterns can be found throughout history in many different cultures and musical traditions. In this dialogue form, musical material is divided between two groups, which will necessarily be situated at two different spatial locations. Call-and-response patterns therefore represent the most basic form of spatial music and they are a fundamental aspect of the earliest formalized system of spatial music, antiphonal choral music.

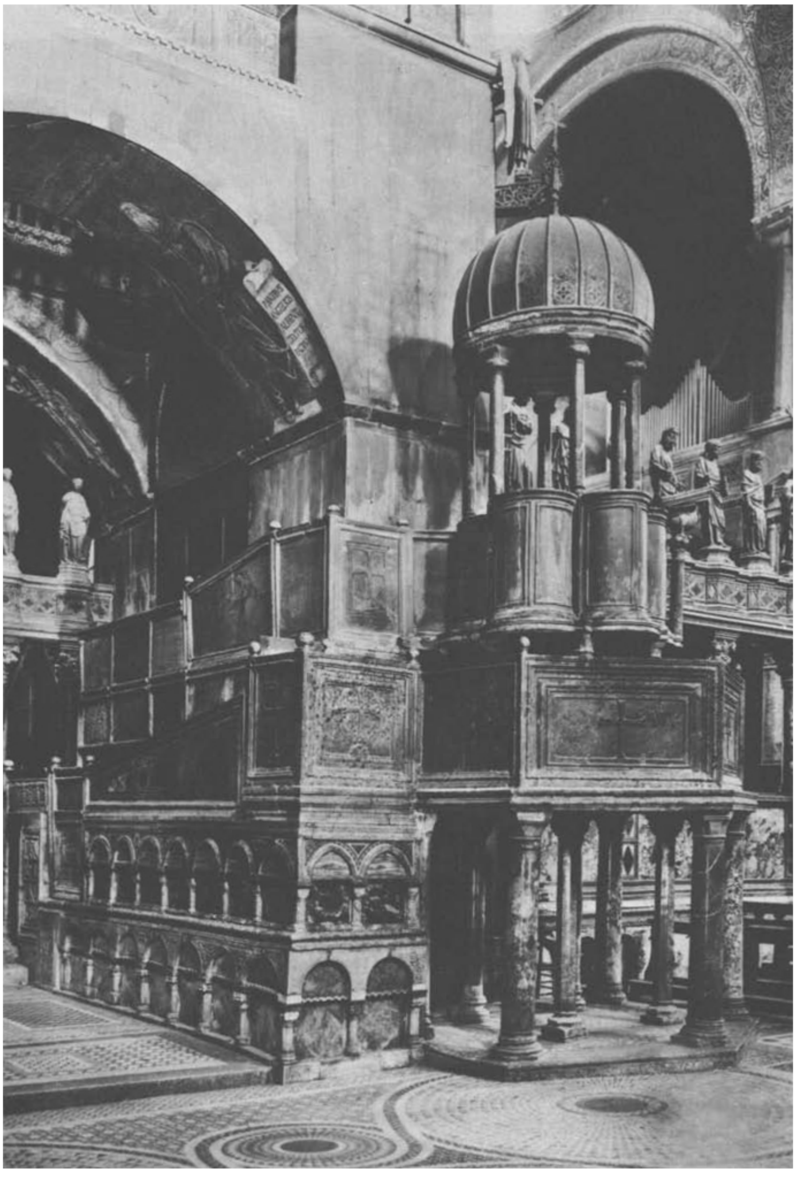

Fig. 7.1 Interior of St Marks Cathedral in Venice

7.1 Early Spatial Music

In the early Christian tradition, the term antiphon was used to denote a sung response to a psalm during a religious service, and this gave rise to the antiphonal style of singing which was widely used in medieval church music. The polyphonic choral music of the sixteenth century and the Renaissance retained this older technique, often using a smaller *capella* choir in conjunction with the main *ripieno* choir. The popularization of this technique is generally attributed to the Venetian School founded by the Flemish composer Adrian Willaert, who became *maestro di cappella* of St. Mark's Cathedral in Venice in 1527 (see Figure 7.1). The existing use of alternating choirs was facilitated here by the unique interior of the Cathedral of St. Marks which contains two spatially separated organs and choir lofts. Composers began to take advantage of the increased distance between the groups and use the spatial separation as a special effect. Bryant points out however that it cannot be assumed that the spatial separation of the choirs, or *cori spezzati*, was an integral part of early music of the Venetian School. The historical records suggest that this was instead an alternative arrangement which was occasionally used as a special effect based on the preference of a performance's music director or special guest [Bryant, 1981]. Indeed, as antiphonal music is largely based on successive, often contrasting phrases which alternate between the two choirs, a greater spatial separation is not, strictly speaking, necessary, as the separation is inherent to the musical material. Willaert's eight-part *Vespers,* composed in 1550 is one of the earliest examples of this music and features the echo and dialogue effects between the two spatially separated groups which are typical of this aesthetic.

By the early seventeenth century the music of the Venetian school of composers had developed beyond this strict ritual alternation between two choirs to a rapid, more sophisticated dialogue between multiple choirs and instrumental groups. Students of Willaert's such as Andrea and Giovanni Gabrieli composed large scale works for multiple groups and choirs and were also the earliest composers to include dynamic markings and specific instrumentation in the score. Bryant suggests that the larger choir would have remained at floor level as they had other ceremonial duties during the mass [Bryant, 1981]. The instrumental groups, who had no active role in the ceremony, would have been placed in an inconspicuous position, most likely in the two elevated organ lofts, while the smaller *cappella* choir would most likely have been situated at some other point away from the main choir. This theory is supported by historical accounts from the time, one of which details the method of synchronizing the various spatially distributed groups. It appears that two additional conductors were employed to relay the beat indicated by the principal conductor, situated at floor level with the main choir, to the musicians in each organ loft. Giovanni Gabrielli's famous work *In Ecclesiis* is an excellent example of this form of polychoral music, using four separate groups of instrumental and singing performers accompanied by organ and basso continuo to create a spatial dialogue and echo effects (see score extract in Figure 7.2).

Fig. 7.2 Echo effects in Giovanni Gabrielli's *In Ecclesiis*

The Venetian school was highly influential across Europe, helped in part no doubt by the invention of the printing press a century before. The English composer Thomas Tallis composed *Spem in Alium* in 1573 for forty separate vocal parts arranged in eight choirs while Orazio Benevoli’s *Festal Mass* was written for fifty-three parts, two organs and basso continuo. Over the next four hundred years, the use of antiphony became rarer with some notable exceptions such as the antiphonal choral effects of J. S. Bach’s *St. Matthew Passion* (1729), and the motivic interplay of spatially separated groups of Mozart’s *Serenade in D for four Orchestras* (1777). In the Romantic era, composers occasionally placed groups of musicians away from the main orchestra for dramatic effect. One example is Berlioz’s *Requiem* (1837), which at its premiere included four brass ensembles positioned at the four cardinal points, along with a massive orchestra of singers, woodwinds, horns, strings, and percussion. Berlioz was aware in advance that this work would be premiered in *Les Invalides*, the gigantic domed cathedral of the military hospital in Paris, and he exploited the characteristics of this space in this new commission. In the famous *Tuba Mirum* section, the invocation of God's fury with the damned is invoked through the consecutive entrance of the four brass ensembles which gradually builds to a dramatic climax of massed timpani and voices (see Figure 7.3). [1]

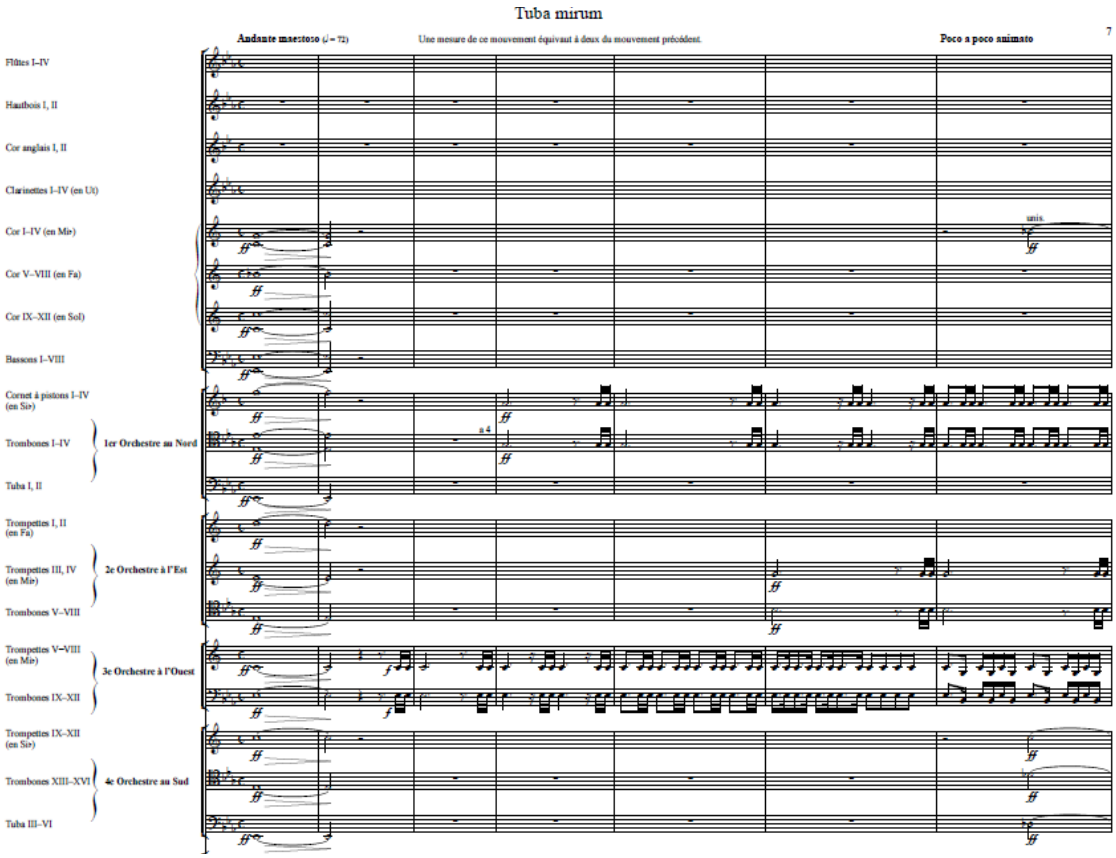

Fig. 7.3 Tuba Mirum section in Berlioz's *Requiem*

Although Berlioz was clearly thinking about the use of space in music (he referred to it as "architectural music"), the spatial distribution of the performers in this case is largely for dramatic effect and is not a critical feature of the work, and this is true of most of the historical examples discussed in this Section. In the early part of the twentieth century, space was sometimes used to create a sense of perspective by contrasting the orchestra on stage with more instruments placed at a distance off-

[1] Interestingly, during the first performance, the conductor Habeneck (a rival of Berlioz) is alleged to have broken for a pinch of snuff during the critical entrance of the brass ensembles, requiring Berlioz himself to leap in and take over the conducting duties [Bromberger, 2007].

stage. The Austrian composer Gustav Mahler often used off-stage musicians in addition to the main orchestra, such as, for example, the brass and percussion in the fifth movement of *Symphony No. 2* (1894) or the off-stage snare drum in *Symphony No. 3* (1896). Other significant composers of the era also used similar effects, although not as frequently as Mahler. Igor Stravinsky made use of tubas *dans le couline* (in the corridor – i.e. in the wings) in the ballet score of *Firebird* (1910) and Strauss featured six trombones *auf dem Theater* in *Die Frau ohne Schatten* (1919). This period was one of great upheaval in Western Art music as many composers begin to move away from the strict functional tonality and defined meters of previous eras. While composers like Stravinsky and later Schoenberg experimented with the basic foundations of musical structure, others retained aspects of traditional music practice but utilized them in a very different way. The American Charles Ives is one such composer who regularly combined traditional tonality with the then new ideas of musical quotation, polyrhythms and meters, and spatial effects. Ives was a contemporary of Mahler (both produced most of their music within the same thirty year period from 1888 to 1918) and although their music is quiet disparate, and derived from very different musical traditions, both composers do exhibit certain similarities [Morgan, 1978]. For example, both composers were interested in the quotation of other musical material within their own compositions and both regularly combined and juxtaposed layers of different and contrasting material. Both composers also retained aspects of functional tonality in their work and made extensive use of overlapping yet unrelated tempi. While both composers used the spatial distribution of performers in their work, it would be Ives who developed this practice further and have the most lasting effect on the development of spatial music, particularly in America.

7.2 The Co-Existence of Dissimilars

Charles Ives was born in Connecticut in 1874 and although he received little acclaim in his lifetime, he is now regarded by many as one of the most innovative composers of his generation. Throughout his life, Ives talked about the influence of his father, a musician and band-leader, on the development of his unorthodox approach to music. He stated that his father had once directed his two sons to sing a tune in one key while he accompanied them on piano in another, while numerous accounts exist of his experiments with the spatial effect of simultaneous marching bands [Ross, 2009; Mortenson, 1987]. This influence can clearly be seen in the music of Charles Ives, which often makes use of overlapping keys and meters and the combination of European Art music with American popular and church music. Over the course of his life, Ives would go on to explore many of the musical innovations which would become associated with modern contemporary music such as polytonality and polyrhythm, tone clusters and microtonality, musical quotation and collage, and also spatial music. Much of Ives' music involves the juxtaposition of various disparate elements and his compositional use of space generally reflected this. *The Unanswered Question* (1908), one of his most famous compositions, uses the spatial distribution of musicians to highlight the three distinct layers of strings, woodwinds and brass. The three layers operate independently at their own tempo and key and this musical separation is further accentuated through the placement of the string orchestra off-stage, the woodwind ensemble on stage, and the solo trumpet positioned at some other distant position, such as a balcony. Throughout the piece the string orchestra performs slow, sustained tonal triads which are punctuated by short trumpet phrases (the question), which are in turn answered in an increasingly incoherent fashion by the woodwinds. The symbolism of this work is beautifully supported by the spatial distribution and layering of the musical material. The slow-moving tonal strings represent the natural world which surrounds the audience and remains in constant yet slow motion, undisturbed by the question and answer dialogue of the trumpet and woodwinds. The trumpet sounds out the question with a clear, atonal phrase which is then answered by the woodwind ensemble. In contrast to the clear question of the solo trumpet, the woodwinds respond with multiple, overlapping and seemingly unrelated phrases and so, no answer is found to the eternal question of the title.

In much of Ives' music, space is used to clarify and define the various overlapping yet independent musical layers, and other composers at the time were also beginning to experiment with the layering of disparate musical material. Italian futurists like Luigi Russolo were exploring collage form and noise while Darius Milhaud's 1918 ballet *L'Homme et son desir* used multiple, spatially distributed ensembles playing independently of each other, sometimes in different metres [Zvonar, 2004]. However, Ives' fourth and last symphony (1910-1916) is certainly one of the most ambitious works that made use of this technique. This work, for a gigantic orchestra with additional off-stage ensembles, was not performed in full until nearly a decade after Ives' death. The second movement juxtaposes so much different thematic material that a second conductor is generally required while the final movement contains a contrasting dialogue between discordant and tonal material. In the conductor's note Ives wrote [Johnson, 2002];

> *"As the distant hills, in a landscape, row upon row, grow gradually into the horizon, so there may be something corresponding to this in the presentation of music. Music seems too often all foreground even if played by a master of dynamics…*
> *It is difficult to reproduce the sounds and feeling that distance gives to sound wholly by reducing or increasing the number of instruments or by varying their intensities. A brass band playing pianissimo across the street is a different sounding thing than the same band playing the same piece forte, a block or so away. Experiments, even on a limited scale, as when a conductor separates a chorus from the orchestra or places a choir off the stage or in a remote part of the hall, seem to indicate that there are possibilities in this matter that may benefit the presentation of music, not only from the standpoint of clarifying the harmonic, rhythmic, thematic material, etc., but of bringing the inner content to a deeper realization.* "

Clearly Ives used the spatial separation of performers to create a sense of distance and perspective, in much the same way as European composers such as Mahler. However, Ives also used this spatial distribution to clarify different layers of independent and potentially dissonant musical material and to facilitate the performance of overlapping yet unrelated musical layers, often at different tempi or metres. While the spatial separation of musical material at different tempi obviously has practical benefits for its performers, the above quote also indicates that Ives had intuitively realised that this spatial separation also benefited the listener. It is unknown whether this insight was informed by the recent development of Gestalt psychology in Germany, or derived from Ives' own experience. However, it has since been shown by Bregman and others that our ability to segregate an audio scene into multiple streams strongly influences our perception of musical parameters such as melody and rhythm [Bregman, 1990]. Bregman's work on Auditory Scene Analysis (see Chapter Two, Section 2.3) emphasized the importance of spatial cues in the segregation of audio streams and he suggested that the spatial separation of a multiplicity of sounds prevents the auditory system from computing dissonances between them. Other studies have also found that a listener's ability to detect and understand the content of multiple signals is improved if the signals are spatially separated signals [Shinn-Cunningham, 2003; Best, 2004] and this would also appear to support Ives' use of space to "clarify the harmonic, rhythmic and thematic material". Although largely ignored for much of his career, Charles Ives would eventually be recognized as a highly creative and innovative composer, and his experiments with spatial music would be an important influence on a number of American composers.

The composer Henry Brant (1913-2008) is one such composer who was influenced by Ives. Over the course of his career, Brant wrote seventy-six works of spatial music (along with fifty-seven non-spatial works [Harley, 1997]) and has become one of the most famous composers of orchestral spatial music. Brant's use of space was clearly influenced by Ives, as illustrated by this extremely admiring 1954 description by Brant of Ives' *The Unanswered Question*.

> *"This unique, unprecedented little work, written in 1908, presents, with extraordinary economy and concentration, the entire twentieth-century spatial spectrum in music, and offers guidelines for solving all the practical problems involved. The spatial-contrapuntal-polytemporal principles so brilliantly exemplified in this piece are the basis for the more complicated spatial superimpositions present in all my own recent large-scale works* [Brant, 1967]*"*.

Brant composed and wrote extensively on the use of space in orchestral music and his first spatial work, *Antiphony I (1953)* contains many of the core ideas which the composer would continue to use throughout his career. In this piece, the orchestra is divided into five groups which are placed at different parts of the auditorium and perform material in contrasting tempi, meter and harmonies. Although the entrance of each group is cued, they then proceed independently at their own speed. The composer states that "a purposeful lack of relationship between the intervals, phrasing, note-values, tone-quality and sonorities of the various lines will necessarily produce a complex result as soon as the lines are combined [Harley, 1997]". Brant presented his ideas about the compositional use of space in a brief article written in 1967 [Brant, 1967] which is summarized as follows;

- The perception of different layers of musical material can be enhanced by the spatial separation of the performers.
- Exact rhythmic coordination is difficult to achieve when musicians are spatially separated by large distances.
- Spatial separation is equivalent to pitch separation but allows for greater complexity as material in the same harmonic range which would merge if produced from the same location, can be separated into distinct musical lines if the sources are separated in space.

- Each situation and listening position is different and there is no single optimum listening position.

Brant is clearly influenced by Ives in his approach, which is based on completely contrasting material, spatially separated and without any rhythmic synchronization between the groups. The use of strings as a continuous static layer in contrast with other musical lines at different spatial locations represents another similarity between this work and *The Unanswered Question*. However, unlike Ives, Brant specified the precise location of each ensemble, something he would continue to do throughout his career. This allowed Brant to explore other spatial effects as illustrated by the spatial distribution used for *Millenium II* (1954) which is shown in Figure 7.4[Harley, 1997]. The piece uses a spatial assembly of ten trombones and ten trumpets, positioned along the side walls of the hall, as well as on-stage percussion and brass, and a single voice, preferably positioned to the rear and above (in a balcony, for example). At the beginning of the piece, the trumpets and trombones enter one-by-one in numerical order, each playing in a different melody in a different key. Brant describes the effect as follows;

> *"there is a compelling impression of the hall tangibly filling up with sound, principally along the walls, but also with some feeling of the centre being progressively saturated, especially as the accumulation proceeds towards its maximum point. The impression of the sound travelling gradually down the first wall is very strong; this impression of moving direction becomes less well defined as the further entrances and accumulations occur* [Brant, 1967]*".*

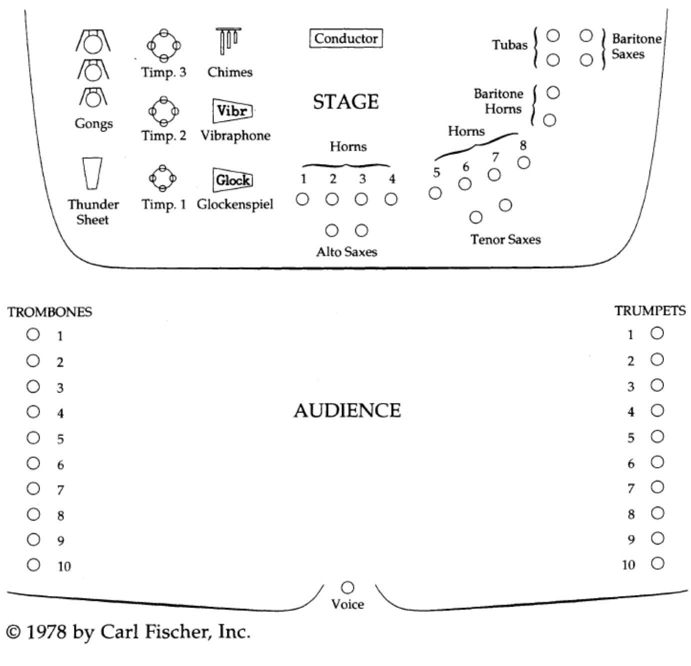

Fig. 7.4 Spatial location of musicians in Brant's *Millennium II* (1954)

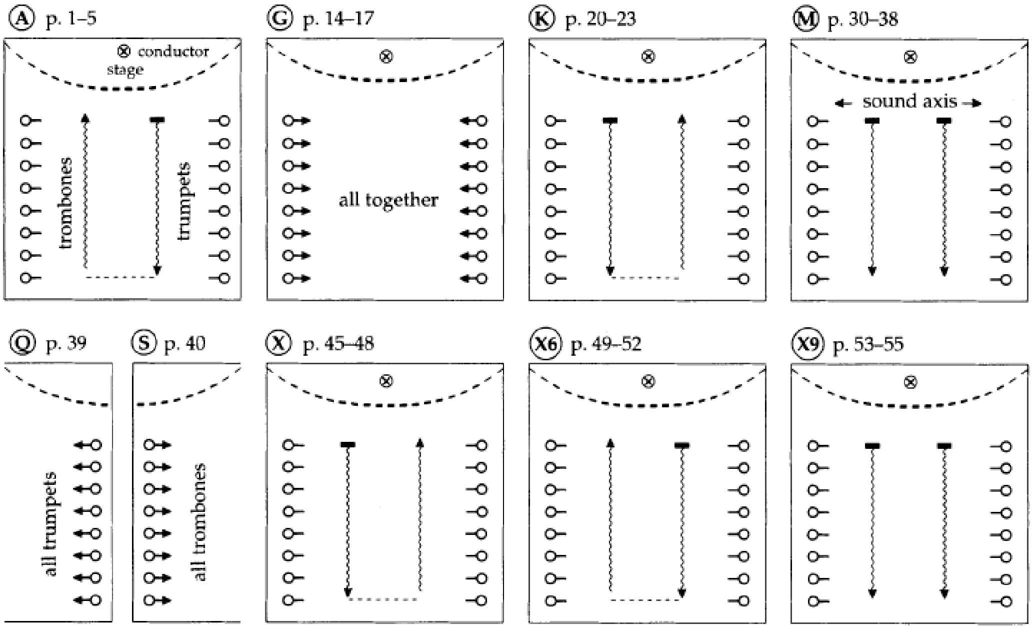

Fig. 7.5 Spatial patterns in Brant's *Millennium II* (1954)

Fig. 7.6 Sound axes in Brant's *Millennium II* (1954)

This piece also includes other spatial effects such as the stepwise introduction of trombone/trumpet pairs, which are described by Harley as sound axes [Harley, 1997]. This spatial movement is linked to pitch, as the entries begin with a very high trumpet note paired with a very low note on the trombone directly opposite, and end with a convergence around middle C (see Figures 7.5 and 7.6). This movement is quite different from the successive entry of the spatial assembly instruments as this time a relationship exists between the material, instead of each instrument having its own independent melody, key and meter. Brant used the term spill to describe the effect of spatially distributed musicians performing similar, harmonically related material. He uses the Tuba Mirum section of Berlioz's *Requiem* (see Figure 7.3) as an example and describes how the common tonality and tone quality of the four groups causes the resulting musical texture to extend from the corners to fill the room [Brant, 1967].

Throughout his long career, Brant continued to compose spatial works of greater and greater scale, culminating in *Bran(d)t aan de Amstel*, the massive spectacle of spatial music which encompassed most of the city of Amsterdam in the 1984 Holland Festival [Harley, 1997]. This huge work involved a colossal number of musicians including numerous bands in public squares, a youth jazz band, two choruses, two brass bands, four street organs and four boatloads of performers moving through the city's canals.

Ives' influence can also be seen in the spatial distribution scheme adopted by the early composers of electronic music in America in the early 1950s. The relatively recent development of magnetic tape was quickly adopted by composers as it greatly facilitated the editing and splicing together of different sounds. The Project for Music for Magnetic Tape was established in New York in the early 1950s and over the next two years this group produced three new electronics works, Cage's *Williams Mix* (1952), Earle Brown's *Octet* (1952) and Morton Feldman's *Intersection* (1953). Each piece was realized using eight unsynchronized monophonic tapes positioned equidistantly around the auditorium. The spatial separation of multiple independent musical layers, in this case electronically generated, is clearly reminiscent of the approach taken by Ives and Brant. In a lecture on experimental music given in 1957, Cage described this approach as follows;

> *"Rehearsals have shown that this new music, whether for tape or for instruments, is more clearly heard when the several loudspeakers or performers are separated in*

space rather than grouped closely together. For this music is not concerned with harmoniousness as generally understood, where the quality of harmony results from a blending of several elements. Here we are concerned with the coexistence of dissimilars, and the central points where fusion occurs are many: the ears of the listeners wherever they are. This disharmony, to paraphrase Bergson's statement about disorder, is simply a harmony to which many are unaccustomed [Cage, 1957]*".*

In a 1992 interview Brant stated that the main function of space in music is "to make complexity intelligible" [Harley, 1997] and Cage's "co-existence of dissimilars" is very reminiscent of this. Brant did, however, distinguish his music from Cage, stating that his approach "is opposed to what later came to be termed 'aleatoric' or 'indeterminate' music, in which accident and chance are looked upon as fundamental musical parameters. When uncoordinated rhythmic layers are combined with spatial distribution, accident is no more a factor than it is in the performance of rubato in a complex Chopin ratio [Brant, 1967]".

The music of Ives, Brant and Cage uses space as a fundamental musical parameter, as in this case, the spatial separation of the different musical lines is crucial for the correct perception of the work. However, although space is a critical aspect of this music, it is nevertheless used in a somewhat limited way, namely just to separate and clarify the various musical lines. While composers like Cage adopted a strictly indeterminate approach, others began to develop more formal systems to organize the spatial aspects of a work. Brant was well aware of the dangers in this approach, saying that "schemes for spatial distribution that are conceived in terms of their visual expressiveness on paper cannot be expected to produce any effect on the aural mechanism". He also discussed the degradation in spatial impression that occurs after a certain level of musical activity is reached, stating that "the impression of the sound travelling gradually down the first wall is very strong; this impression of moving direction becomes less well defined as the further entrances and accumulations occur [Brant, 1967]". However, Brant did use geometrical patterns to map out spatial trajectories, such as the sound axes of *Millennium* discussed earlier, and in the years to follow many more composers would also turn to geometric abstraction in an attempt to systematically organize the spatial relationships within a composition. This approach was undoubtedly influenced by the development of electronic music as now, for the first time, sounds could be entirely removed from their source and reproduced at will. The different aesthetics of electronic music which developed in the middle of the twentieth century would profoundly influence the use of space, not only in electronic music, but also in the way composers employed space in orchestral music.

8 Musique Concrète & Elektronische Musik

Fig. 8.1 Pierre Henry performing with the potentiomètre d'espace, Paris, 1952

The development of recording and amplification technology allowed composers to, for the first time, completely separate a sound from its physical source, and hence its physical location in space. The French composer and engineer Pierre Schaeffer conducted some of the earliest experiments with music based on recorded sounds while working at the French national broadcaster Radiodiffusion Française (RDF) in 1936. He continued this work after the war and by 1948 Schaeffer had completed a number of works in this new style of *Musique Concrète*. Compositions such as *Etude Violette* (1948) and *Aux Chemin de Fers* (1948) were broadcast on French national radio. In 1949, Schaeffer met the composer Pierre Henry and the two went on to found the first purpose-built electroacoustic music studio at the *Groupe de Recherche de Musique Concrète* (GRMC) at RDF in 1951. The two also collaborated on a number of compositions such as *Symphonie pour un homme seul* (1950) and *Orphee 51 ou Toute la Lyre* (1951), initially using turntables and then later magnetic tape. In 1952, Schaeffer published *À la Recherche d'une Musique Concrète* (The Search for a Concrete Music) which summarized the basic principles and working methods of *Musique Concrète* at that point. In this text, Schaeffer introduced the idea of the movement of sound along *trajectories sonores* (sonic trajectories) and the creation of spatial relief through the contrast of static spatial locations and dynamic mobile sound sources which are controlled manually by the performer. This new approach was demonstrated for the first time in a performance in Paris in 1951 which featured multiple monophonic turntables routed to four loudspeakers positioned to the left, right and rear of the stage, and a fifth loudspeaker placed overhead. During the performance, four tracks were routed to each loudspeaker while a fifth was spatially diffused live by a performer using the *potentiomètre d'espace,* a highly theatrical system which controlled the spatial distribution of the fifth track (see Figure 8.1).

The aesthetic of *Musique Concrète* is often discussed solely in terms of its use of recorded sounds. However, this term reflects the compositional working method as much as the source material. The traditional compositional process begins with an abstraction, the score, which is then eventually realized by the performer. *Musique Concrète* inverts this approach, as the composer now begins with real, concrete sounds which must be then manipulated to create an abstract piece of music. For Schaeffer, it is the sound itself which is of utmost importance and all of the activities of the composer and the listener should be directed towards the inherent properties of the sound itself. Schaeffer's compositional aesthetic was therefore incompatible with traditional approaches to composition, or with the principles of serialism, which more and more composers began to adopt in the post-war period. Serialism developed initially from the twelve-tone music of composers such as Arnold Schoenberg, Alban Berg and Anton Webern and was adopted by many composers as a means of departing from the functional tonality of previous eras. Later composers generalized this idea and began to use a series of ordered elements, such as a twelve-tone pitch set, as a structuring principle to order and manipulate other parameters within the composition such as rhythm and dynamics. Since serialism presented itself as a denial of tonality, Schaeffer saw no point in applying serial methods to concrete material [Palombini, 1993], as this organizing principle is imposed upon the sounds, instead of being derived from them. Nevertheless, Schaeffer's ideas were well received at a lecture at Darmstadt festival in 1951 and, perhaps due to the influence of visiting composers like Boulez, Messiaen and Stockhausen, serialist tendencies, although resisted by Schaeffer, began to emerge within the GRMC [Palombini, 1993]. This eventually lead to the resignation of Schaeffer, Henry and others from the GRMC in 1958 and the founding of a new collective, the *Groupe de Recherches Musicales* (GRM), which was later joined by composers such as Luc Ferrari, Iannis Xenakis, Bernard Parmegiani, and François Bayle.

Meanwhile in Germany, another electronic music studio was established which would be far more amenable to the tenants of serialism. The physicist Werner Meyer-Eppler had published a thesis in 1949 on the production of electronic music using purely electronic processes [Meyer-Eppler, 1949] and in 1951, Meyer-Eppler, with Robert Beyer and Herbert Eimert, established a new studio for this purpose in Cologne at the *Nordwestdeutscher Rundfunk* (NWDR). Their approach to *Elektronische Musik*, differed from *Musique Concrète* in its use of synthesized sounds rather than recorded acoustic sounds. However, both aesthetics also differ fundamentally in terms of their basic approach to electronic music composition. Composition with sound synthesis is inherently more suited to abstract structuring principles such as serialism, as the material can be deliberately generated to fit the preconceived score. This is much more difficult with the complex acoustic sounds of *musicque concrète* and indeed Schaeffer struggled to find a formal compositional system which originated from the intrinsic properties of the sounds. Schaeffer's disappointment is evident in this quote from an interview in 1986;

> *"I fought like a demon throughout all the years of discovery and exploration in musique concrète. I fought against electronic music, which was another approach, a systemic approach, when I preferred an experimental approach actually working directly, empirically with sound. But at the same time, as I defended the music I was working on, I was personally horrified at what I was doing. . . . . . I was happy at overcoming great difficulties - my first difficulties with the turntables when I was working on Symphonie pour un Homme Seul, my first difficulties with the tape recorders when I was doing Etude aux objects - that was good work, I did what I set out to do. My work on the Solfege - it's not that I disown everything I did - it was a lot of hard work. But each time I was to experience the disappointment of not arriving at music. I couldn't get to music, what I call music. I think of myself as an explorer struggling to find a way through in the far north, but I wasn't finding a way through. "* [Hodgkinson, 2001]

The dichotomy between these two approaches is characteristic of the divergent approaches to electronic and electro-acoustic music composition and performance that has existed throughout its fifty year history, and this is equally true of their respective approach to the use of space. Is the focus on the sound itself or the relationship between sounds, does the spatial movement or distribution arise from the sound itself or is it imposed upon the sound?

The NWDR studio in Cologne swiftly became one of the most well-known electronic music studios in the world, helped in no small part by the growing fame of the composer Karlheinz Stockhausen, who had joined the studio in 1953. Stockhausen had worked briefly with Schaeffer in 1952 and produced a single *Konkrete Etüde* in 1952. Stockhausen's approach to composition at this time was based upon the ideas of total serialism, meaning the application of a measurable scale of proportions (the series), not only to pitches, but also to non-pitched parameters such as timbre, rhythm and also space. Schaeffer's *musicque concrète* would eventually lead to an aesthetic in which space is used in a performance to highlight and exaggerate the pre-existing content of the music, and this will be looked at in more detail later in this chapter. The alternative legacy of composers like Stockhausen can be seen in the large number of abstract schemes which composers began to use to organize the spatial distribution of material.

8.1 Stockhausen and the Serialization of Space

The German composer Karlheinz Stockhausen (1928-2007) is widely regarded as one of the most important and influential composers of the twentieth century. Throughout his career, Stockhausen composed numerous works of spatial music and also published various treatises on electroacoustic composition and the use of space in music. Although often a controversial figure, Stockhausen was highly influential, particularly in the development of electronic music, and was one of the few *avant-garde* composers to become well known to the general public.

Stockhausen completed his undergraduate studies at the Cologne Conservatory of Music in 1951 and went on to study serialist techniques at the Darmstadt composition school. After studying for a short time in Paris with Olivier Messiaen and Pierre Schaeffer, Stockhausen returned to Germany in 1953 and took a position as an assistant to Herbert Eimert at the NWDR studio in Cologne. In his early career, Stockhausen primarily composed "totally organized" serial music, in which each musical parameter is controlled by a series of numerical proportions. Stockhausen believed that the precedent for this extension to the serial principle could be found in the works of Webern and he made the following comment about Webern's *Concerto,* *Opus 24*, "What is essential is not a uniquely chosen gestalt (theme, motive), but a chosen sequence of proportions for pitch, duration and volume [Morgan, 1991]". Electronic music would be the ideal medium to implement these ideas as it allowed the composer to create the musical material and hence organize timbre and space according to a series of proportions. These ideas were implemented by Stockhausen in the electronic work *Gesang der Jünglinge* (1955/1956) which has been described as "the first masterpiece of electronic music [Simms, 1996]" and was also the first piece to serialize the projection of sound in space [Smalley J., 2000]. In this piece, Stockhausen attempts to forge a connection between recordings of a boy soprano and electronically synthesized sounds ranging from white noise to sine tones. Stockhausen categorized the vocal recordings into basic phonetic components such as noise-like plosive consonants and vowel sounds which resemble pure tones. These were then combined with artificial consonant and vowel-like sounds created from layered sine waves and filtered white noise to produce a range of material which fills a continuum of timbres from pure tones to noise. This positioning of material within a continuum of timbres illustrates Stockhausen's conception of serialism as a graduated scale of values between two opposing extremes which he later described in an interview in 1971;

> *"Serialism is the only way of balancing different forces. In general it means simply that you have any number of degrees between two extremes that are defined at the beginning of a work, and you establish a scale to mediate between these two extremes. Serialism is just a way of thinking."* [Cott, 1973]

Although both recorded and synthesized sounds are used in *Gesang der Jünglinge*, every sonic event and parameter (including spatial locations and the comprehensibility of the text) was controlled and organized serially. The original work was performed using five groups of loudspeakers at its premiere in 1953 but was subsequently mixed down to four tracks. No five-track tape machines existed at this time, so a four-track machine was used to feed four loudspeakers positioned around the audience while an additional tape machine was used to feed a fifth loudspeaker positioned on stage (see Figure 8.2 [Smalley J., 2000]). Although it is known that Stockhausen attempted to organize space in this work in the same way as every other parameter, the exact means by which serial techniques were applied to the spatial distribution and movement is not entirely known [Smalley J., 2000]. The spatial distance, especially of the boy's voice, is clearly an important aspect of this work, and the composer relates this parameter to the comprehensibility of the voice. When the voice is positioned close-by with little reverb, the text is clearly comprehensible and is the primary focus of attention. However, as the voice recedes into the distance and the level of reverberation increases, comprehensibility decreases and the complex interaction of the voice with itself and other similar electronic elements becomes the primary focus of concern [Moritz, 2002].

Fig. 8.2 *Gesang der Jünglinge* premiere at WDR Cologne, 1956

Stockhausen's theoretical writings from this time are primarily concerned with the difficulties in applying serialist proportions to non-pitched parameters such as space and timbre. Stockhausen was well aware of the difficulties inherent in this approach, particularly in terms of timbre, writing that "we only hear that one instrument is different from another, but not that they stand in specific relationship to one another", and it is clear that a similar problem exists in terms of the perception of spatial locations and movements. Stockhausen's solution is described in the well-known text, *Wie die Zeir Vergeht* (how time passes) which was published in 1955 [Stockhausen, 1959]. In this essay, Stockhausen presents a theory of “the unity of musical time” in which every musical parameter is considered at different temporal levels. So, for example, the macro-level rhythm of an individual musical phrase can be related to the specific micro-level subdivision of the spectrum by its harmonic structure, which is also similarly related to timbre [Morgan, 1991]. At even greater durations, the entire composition can be considered as a timbre with a spectrum derived from the pitch, rhythm and dynamic envelope of the combined individual musical phrases. This new approach moved the focus away from the pointillistic note relationships of early serialism and onto what Stockhausen referred to as "group composition", which emphasized the overall character of large groups of proportionally related material, rather than on the relationship between individual pitches. This approach is demonstrated in *Gruppen* (1955-57), for three orchestras positioned to the left, in front and to the right of the audience. The spatial separation of the three orchestras was initially devised to clarify the carefully constructed relationships between the three layers of material, which is clearly reminiscent of the approach adopted by Brant and Ives. However, in certain passages, musical material is passed from one group to another and Stockhausen notes that similar orchestration was deliberately used for each of the three groups in order to achieve this effect. The spatial movement was produced using overlapping crescendos and decrescendos (see Figure 8.5) which are clearly reminiscent of stereophonic panning and illustrates how Stockhausen's experience with electronic music composition influenced his composing for instruments. Indeed, Stockhausen originally intended to write *Gruppen* for both orchestral and electronic forces but the electronic element was eventually abandoned due to practical and economic constraints [Misch *et al*, 1998]. The composer describes the spatial aspects of this work as follows;

> *"The spatial separation of the groups initially resulted from the superimposition of several time layers having different tempi – which would be unplayable for one orchestra. But this then led to a completely new conception of instrumental music in space: the entire process of this music was co-determined by the spatial disposition of the sound, the sound direction, sound movement (alternating, isolated, fusing, rotating movements, etc.), as in the electronic music Gesang Der Jünglinge for five groups of loudspeakers, which was composed in 1955/56* [Moritz, 2002]*".*

The spatial segregation of the three groups clearly helps to distinguish the different layers of material but the more elaborate spatial effects also seem to be quite effective. Various reviews of performances of this work have commented on the dramatic effect of a single chord in the brass instruments travelling around the hall, and this spatial movement is also quite apparent in stereo recordings of this piece (in which the three groups are generally panned hard left, centre and hard right). The approach to space used in *Gruppen* was further developed in *Carre* (1959-1960) which was composed by Stockhausen and his assistant at the time, the British composer Cornelius Cardew. This piece was composed for four orchestras which as the name suggests (*Carre* literally means square in French) were arranged in a square around the audience, as shown in Figure 8.6 As with *Gruppen*, similar orchestration is used for each group with the addition of a mixed choir of eight to twelve singers. The vocal text consists largely of phonemes and other vocal sounds chosen for their relationship to the instrumental sounds.

Fig. 8.3 Rehearsal of *Gruppen* in Cologne, March 1958

Fig. 8.4 Dress rehearsal of *Carre* in Hamburg, October 1960

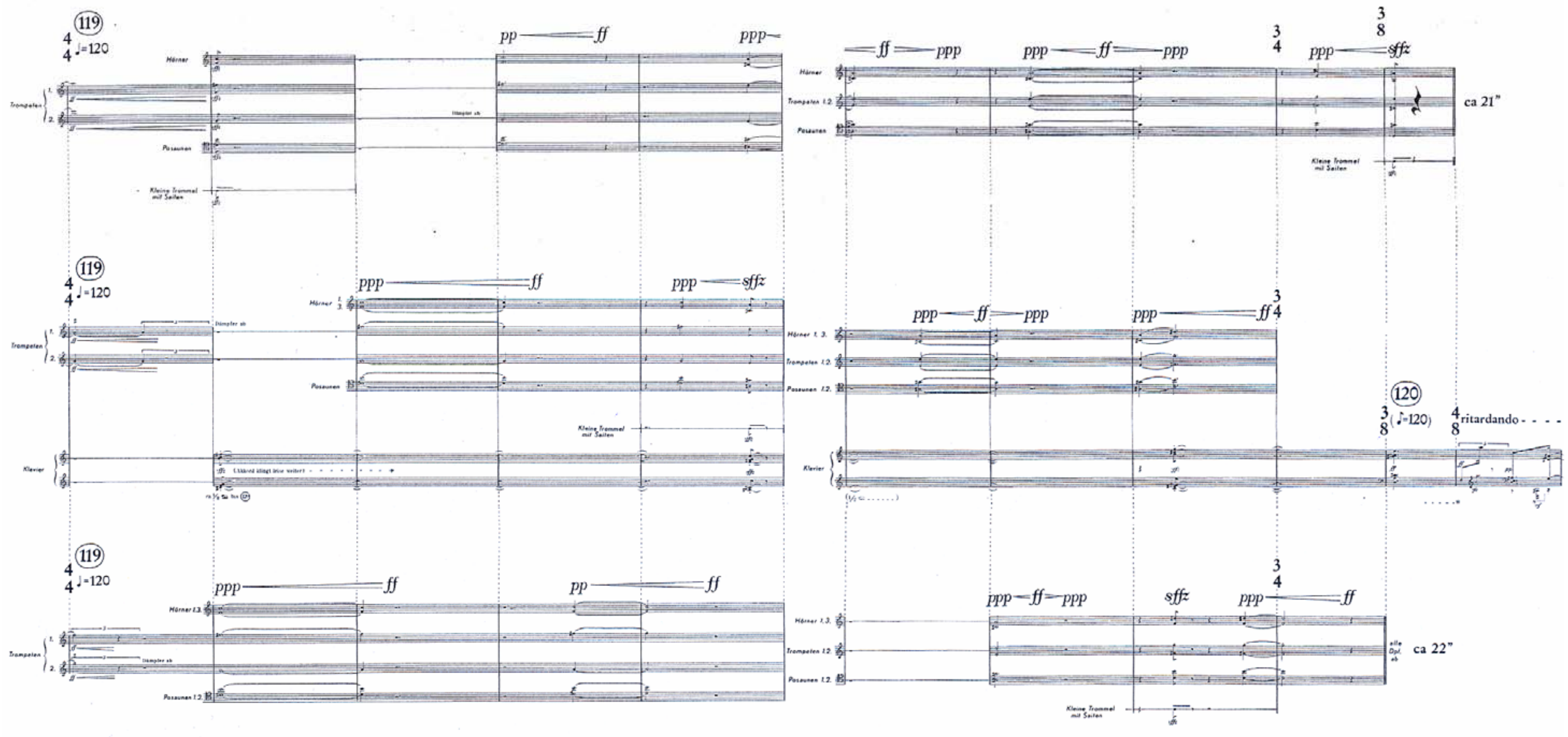

Fig. 8.5 Spatial movement in *Gruppen*

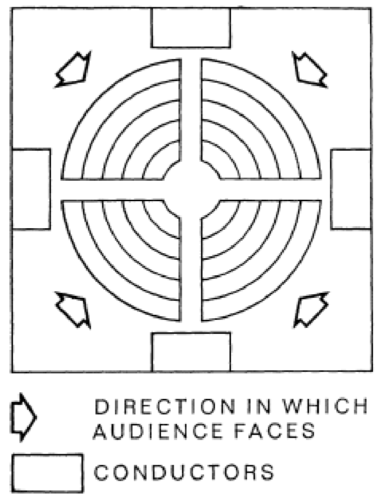

Fig. 8.6 Orchestra & audience layout in Stockhausen's *Carre*

In 1958 Stockhausen published two articles in the German music journal Die Riehe entitled *"Electronische und Instrumentale Musik* (Instrumental and Electronic music) [Stockhausen, 1975a]" and "*Musik in Raum* (Music in Space) [Stockhausen, 1975b]". In the latter essay, Stockhausen briefly discusses the history of Western spatial music before going on to outline the functional use of space in his own work. The composer describes two difficulties which often occur in serialist music. Firstly, the extensive use of serialist processes results in music in which every parameter is constantly changing, and this often results in a rather static, pointillistic texture. Secondly, in order to articulate longer time-phrases, one parameter must remain constant and dominate, but this is in direct contradiction with the serialist aesthetic. Stockhausen suggests that the spatial distribution of sounds can be used to articulate longer time-phrases and structure the material, and also to clarify the complex relationships between the different layers [Stockhausen, 1975b]. Stockhausen also discusses the perception of distance in terms of air absorption, reverberation and early reflections which, despite the typically characteristic terminology, compares relatively well to the current research reviewed earlier in this thesis. Stockhausen discusses the relative nature of the perception of distance, and notes the importance of source recognition in this context and the difficulty in estimating the distance of unknown, synthesized sounds. Consequently, Stockhausen concluded that the distance of a source is a secondary musical parameter which is determined by the timbre and loudness of the source signal, and that the direction of the source should be the primary spatial parameter which is related to overall serial structure. Interestingly

Stockhausen suggests that proportional changes in location should be used instead of fixed scale steps and the appropriate minimum change in value should be derived experimentally.

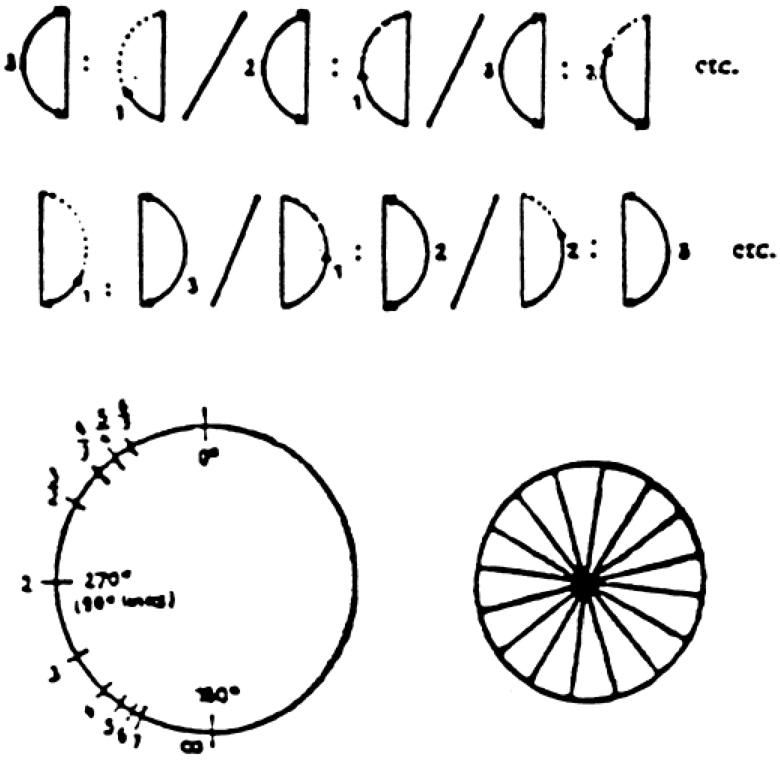

Fig. 8.7 Spatial intervals and directions from Stockhausen's *Musik in Space*

Stockhausen's next work, *Kontakte* (1958-60), would complete the idea for a combination of instrumental and electronic music originally considered for *Gruppen*. This famous work can be performed as a four-channel tape piece or as a mixed-media work for four-channel tape, live piano and percussion. As the title suggests, this work explores the various points of contact between static instrumental sounds and dynamic electronic textures, between different spatial movements, and also the temporal relationship between pitch and rhythm. At one point in this piece, Stockhausen dramatically illustrates his theory of the "unity of musical time" as a high-frequency, clearly pitched tone smoothly transforms into a slower rhythmical procession of clicks. Although direction, or more precisely changes in direction, is the primary spatial element in *Kontakte*, distance is also used to support dynamic aspects of the music, such as the fast oscillating spatial movements which emphasize certain loud dynamic sounds. In addition, spatial depth is suggested through the contrast of a dense layer of material in the foreground which is suddenly removed to reveal another more distant layer in the background. In order to generate dynamic spatial trajectories and rotation effects, Stockhausen developed a rotating loudspeaker mechanism which is surrounded by four microphones (see Figure 8.8). Sounds played back with the rotating loudspeaker are recorded and reproduced using four corresponding loudspeakers positioned around the audience. The physical rotation of the loudspeaker results in a Doppler shift, time varying filtering, phase shifts and other distortions which are difficult to accurately reproduce electronically [Roads, 1996]. Six distinct spatial movements are used in *Kontakte*, namely rotation, looping, alternation, a static distribution with a duplicated source, a static distribution with different sources, and single static sources. These spatial motifs are implemented at different speeds and directions and interact with each other, and with the static instrumental performers. This series of spatial movements is treated in the same way as pitch and rhythm using a system based on the change in speed and angular direction. In fact, Stockhausen equates the spatial movements with rhythm, arguing that changes in spatial location can articulate durations in exactly the same way [Cott, 1973]. The interaction between the instruments and the electronics is one of the most important aspects of this work and, as in *Gesange der Junglinge*, Stockhausen effectively positions the concrete sounds, now a piano and percussion instead of a boy soprano, within the electronic texture. Dramatic crescendos in the instruments seem to trigger electronic textures while at other times the pitch and rhythmic motion of the instrumental parts reflect the spatial motion of the electronic sounds.

Fig. 8.8 Stockhausen with his rotating loudspeaker mechanism

8.2 Architectural Spatial Music

> *New halls for listening must be built to meet with demands for spatial music. My idea would be to have a spherical chamber, fitted all round with loudspeakers. In the middle of this spherical chamber, a platform, transparent to both light and sound, would be hung for the listeners. They could hear music, composed for such adapted halls, coming from above, from below and from all directions.* [Stockhausen, 1975b]

One of the high points of *avant-garde* electronic music was undoubtedly the extravagant international expositions of the late 1950s and 60s which featured multimedia art works and music and performances by such prominent composers such as Karlheinz Stockhausen, Edgard Varèse and Iannis Xenakis. An elaborate multimedia environment was commissioned by Philips for the 1958 Brussels World's Fair to showcase their new advances in audio and visual technology. The pavilion (see Figure 8.9) was designed by the architect and composer Iannis Xenakis in the form of an "electronic poem" [Zvonar, 2004]. It contained within its unusual hyperbolic paraboloid structure a long unbroken projection surface with elaborate lighting and projection equipment and an eleven channel multi-track tape system which could be routed to over four hundred loudspeakers. The main musical element of the program consisted of Edgard Varèse' tape piece *Poème Electronique*, which was synchronized to the visual effects and dynamically distributed through the space in nine different "sound routes" using a switching system controlled with an additional tape track.

The architect and designer of the Philips Pavilion, Iannis Xenakis, was also a composer and his tape piece *Concrète PH* was played as an interlude between shows. Xenakis had moved to Paris from his native Greece in 1947 and took a position as an assistant to the architect Le Corbusier. In his spare time, Xenakis studied composition and produced his first major work, *Metastasis* in 1954 after studying with Olivier Messiaen. An architectural influence is evident in the score for this work which features massed string glissandi in which each individually notated instrument is arranged according to the large scale, deterministic structure (see Figure 8.10). In a 1954 article, Xenakis describes this approach as follows, "the sonorities of the orchestra are building materials, like brick, stone and wood. . . the subtle structures of orchestral sound masses represent a reality that promises much [Hoffman, 2001] ". Xenakis went on to use the structural design of *Metastasis* as the basis for the wall curvature of the Philips Pavilion and a similar hyperbolic paraboloid structure is evident in both the Pavilion (Figure 8.9) and the score (Figure 8.10).

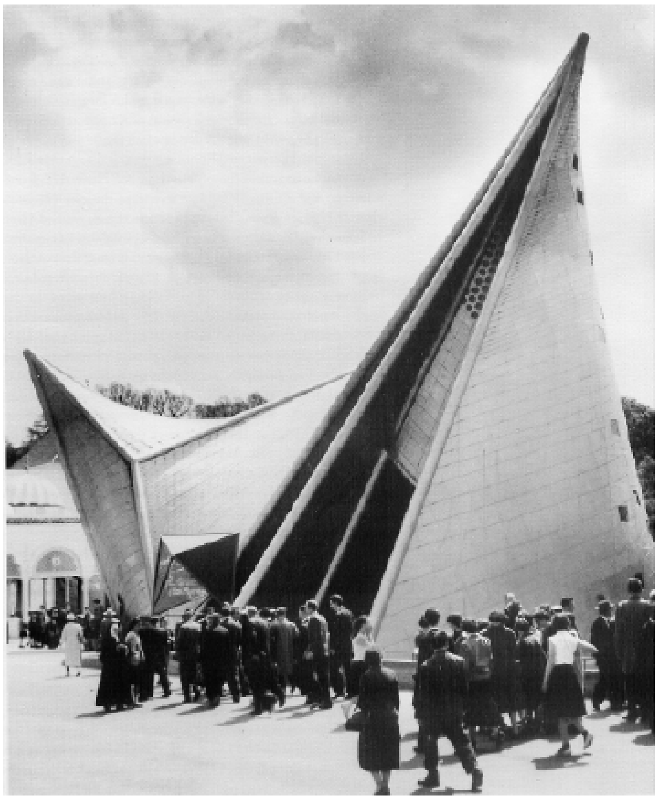

Fig. 8.9 The Philips Pavilion at the 1958 Worlds Fair in Brussels

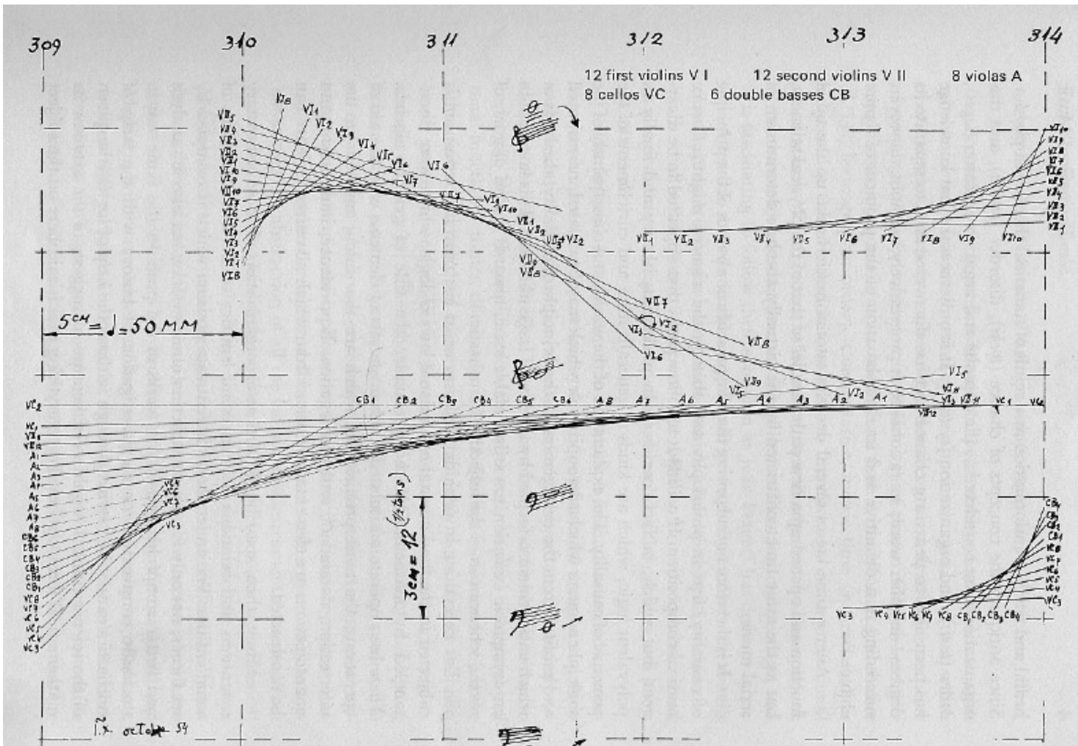

Fig. 8.10 String glissandi, bars 309-14 of *Metastasis* by Iannis Xenakis

Many of Xenakis' compositions make use of abstract mathematical theories such as probability theory, game theory or the kinetic theory of gases. These mathematical functions were used to structure individual elements within the overall design, which Xenakis often considered in spatial, almost architectural terms. Xenakis rejected serialism as a compositional principle, but his mathematically formalized organization of pitch, duration, timbre, dynamic and spatial location into a unifying overall structure [Xenakis, 2008] is reminiscent of total serialism. However, his use of formalized mathematical techniques to design large-scale geometrical sound masses is very different from the moment to moment form of serialist composers like Stockhausen. Xenakis's geometrical conception of spatial music was perhaps influenced by his collaborator at the 1958 Worlds Fair, Edgard Varèse, who made the following comment about the future of spatial music in 1936;

> "*When new instruments will allow me to write music as I conceive it, the movement of sound-masses, of shifting planes, will be clearly perceived in my work, taking the place of the linear counterpoint. When these sound-masses collide, the phenomena of penetration or repulsion will seem to occur. Certain transmutations taking place on certain planes will seem to be projected onto other planes, moving at different speeds and at different angles. There will no longer be the old conception of melody or interplay of melodies. The entire work will be a melodic totality. The entire work will flow as a river flows* [Varèse, 1998]. "

Xenakis continued to create large-scale multimedia art works after the Brussels Worlds Fair and he referred to this new form of multimedia art as the polytype, from the Greek *polys* (many, numerous) and *topos* (place, space, location) [Harley, 1998b]. The polytype therefore refers to the many small elements of light and sound which create and articulate a larger space such as the elaborate audiovisual system in the Philips Pavilion. Xenakis also suggested that geometrical shapes could be created through the sonic projection of sounds in defined trajectories around a loudspeaker array. So for example, a single sound played successively and with a slight overlap through a circular array of loudspeakers would produce a circle, while many short impulses played through a group of loudspeakers would produce a sonic surface. Xenakis utilized Peirre Schaefer's concept of spatial relief and divided the spatial distribution between *stereophonie statique* (sound emanating from numerous static points in space) and *stereophonie cinematique* (sounds produced by multiple, mobile sources) [Harley, 1998b]. The composer suggested that the different spatial patterns and distributions could then be structured and related to create a form of spatial counterpoint.

Xenakis attempted to implement these spatial designs in the orchestral work *Terretektorh* (1965-66). In this work, the musicians are divided into eight groups and arranged in six concentric circles (as shown in Figure 8.11 [Hoffman, 2001]) while the audience is unusually distributed amongst the musicians. Various geometrical patterns and movements are implemented within the five main spatial distributions which are used to structure the piece, namely;

- stochastically distributed points
- sound-planes with internal movement
- static sounds
- densely woven individual lines
- continuous glissandi

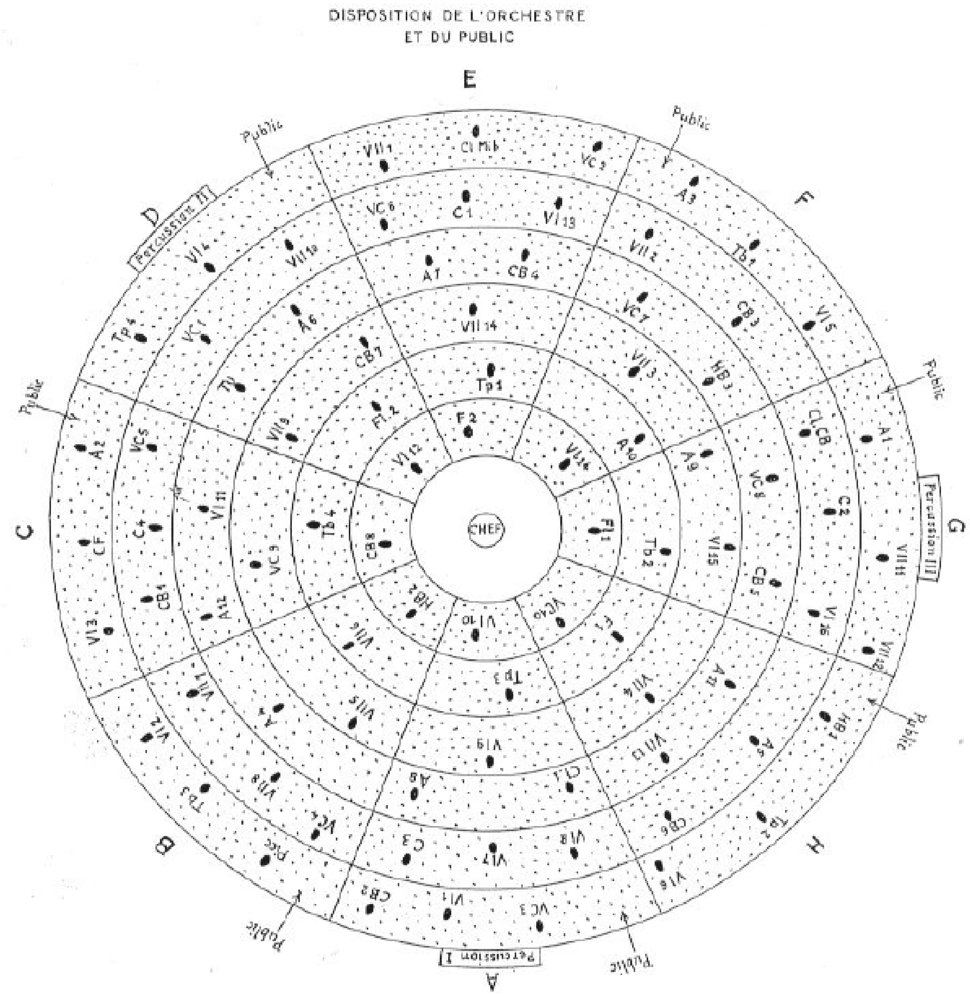

Fig. 8.11 Orchestral disposition of *Terretektorh*

Xenakis would go on to compose numerous large scale multimedia works in which abstract geometrical designs are used to control the distribution of both instrumental and electronic sounds, and lights and projected visuals. *Hibiki-hana-ma* for multi-channel tape was composed for the 1970 Worlds Fair in Osaka and was performed in the Japanese pavilion through eight hundred loudspeakers arranged above, around and under the audience. Stockhausen was also present at Osaka 70 and his music was performed twice a day for 183 days in the German Pavilion by twenty instrumentalists and singers (see Figure 8.12). The composer himself controlled the sound projection from a position in the centre of the spherical venue which contained fifty-five loudspeakers arranged in seven rings, and six small balconies for the musicians. The design of the auditorium was largely based on Stockhausen's specifications, which are clearly expressed in the quote at the beginning of this section. As with his earlier works, Stockhausen again used spatial trajectories to articulate different musical events. However, the spatial distribution was now implemented in real-time. This approach is somewhat similar to the live spatial diffusion practised by Pierre Schaefer, albeit with a greater emphasis on geometrical paths. Stockhausen describes the experience as follows:

> *"I can create with my hand – up to six or seven revolutions a second. And you can draw a polyphony of two different movements, or let one layer of sound stay at the left side, then slowly move up to the centre of the room, and then all of a sudden another layer of sound will start revolving like mad around you in a diagonal circle. And the third spatially polyphonic layer will just be an alteration between front-right and back-left, or below me and above me, above me and below me, alternating. This polyphony of spatial movements and the speed of the sound become as important as the pitch of the sound, the duration of the sound, or the timbre of the sound.* "[Cott, 1973].

Although Stockhausen would continue to argue for the construction of concert halls specifically for the performance of spatial music, such as the pavilion in Osaka, the general trend from the seventies onwards was toward more generalized multichannel solutions such as Quadraphonics and Ambisonics. Stockhausen was one of the first composers to adopt an arrangement of eight loudspeakers and he used this arrangement extensively for the rest of his career. His first composition for this arrangement, *Sirius* (1975-77) utilized manual diffusion of electronic sounds, live instrumentation in the form of trumpet, soprano, bass clarinet, and bass, and an eight-channel version of the rotational table used in *Kontakte*. In an interview after the premiere of *Sirius,* the composer described the use of rotational effects in this work.

> "*Sirius is based entirely on a new concept of spatial movement. The sound moves so fast in rotations and slopes and all sorts of spatial movements that it seems to stand still, but it vibrates. It is [an] entirely different kind of sound experience, because you are no longer aware of speakers, of sources of sound – the sound is everywhere, it is within you. When you move your head even the slightest bit, it*

changes color, because different distances occur between the sound sources. ”
[Felder, 1977]

Fig. 8.12 Karlheinz Stockhausen performing at OSAKA 70

8.3 The Perception of Abstract Spatial Designs

“The predominance of abstract designs over audibility of their sonorous results is the basic problem in post-war modernist music [Harley, 1998a]. “

The development of electronic processes which could dynamically move sounds through space encouraged composers such as Stockhausen to attempt similar dynamic processes in instrumental spatial music. In “*Musik in Raum”*, Stockhausen suggests that spatial distribution could be used to structure and articulate different layers of material, thereby avoiding the homogenous pointillistic textures common to many twelve-tone compositions. In many respects, this is similar to Brant’s conception of spatial music, yet the American was highly dismissive of Stockhausen and he criticized *Gruppen* as not really being spatial because “all the orchestras have brass, woodwinds, and percussion, so the direction and the tone quality cannot indicate the source of the material [Harley, 1997]”. However, this criticism is not entirely justified as Stockhausen is clearly using the spatial distribution of the musicians to articulate different layers of material, albeit with similar instrumentation. While the delineation of spatial locations with different timbres may be necessary when many disparate layers of material are operating concurrently, such as in the music of Henry Brant, this is not necessarily required when each group produces intermittent, rhythmically and harmonically distinct passages of music, as in the moment to moment form typically employed by Stockhausen.

One of the most contentious aspects of Stockhausen’s treatise on spatial music was his proposal for the serialization of direction based upon the division of a circle into a series of proportions. Critics argue that the absolute directions specified on paper will not be perceived by the audience [Harrison, 1999; Harley, 1998a], and the tests discussed earlier in this thesis would seem to support this view. The results of these tests illustrate the significant variations in perceived source direction which occur with electronic spatialization systems and suggest that a reliable and absolute perception of direction is difficult to achieve electronically. This problem is exacerbated in a performance setting when the audience is seated at different locations within the loudspeaker array. The apparent contradiction between the compositional design and the perceptible results in the music of Stockhausen, and other composers, is not only evident in their use of space, but also in other areas, as illustrated by the quote at the beginning of this section. Some composers questioned whether the equal division of parameters such as pitch, motion, duration and form on paper, translates to an equal division in the perception of the listener [Bernard, 1987]. The twelve-tone music of the Vienna school which preceded serialism had primarily focussed on a serialized pitch row which was, in theory at least, evident in the resulting pointillistic texture. By 1955, however, Stockhausen had begun to use these serialist procedures to control the overall character of large groups of material, in what he referred to as "group composition". In this aesthetic, it is questionable whether the audience is intended to perceive the absolute and tightly controlled internal relationships within the composition. Instead perhaps, it is the overall result of these procedures which is important. Serialism began as a negation of tonality, and serialist procedures serve to eliminate any perceivable melody or theme, or any defined tempo or rhythm. These purely negative goals could also be achieved by simply randomizing every parameter, yet the controlled bursts of activity in *Kontakte* or *Gruppen* sound anything but random. In effect, the serialist procedures eliminate any repetition, of pitches, rhythms or indeed spatial movements, while also maintaining a definite coherence in the audible material. In these works, direction is therefore divided into a series of proportions in the same way as every other parameter, and it is the consistency of this approach which produces the non-repeating, yet entirely coherent material in these works. The precisely specified spatial locations presented in "*Musik in Raum"* are therefore perhaps not intended to be perceived in an absolute sense, but rather as a way to remove any recognizable or re-occurring spatial motifs. Stockhausen therefore uses space to create a sense of perspective in the listener which is not in fact fixed, as in the classic concert listening experience, but varying. This approach ties in with the composer's overall aesthetic in which the music is similarly removed from a definite tonal, harmonic or rhythmical centre [Cott, 1973].

Stockhausen's use of space, particularly in his early career, suggests that the composer was well aware of the perceptual issues with his approach. "*Musik in Raum"* contains a detailed and relatively accurate assessment of the perception of auditory distance and also suggests that a perceptually appropriate scale of directions should be determined by listening tests. In a later interview, Stockhausen clearly recognizes the importance of the source material in terms of its perceived spatial movement:

> "*I can say in general that the sharper the sound and the higher the frequency, the better it moves and the clearer its direction. But I'd also say that the more a sound is chopped, the better it moves in space. And the sharper the attack of each segment of a sound event, the better it moves. Whereas a continuous or harmonious sound or a spectrum of vowels doesn't move very well*". [Cott, 1973]

Stockhausen integrated certain spatial parameters, such as direction, speed and angular movement into the overall serial structure but other parameters such as distance are used in a much more intuitive and indeed dramatic manner. For example, while the directional movements in *Kontakte* are serially controlled, distance is used to dramatically highlight certain climactic passages. Stockhausen's theatrical and dramatic use of space is also evident in the *Helicopter String Quartet* (1995), which uses the sound of the quartet (who are distributed amongst the helicopters) is mixed with the sound of the helicopter rotors. Stockhausen's sophisticated use of spatial depth and distance is sometimes ignored in certain ideologically driven criticisms (to which Stockhausen was no stranger) which concentrate solely on the serial control of direction discussed in *Musik in Raum.* Consider the following statement:

> "*Stockhausen dismisses the use of space by Gabrieli, Berlioz, and Mahler as being too theatrical, and argues instead that direction is the only spatial feature worthy of compositional attention because it could be serialized.* "[Harley, 1998a]

Although Stockhausen does state that direction is the only spatial parameter suitable for serialization, he clearly does not consider this to be "the only spatial feature worthy of compositional attention [Harley, 1998a] ". As discussed previously in this Chapter, distance is used extensively in *Kontakte* and *Gesang der Jünglinge* but, as this parameter is highly subjective, and dependent on multiple parameters such as amplitude, timbre and the nature of the source signal, it is not suitable for serial control.

Stockhausen's acoustic spatial music also displayed an awareness of the limitations of the medium. His use of spatial movement in *Gruppen* was one of the first attempts to replicate stereophony using acoustic instruments and is particularly successful in this regard due to the relative restraint displayed by the composer. Spatial movements are implemented, often in isolation, as short, distinct gestures, and this helps to clarify the perceived movement. A more complex spatial movement is in fact implemented only once in the dramatic and famous passage shown in Figure 8.4 where a single travelling chord rotates around the distributed brass instruments. The results of experiments presented in Chapter Six indicate that the presence of an additional distracting stimulus reduced the localization accuracy for the primary source [Blauert, 1997]. This suggests that when these spatial movements are isolated in this way, they are much more likely to be clearly perceived by the audience. However, if multiple complex trajectories are occurring simultaneously, it will be much harder for the listener to determine the precise trajectory of each source. In addition, the fragility of movements created using stereophonic processes has been clearly demonstrated and this too imposes perceptual limitations on the complexity of the spatialization process. It would therefore appear to be quite difficult to justify the elaborate and visually-orientated spatial schemes implemented by composers such as Varèse and Xenakis. Xenakis appeared to reach the same conclusion much later in his career, as evident in the following comment made by the composer in 1992.

> *"In reality, sound movements are usually more complex and depend on the architecture of the performance space, the position of the speakers and many other things. When you want to reproduce such a complicated phenomenon with live*

musicians playing one after another with amplitude changing in the same way that you change the levels in a stereo sound projection, sometimes it will work and sometimes it will not. It depends on the speed of the sound as well as on the angle of two loudspeakers or musicians, that is, on the relative position of the listener. These two considerations are equally important.
Xenakis "Music, Space and Spatialization, 1992: 6–7)

While the spatial movement of sound obviously lends itself to graphical geometrical representation, there is a danger in simply equating our visual perception with our audible perception. A visual representation of a circle is outside of time, in that the entire shape can be comprehended all at once, however, this is not the case when a sound is moved around the audience in a circular fashion. In this case, the perception of the sound "shape" will only become clear once the sound has moved through a full 360 degrees rotation. Creating a spatial counterpoint between different trajectories based on a graphical representation is therefore difficult, as the precise "shape" is not instantaneously clear as it is in the visual realm. Harley illustrates this problem in an analysis of the *Genesis cycle* (1962–3), a large-scale orchestral work by Henryk Gorecki. The chaotic and highly dissonant material in this piece is organized spatially into blocks of material, in a similar fashion to Brant and the group compositions of Stockhausen. However, this piece also attempts to create geometric spatial trajectories using triangular, polygonal and rectangular arrangements of the orchestra, as shown in Figure 8.13. This elaborate seating plan necessitated the repositioning of the audience and performers before each of the three movements. However, Harley correctly suggests that the slight effect of the different layouts is far less important than the symmetrical movement of material from left to right, or from front to back [Harley, 1998a], which could be achieved with any of these arrangements. The highly visual conception of spatial music envisaged by composers like Edgard Varèse and Iannis Xenakis is perhaps achievable when combined with a synchronized visual component. However, it is difficult to see how distinct sonic "shapes" can be reliably produced even in isolation, with either acoustic or electronic techniques. The creation of a spatial counterpoint between multiple shapes and trajectories appears even less realizable.

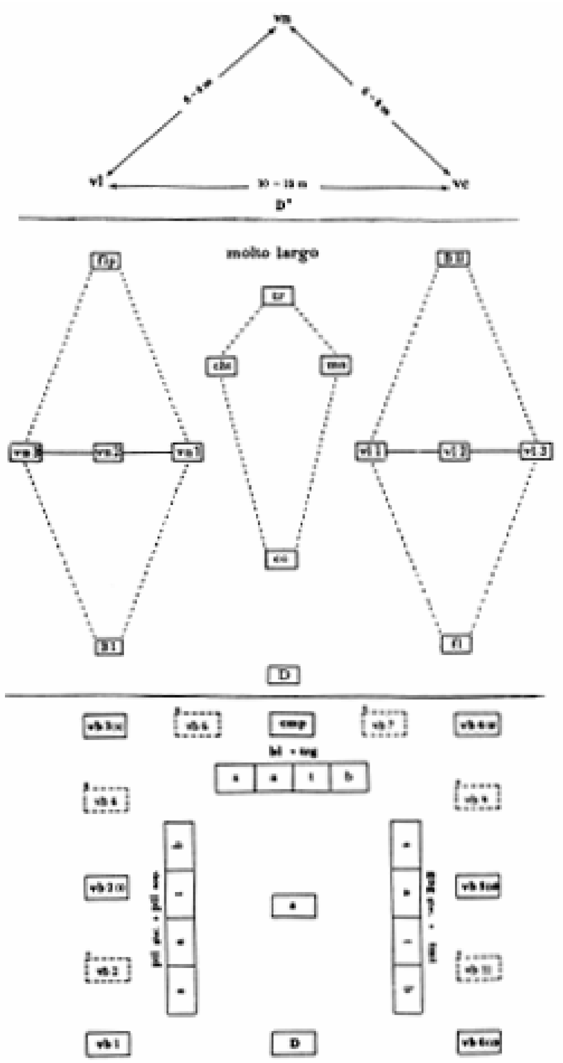

Fig. 8.13 Layout of instruments in Gorecki's *Genesis Cycle* 1974

As mentioned earlier, the aesthetics of *Musique Concrète* and *Elektronische Musik* which developed in the 1950s can be used to illustrate two divergent approaches to electronic music, and also to the use of space in music. The latter approach, which was discussed in the preceding Section, represents a continuation of traditional modes of composition in which abstract structuring systems are used to organize sounds and their position in space. The alternative approach, which has its origins in the work of Pierre Schaeffer, Pierre Henry and *Musique Concrète,* uses the sounds themselves as a starting point, and the large scale structure and spatial distribution is derived from the intrinsic properties of the sounds. Over the past fifty years, this alternate approach has developed into a set of performance practices based upon the manual diffusion of a stereo source to a loudspeaker orchestra and this aesthetic will be examined in the following section.

8.4 *Diffusion, and the legacy of Musique Concrète*

The legacy of *Musique Concrète* can still be seen in the international community of composers who create stereo electroacoustic works specifically for live diffusion to a loudspeaker orchestra. The term *acousmatic music,* meaning sound which is heard without a corresponding visible cause, is often used to describe this aesthetic. However, the exact definition of this term is disputed as it may simply refer to the means of production, i.e. music for loudspeakers alone, or to a particular style of composition [McFarlane, 2001].

Spatial diffusion was used in the very earliest performances of *Musique Concrète.* Pierre Schaeffer had introduced the idea of the movement of sound along sonic trajectories and the creation of spatial relief through the contrast of static spatial locations and dynamic sources which would be controlled manually by a performer using the *potentiomètre d'espace.* Over the next two decades two-channel stereo and magnetic tape were adopted as the medium of choice, and composers, led by Pierre Henry, began to formulate a performance practice based on the diffusion of a stereo source to a large number of loudspeakers using a special mixing desk. Unlike the multichannel approach adopted by Stockhausen and others, this loudspeaker orchestra consists of a diverse range of speakers, chosen for their specific tonal characteristics. This aesthetic focuses on the temporal, spectral and spatial development of sounds rather than the relationship between these parameters, and the diffusion process is therefore used to exaggerate the dynamic, spectral and spatial content of the musical material already present in the work. In early performances, the intervention of an engineer was already required due to the technical limitations of magnetic tape (such as tape hiss and a limited dynamic and spectral range) and in time, these technical considerations would come to include the diffusion process. The first formalized system based on this approach was the Acousmonium at the *Groupe de Recherches Musicales* (GRM), the collective founded by Schaeffer and Henry in 1958. The Acousmonium was developed by the composer and technician François Bayle, who took charge of the GRM in 1966, and the engineer Jean-Claude Lallemand [Gayou, 2007] and was conceived as a continuation of Jacque Pullin's work on the *potentiomètre d'espace*. The first concert was held in Paris in 1974 and featured a performance of Bayle's *Experience Acoustique*.

Fig. 8.14 An early concert with the Acousmonium

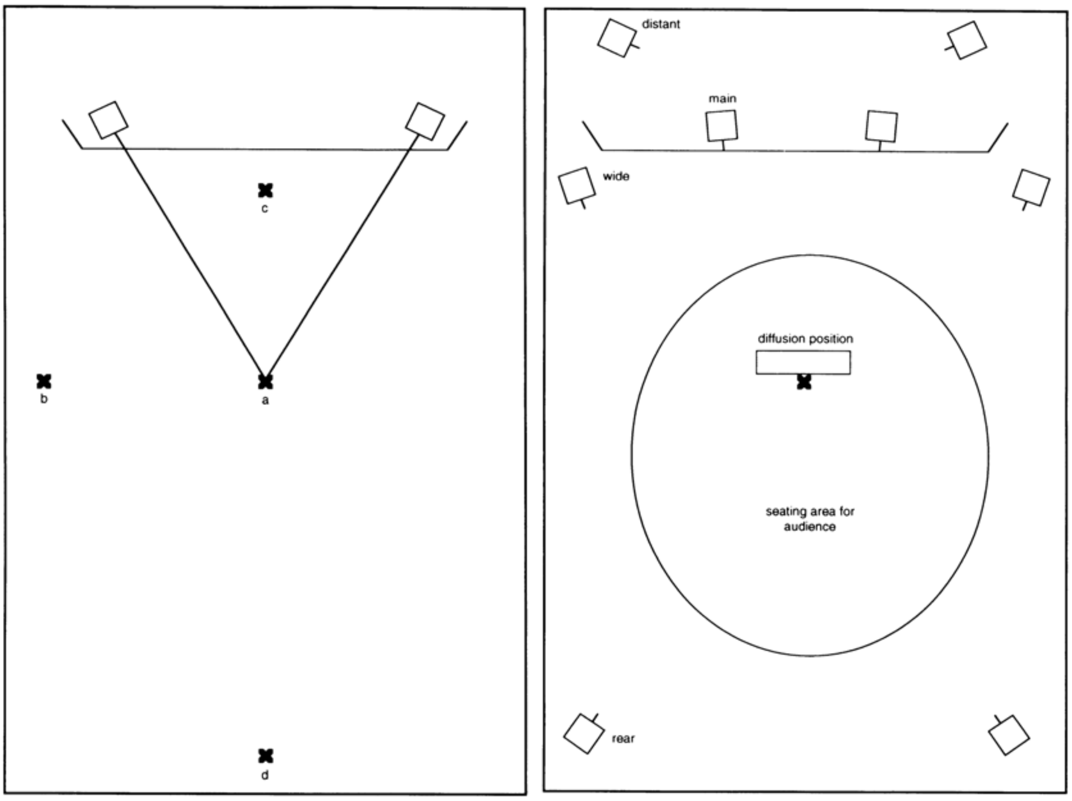

Fig. 8.15 Standard stereo setup (left), diffusion setup (right)

The approach developed by Henry and Bayle was adopted by other composers and institutions such as the *Groupe de Musique de Bourges* in Belgium. The Birmingham ElectroAcoustic Sound Theatre (BEAST) at the University of Birmingham is another important centre for research in this area. Although these groups differ in terms of the precise technical setup and loudspeaker layout, they use many similar working methods and techniques. The source material is generally a stereo CD, and is often the commercial CD release of the piece. The stereo track is routed to a special mixing desk which allows the diffusion engineer to control the routing of the stereo track to different loudspeaker pairs. Often this will be a commercial desk, reverse-engineered so that it takes a stereo signal as the input, and each individual fader channel feeds a different loudspeaker pair.

The diffusion process is as much concerned with adapting the material for the particular performance space as it is with the spatial articulation of the material. For this reason, the loudspeaker pairs are often specifically arranged in an attempt to preserve the stereophonic image for as much of the audience as possible. Jonty Harrison, who works at the University of Birmingham, illustrates this approach using the two layouts shown in Figure 8.15 [Harrison, 1999]. In a normal two-channel stereo system, only listeners seated in the sweet spot (point (a) in Figure 8.15 (left)) will perceive the stereo image correctly. At point (b), the stereo image will collapse to the left speaker due to the precedence effect, while at point (c) listeners will experience a significant hole-in-the-middle effect due to the wide loudspeaker angle at this location close to the loudspeakers. Meanwhile a distant listener at point (d) will perceive a drastically narrowed image. Diffusion systems attempt to overcome these problems through the introduction of additional pairs of loudspeakers and a typical layout is shown in Figure 8.15 (right). In this arrangement, the main stereo pair has been narrowed to reduce the hole-in-middle effect for listeners close to the stage. This is supported by another pair of similar loudspeakers positioned at a wider angle which can be used to increase the image width as necessary. Additional distance effects are supported through the use of a distant loudspeaker pair, positioned at the back of the stage and angled across the stage. Finally, a rear pair is added so that the stereo image can be extended out from the stage and around the audience. This group of loudspeakers is described as the "main eight" in the BEAST system and this layout is described by Harrison as the absolute minimum arrangement required for the playback of stereo tapes [Harrison, 1999]. More loudspeakers are often added to the main eight to further extend the capabilities of the system. For example, additional side-fill loudspeakers are often added to facilitate the creation of smoother movements from front to back. Various elaborate systems have been developed, such as the

BEAST system in Birmingham (shown in Figure 8.16) and the Acousmonium at GRM.

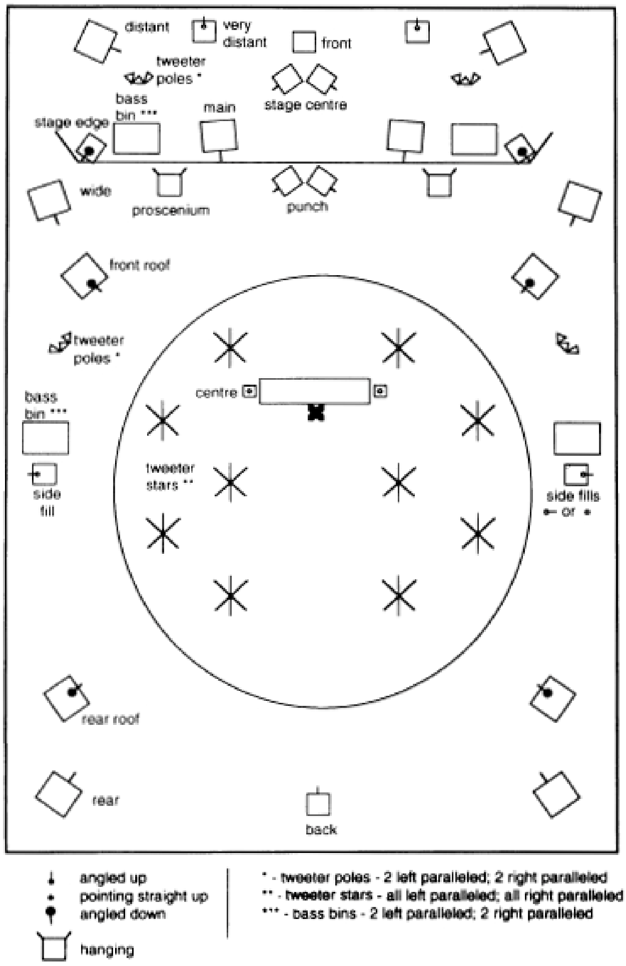

Fig. 8.16 The full BEAST system

Modern diffusion practice also attempts to articulate musical material by performing different passages through different sounding pairs of speaker. Composers do not generally provide a score or notation that indicates how the work should be diffused, instead it is up to the diffusion engineer to interpret the work and “perform” the piece in a way that highlights the musical content and also adapts it to the specific loudspeaker array and acoustics of the particular performance venue. The composer and theorist Denis Smalley describes this approach as follows;

> *“Sound diffusion is the projection and the spreading of sound in an acoustic space for a group of listeners – as opposed to listening in a personal space (living room, office or studio). Another definition would be the sonorizing of the acoustic space and the enhancing of sound-shapes and structures in order to create a rewarding listening experience* [Smalley *et al*, 2000]*”.*

Diffusion has traditionally been controlled manually but some work has been carried out with automated processes and digital spatialization techniques [Truax, 1999; Moore, 1983]. Harrison suggests that this approach reflects the primarily manual processes used in the studio by composers in the early days of *Musique Concrète* [Harrision, 1999]. Others, like Smalley have compared the spatial gestures produced by the diffusion engineer to the physical gestures of traditional instrumental performers [Smalley, 2007]. However, although the diffusion process clearly adds something to the performance, the diffusionist can only emphasize the pre-existing material in the work. Clearly then for a successful performance of a diffusion piece, the composer must organize the material in a way that supports its eventual performance. Denis Smalley's theory of spectromorphology details how a gestural approach can be used as a structuring principle in electroacoustic, and particularly acousmatic music.

8.5 Spectromorphology and the Gestural use of Space

Denis Smalley developed the theory of spectromorphology as a means of describing and analysing the electroacoustic listening experience. Smalley studied with François Bayle at GRM and his compositional aesthetic and theoretical writings developed from the theories of Musique Concrète described by Pierre Schaeffer in *Traite des objects musicaux* [Smalley, 1997]. Schaeffer's approach emphasized *reduced listening,* that is the conscious focus on the intrinsic properties of the sound and the ignoring of extrinsic, referential properties. However, this form of listening is highly unnatural and difficult to maintain, as the natural instinct of any listener is to identify the source and cause of the sound. Smalley uses the term source-bonding to describe this natural tendency to relate sounds to supposed sources and causes, or to relate sounds to each other due to a shared origin. In instrumental music, this process is determined by the physical interaction of the performer with the instrument as the spectromorphological profile of the sound indicates how the instrument is excited by the performer. Smalley suggests that in the case of acousmatic music, both the source and cause of the perceived sound may be detached from a directly experienced or known physical gesture, a concept he refers to as gestural surrogacy. First order surrogacy refers to sounds produced by human gestures such as traditional instrumental music and singing. Second order surrogacy retains some aspects of human gesture but the resulting spectromorphology is not identifiable as a known musical instrument. Remote surrogacy applies when the sound is not related to human gestural activity or any other known source, such as unusual synthesized sounds. Smalley suggests that the way electroacoustic music uses technology to develop musical gesture beyond the note-based instrumental paradigm is one of its greatest achievements [Smalley, 2007]. He describes gesture as an energy-motion trajectory which can be shaped by the composer through the manipulation of the attack and decay envelope of the sound object. The movement of the sound in space should therefore support and emphasize the inherent spectromorphological profile of the sound object. Smalley suggests that weak gestures which are stretched out in time or evolve slowly are detached from human physicality and are instead perceived in a more environmental sense. This results in a shift in attention away from the forward motion of a distinct gesture to a static environmental texture in which the internal activity of the sound object is the primary focus of attention [Smalley, 2007]. A composition could therefore utilize either a gesture-carried structure which implies a degree of forward motion, triggered by some external impetus, or a texture-carried structure which focuses more on the internal activity of the sound which appears to act without any obvious external stimulus. In this way Smalley suggests that gesture and texture can be used as forming principles in a composition and this process is illustrated in Smalley's acousmatic composition *Empty Vessels* (1997). This work is entirely constructed from environmental recordings made by the composer in his garden. The microphones were placed inside a number of large garden pots so the resulting recordings capture both the internal resonances of the pots and also the external environmental sounds of birds, rain and planes. Additional recordings were also made of just the environmental sounds without the filtering effect of the pots. The work begins with a struck chord of unknown origin which slowly transforms into the recorded environmental sounds. This smooth transformation is achieved through the careful matching of the initial struck chord with the resonant drone of the garden pots present in the original recordings. The descending glissandos of planes flying overhead are also smoothly transformed into abstract processed drones which again are related back to the environmental recordings via the internal resonances of the garden pots, the empty vessels of the title.

Smalley recognizes that the perception of the listener will be influenced not only by the spectromorphological profile created by the composer but also by the acoustic of the listening room. He therefore divides the perceived space into the *composed space* (which contains the spatial cues created by the composer), and the *listening space* (the space in which the composed space is heard), as shown in Figure 8.17. The *composed space* consists of both the internal space of the sounding object, such as the resonances of a struck enclosure, and the external space of the environment containing the sound object, which is made apparent through reflections and reverberation. The *listening space* is also subdivided into the personal space, which relates to the precise position of the listener within the venue, and the diffused space created by the various loudspeakers distributed around the venue. Smalley suggests that, in his experience, the perception of a number of parameters will be different when the work is transferred from a single listener in a studio to a large performance space [Smalley, 1997]. Spatial depth which is captured using stereo recording techniques or synthesized using artificial reverberation can easily create images which appear to originate from outside the array for a single listener. However, in a larger space these images may instead become superimposed within the space, rather than beyond the physical loudspeaker positions. Similarly, creating a sense of spatial intimacy [Smalley *et al*, 2000] becomes much more difficult as the size of the listening area increases. The spatial variants suggested by Smalley are shown in Figure 8.17 [Smalley, 1997].

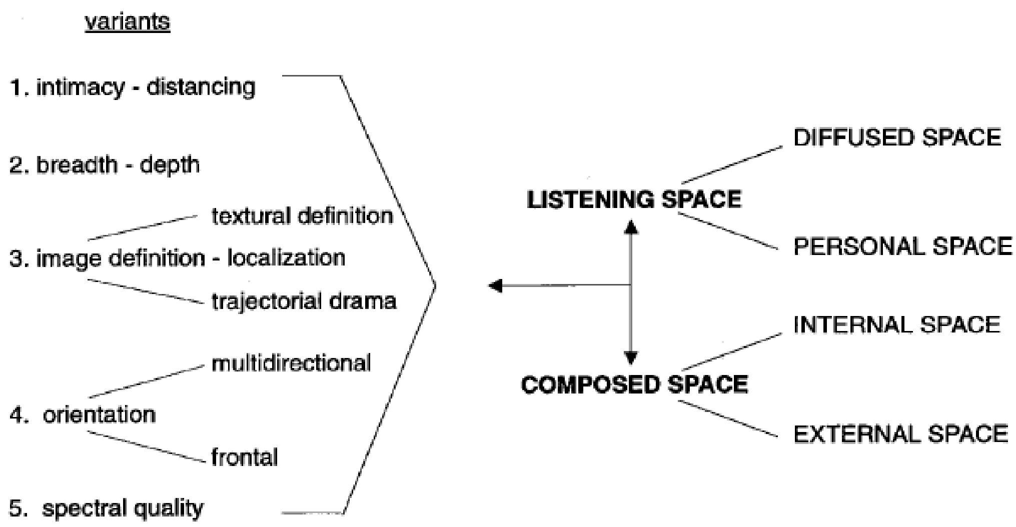

Fig. 8.17 Denis Smalley's perceived space

Although spectromorphology clearly originated from the acousmatic tradition and stereo diffusion, this approach can be applied to other compositional aesthetics and technical formats. Spectromorphology can be used to relate and structure a wide variety of sounds as either gesture or texture, performance or environmental. Human activity, whether in the form of traditional instrumental performance or the manipulation of non-traditional instruments (such as the garden pots in *Empty Vessels),* can be related to synthesized sounds via the shaping of dynamic spatial and spectromorphological gestures. In addition, environmental recordings can be related to continuous synthesized or processed sounds without accented attack and decay envelopes. Soundscape composition [Truax, 2008] is a compositional aesthetic which utilizes the latter approach and is predominantly based upon environmental sounds and textures. Luc Ferrari's *Presque Rien No. 1* (1970), which consists solely of layered recordings from a day at the beach with a minimum level of manipulation of the material, is a well known example of this style of composition. *The Vancouver Soundscape* (1970) by the World Soundscape Project similarly consists of direct reproductions of natural soundscapes with a minimum level of human intervention. Since the seventies, soundscape composition has developed beyond this minimalist approach to include digital synthesis and multichannel techniques and the Canadian composer Barry Truax is one of the chief exponents of this aesthetic. His early work used granular synthesis and a quadraphonic speaker system to create highly textural works with a strong environmental character. In later works such as *Pacific* (1990), or *Basilica* (1992), Truax used granulated environmental recordings with an octophonic, eight-channel array and multiple decorrelated granular streams to create an immersive sonic environment. Truax has argued that the avoiding the representational meaning of environmental sounds is difficult, stating that "environmental sound acquires its meaning both in terms of its own properties and in terms of its relation to context" [Truax, 1996]. Despite this difficulty, environmental recordings have also been used in a more symbolic fashion in which the different recorded sounds and spaces are used to tell a sort of narrative. Various composers and theorists have explored this compositional aesthetic and suggested various symbolic interpretations of different spaces and movements [Trochimczyk, 2001; Wishart, 1985].

The gestural use of space suggested by Denis Smalley in his theory of spectromorphology originated in the acousmatic tradition of stereo diffusion, but this idea is equally applicable to compositions for multi-channel loudspeaker arrays, or mixed-media electroacoustic works. The notion of gestural shaping also suggests an obvious approach to works which combine acoustic and spatialized electronic sounds as the spectromorphological profile of the synthetic sounds can be deliberately designed to match or mimic the instrumental gestures of the performers.

9 Electroacoustic Spatial Music

The addition of live performers introduces a number of difficulties in the performance of spatial music. The spatial distribution of performers around the audience can be very effective but is logistically challenging and highly dependent on the specific layout of the performance space. Compositions for live performers on stage and spatialized audio are particularly challenging as the static location of the live performers provides a frontally-biased visual and audible focus which can conflict with the non-visual and spatialized audio. Linking the musical gestures of the instrumental performer with the spatial gestures of the electronic part is a significant challenge, as the spatialization process is often not related to the musical instrument in any obvious way. In addition it is rarely practical or possible for a single performer to concurrently play and diffuse a musical instrument

In general, electroacoustic music of this sort has taken one of two different approaches. The first and most common approach uses traditional instruments in conjunction with a previously prepared electronic track and this mixed-media approach was really the only technologically feasible solution prior to the development of digital computing technology. As digital processing power increased, the processing and spatialization of instruments in real-time became feasible and by the end of the 1980s various composers had completed major works for live instruments and real-time electronic processing.

The alternative approach, which will be looked at later in this chapter, attempts to incorporate the spatialization process into the design of entirely new musical instruments. This approach has a number of potential benefits but is highly dependent on specific, idiosyncratic devices.

9.1 Fixed Media Electroacoustic Spatial Music

Stockhausen's *Kontakte* (1958-60) is one of the earliest and most successful electroacoustic compositions to include live instrumental performers and spatialized electronic sounds, although this piece can be performed either as a tape piece or as a mixed-media work. Stockhausen overcomes the conflict between the static performers and the spatially dynamic electronic part in various ways. Firstly, the scored instrumental material is tightly integrated with the synthesized electronic sounds within the overall serialist structure, as discussed in detail in Chapter Eight. In addition, various points of contact between the static instrumental sounds and dynamic electronic textures are created through crescendos in one part which provoke a response from the other. Harrison suggests that the unusual loudspeaker arrangement also contributes to the success of this piece. As the four loudspeakers are placed in a cruciform arrangement rather than in the corners of the space, this means that three of the four loudspeakers and the live instruments (piano and percussion) are in front or to the sides of listeners with only one loudspeaker behind [Harrison, 1999]. Therefore, the instrumental sounds and most of the electronic parts are produced largely from in front of the audience, this helps to unify the two parts into an integrated whole.

Although *Kontakte* successfully integrates the static instrumental performers and the dynamic spatialized electronic parts within the work as a whole, in much of the piece the static spatial locations of the performers is actually deliberately contrasted with the dynamic spatial trajectories in the electronic part. The various points of contact between the two spaces is then heightened for dramatic effect in much the same way as other elements in this piece such as pitch and rhythm. The composer himself wrote that in works which combine electronic and instrumental music, there remains the problem of "finding the larger laws of connection beyond that of contrast, which represents the most primitive kind of form [Stockhausen, 1975a]".

In electroacoustic works such as *Kontakte*, the performers must perform while listening to a click track in order to synchronize with the tape part. However, some composers were uncomfortable with the rigidity of this arrangement, and the French composer Pierre Boulez was a particularly vocal critic, stating that “give and take with the tempo of a piece is one of the basic features of music”. It is perhaps unsurprising that Boulez was so attuned to this particular limitation considering his career as a conductor and he would go on to compose for one of earliest systems for the real-time processing and spatialization of sounds, an approach which would “capture all the spontaneity of public performance, along with its human imperfections [Boulez *et al*, 1988]”.

9.2 Real-Time Electroacoustic Music

The French composer Pierre Boulez was, along with Karlheinz Stockhausen, one of the chief exponents of serialism and modernist contemporary music. Like Stockhausen, Boulez was profoundly influenced by the twelve-tone compositions of Anton Webern and the total serialism of Olivier Messiaen. Boulez did not at first share the German's interest in electronic or spatial music, preferring instead to compose acoustic works in a totally serialized, punctual style. By the late fifties and early sixties however, Boulez's compositional approach had become less strict and his music began to incorporate electronic sounds and spatial effects. He was particularly interested in the combination of electronic and acoustic sounds, the "fertile land" from his 1955 article on the future of electronic music, "*À la limite du pays fertile (At the edge of the fertile land*)" [Häusler *et al*, 1985]. His first electroacoustic work, *Poésie pour Pouvoir* (1958), was scored for three orchestras and five-track tape with an electronic part constructed from recorded and processed instrumental sounds. This arrangement was perhaps influenced by Stockhausen's *Gruppen*, as Boulez was one of the three conductors at the premiere of this piece. As with many of his compositions, *Poésie pour Pouvoir* was later withdrawn by Boulez due to his dissatisfaction with the electronic part. Boulez was particularly unhappy with the limitations imposed by the click track, which was required to synchronise the human performers with the tape. In an interview in 1985, Boulez described this issue as follows;

> *"I have conducted pieces of other composers who use tape and have learned that one is at a disadvantage because one must constantly take the synchronization into account, leaving no room for the coincidences of interpretation. I don't mean this only in the sense that one can select structures at will, but that the gesture of interpretation — as I would like to call it — is completely paralyzed, because one must pay attention to too many things that have little to do with the actual performance. The time of a tape recording is not psychological time but rather chronological time; by comparison, the time of a performer — a conductor or an instrumentalist — is psychological, and it is really almost impossible to interconnect the two. "* [Häusler *et al*, 1985]

Faced with these difficulties, Boulez' abandoned his work with electronic sounds and it would be over twenty years before Boulez would return to this area. His opportunity to do so came in 1970 when he was asked by the French president Georges Pompidou to create and head a new institute for musical research and composition, the *Institut de Recherche et Coordination Acoustique/Musique*

(IRCAM). The institute opened in Paris in 1977 and, along with the Centre for Computer Research in Music and Acoustics (CCRMA) at Stanford, IRCAM quickly became one of the most important centres for research in digital music technology.

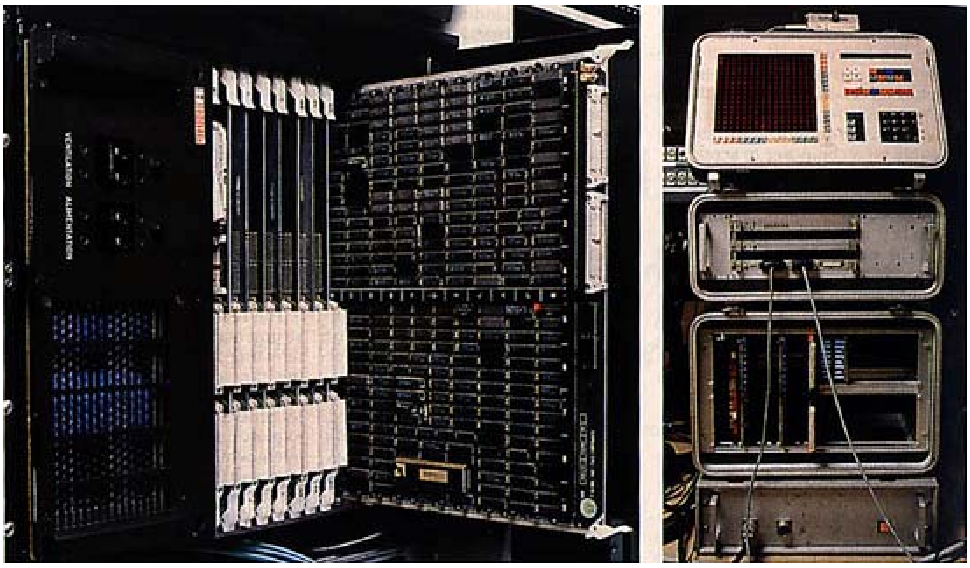

Fig. 9.1 IRCAM's 4X (left) and Matrix 32 (right) digital processors

By 1980 IRCAM had developed the first prototype of a real-time digital system which was expressly designed for the live performance and spatialization of electroacoustic music. The final version of this system was completed in 1984 and consisted of the 4X processor (Figure 9.1 left) and the Matrix 32 (Figure 9.1 right). The 4X was a real-time digital signal processor for the analysis, transformation and synthesis of sounds in real-time which was designed at IRCAM by Giuseppe Di Giugno and Michel Antin [Boulez *et al*, 1988]. The 4X could store, manipulate and recall up to four seconds of digital audio using various software modules, or patches, and a real-time operating system and event scheduler (developed by Miller Puckette, Michel Fingerhut and Robert Rowe). The Matrix 32 was designed by Michel Starkier and Didier Roncin and functioned as a programmable and flexible audio-signal traffic controller for the dynamic routing of audio signals between the instrument microphones, the 4X and the loudspeakers. Boulez would go on to use both devices in the electroacoustic composition *Répons*, a flagship project for IRCAM, which was premiered as a work-in-progress in October 1982 at the Donaueschingen Festival in Germany [Häusler *et al*, 1985]. *Répons* is scored for a 24-piece orchestra, six solo instruments (harp, glockenspiel, vibraphone, cimbalom and two pianos), six loudspeakers and IRCAM's real-time digital signal processing system. Boulez was assisted by Andrew Gerzso in the design of software patches for the 4X processor and several technicians were required during the performance to manage the various devices. Boulez positioned the instruments and loudspeakers in *Répons* using a very similar layout to that of the earlier work, *Poésie pour Pouvoir* (see Figure 9.2). According to the composer, the un-amplified and unprocessed orchestra is situated in the centre of the auditorium in order to provide a clear visual focus for the audience [Boulez *et al*, 1988]. The audience is therefore seated in-the-round, surrounding the orchestra, and within the circle formed by the six soloists and the loudspeaker array. The title of this piece comes from a medieval French term for the antiphonal choral music discussed earlier in this thesis. This call-and-response process is implemented throughout *Répons* and the dialogue between the soloists and the main orchestra is particularly reminiscent of medieval choral antiphony. Boulez relates this process of multiple responses (the orchestra) to a single question (the soloist) to the notion of the multiplication and the proliferation of sounds. This idea is also implemented electronically as various processes multiply single instrumental notes into a multitude of notes, chords or timbres, all of which are related to the original in a clear way.

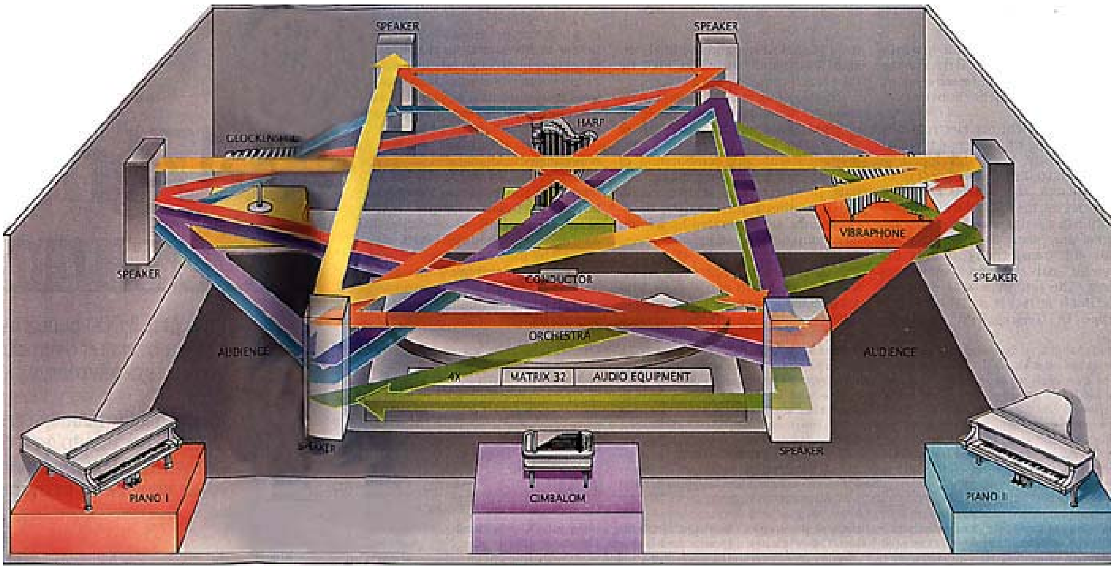

Fig. 9.2 Layout and spatialization diagram for Répons

After an opening movement performed solely by the central orchestra, the six soloists enter dramatically with different but concurrent, short arpeggios. These instrumental arpeggios are captured by the 4X system and circulated around the loudspeaker array in the trajectories shown as coloured arrows in Figure 9.2. The composer describes the effect of this passage as follows;

"The attention of the audience is thereby suddenly turned away from the centre of the hall to the perimeter, where the soloists and the speakers are. The audience hears the soloist's sounds travelling around the hall without being able to distinguish the paths followed by the individual sounds. The overall effect highlights the antiphonal relation between the central group and the soloists by making the audience aware of the spatial dimensions that separate the ensemble from the soloists and that separate the individual soloists as well. " [Boulez *et al*, 1988]

The various spatial trajectories are therefore not intended to be perceived directly but are instead used to distinguish the different electronic gestures associated with each of the six soloists. This effect is further exaggerated through the use of different velocities for each spatial trajectory, which are determined by the amplitude envelope of the original instrumental gesture. Each instrument has a similar amplitude envelope (see Figure 9.3) consisting of a sharp attach and an slowly decreasing decay, however, the duration of the decay depends both on the pitch of the note and also the instrument on which it is played [Boulez *et al*, 1988]. Therefore, the velocity of each trajectory decreases proportionally to the decay of the associated instrumental passage, linking the instrumental gesture of the soloist to the spatial gesture of the electronic part.

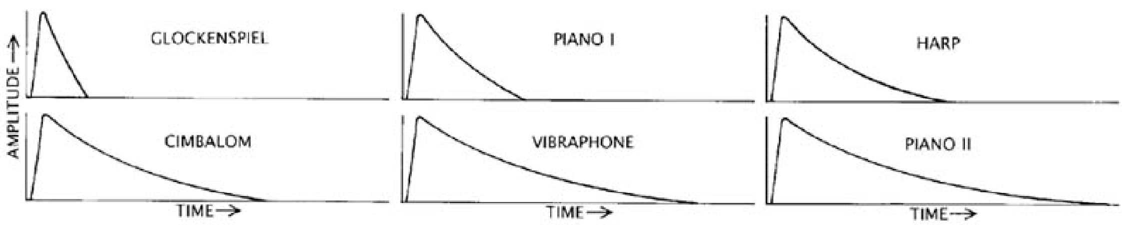

Fig. 9.3 Amplitude envelopes for the six soloists in <u>Répons</u>

The dynamic spatialization is implemented using an unusual switching mechanism rather than a panning scheme such as stereophony or Ambisonics, perhaps to ensure that the system responds immediately to the live input without any delay, or latency. The 4X system controls the switching of the input signal from loudspeaker to loudspeaker with a series of flip-flop modules. These modules are controlled via a timing signal whose frequency changes in proportion to changes in the amplitude of the sound's waveform envelope, which is continuously generated by an envelope follower module [Boulez *et al*, 1988]. Therefore, as the input signal decays, each flip-flop module holds the signal at each speaker for longer before switching to the next loudspeaker, slowing the spatial trajectory. This process illustrated in Figure 9.4 for a system with four loudspeakers, and hence four flip-flop units.

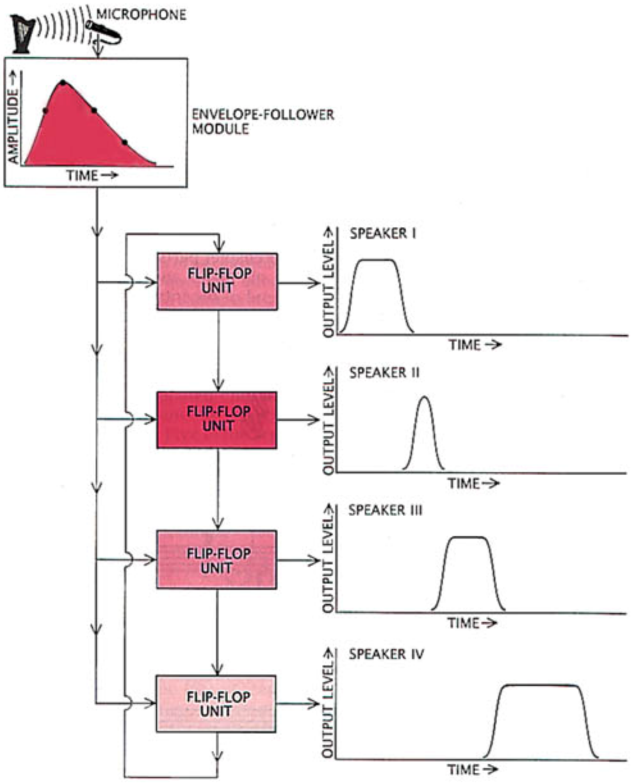

Fig. 9.4 The spatial switching mechanism used in *Répons*

The themes of multiplication and displacement are further developed through the recording, processing and playback by the 4X processor of arpeggiated chords performed by each soloist. Boulez developed the idea of an arpeggio as a displacement in time and pitch which may be applied, not just to the notes of a chord, but also to the spreading of the electronic material in pitch, time and space. Boulez describes this process as an arpeggio (the multiple recorded and processed copies created by the 4X) of an arpeggio (the sequentially played notes of the instrumental chord) of an arpeggio (the individually cued soloists) [Boulez *et al*, 1988]. The speaker switching mechanism could also be interpreted as a form of spatial arpeggio and this is perhaps another reason why this switching mechanism was used instead of a standard panning algorithm.

As with many large-scale works of spatial music, *Répons* has been performed relatively infrequently. As always, finding a venue which can accommodate the various distributed musicians and loudspeakers is difficult, and the dependence of this piece on highly specific hardware is an additional limitation in this case. IRCAM's real-time processor was, however, highly prescient in its combination of a general purpose digital processor and customised software to implement the required processing. As computing power became more readily available and affordable, IRCAM moved away from hardware devices such as the 4X toward the development of software for general purpose personal computers. The real-time scheduler and operating system developed for the 4X was later developed by Miller Puckette into the graphical programming language Max/MSP which is now used extensively in electroacoustic performances. Various other software packages have also been developed by IRCAM for audio analysis and synthesis including the spatialization package Spat, which was discussed previously in Chapter Three.

9.3 Real-Time Spatialization

Although modern personal computers are much more powerful than IRCAM's 4X processor, many of the same issues still apply. The interaction between the instrumental performer and the electronic part is crucial in any electroacoustic composition, and creating a clearly perceivable relationship between these two mediums remains a significant challenge. While a clear contrast between these two elements can be used as a compositional element, this approach is somewhat limited. *Dialogue de l'ombre double* (1985) was completed by Boulez shortly after *Répons* and this work for solo clarinet and pre-recorded tape illustrates how the contrast between these two mediums can be successfully used as a compositional device. The "dialogue" of the title is between the live clarinettist and the pre-recorded and electronically processed clarinet on tape, and Boulez dramatically heightens the contrast between these two elements using lighting and spatial effects. During instrumental passages, the un-amplified clarinettist is positioned in the middle of the hall and audience while the electronic responses on tape are produced using six loudspeakers positioned around the audience. Lighting is used to further heighten this contrast as during instrumental passages the clarinettist is brightly lit with a spotlight which is turned off for the electronic response of the tape part.

The composer Denis Smalley, when asked how his approach to spatial diffusion changes when additional human performers are added gave this response.

"The same factors are present, but they are considered differently because the focus of a live performance visually and musically - is the performer. So, I don't want to use as full a diffusion system as I do for tape pieces, because overdoing the diffusion will tend to undermine the carefully considered musical relationships between the live performer and the content of the acousmatic domain. - If you have more than one performer, there are different considerations, because the visual and sonic weight in front, on stage, is increased. Therefore, I think that there is more for the eye to focus on and follow, and the acousmatic possibiities become reduced. For example, one can't have a lot of visual silences on stage, where people are sitting doing nothing. "
[Smalley *et al*, 2000]

Smalley suggests that this problem is apparent in performances of Edgar Varèse's *Deserts* (1950-54) for wind, percussion and electronic tape [Smalley]. In two sections of this work, the musicians on stage sit quietly, doing nothing and the electronic tape part dominates. Smalley suggests that this creates a visual silence on stage that contrasts uncomfortably with the required acousmatic perception of the tape, as the human performers provide an unavoidable visual focus for the audience which conflicts uncomfortably with the acousmatic electronic part. The extent to which this is a problem in a particular work is however somewhat subjective, and other composers such as Natasha Barrett have suggested that this represents a broader dichotomy between the sonic demands of the composition, and its performance.

"Barrett: *Sometimes. For example, unless I am composing a purely theatrical work, my initial agenda is sound, whereas the performers' primary agenda is performance, and they are not necessarily the same thing. This is, of course, not always the case, but either way we need to integrate the demands of both performance and sound. For example, in Symbiosis, there is—toward the end—a stretch of three minutes for solo computer.*
Otondo: *There are also parts where it is almost purely instrumental.*
Barrett: *Exactly, and in Symbiosis, this is one way in which I balance the elements: to give both parts solos and to have both coming together at important meeting points. In those three minutes of solo computer, the performer has to find a way to "compose" herself. Everybody is looking at her, and she is a performer having to listen. In that section the theater emerges as she demonstrates the act of listening and not just sitting there waiting for the next thing to happen. You see a string quartet play a piece with computer or tape, particularly when they play with a click track, and they look like they are thinking, "When do I come in next? What time is it?*"
[Barrett *et al*, 2007]

In the opinion of the author, the contrast between the visible and static performer, and acousmatic, spatially dynamic sounds can be problematic, but only if this issue is not addressed by the composer within the overall aesthetic. In addition, this is much less of a problem when the performers are distributed around the audience, or when smaller numbers of musicians are involved.

As stated earlier, perhaps the biggest issue is forging some connection between these two elements beyond that of contrast. Smalley suggests that the invisible acousmatic content in a mixed-media work must share gestural or behavioural characteristics with the instrumental part if both elements are to be regarded as equally significant and since the acousmatic component is competing with the visible actions of the performer, it must have a strongly articulated presence [Smalley, 1996]. While this is certainly true, the amplification of the instrumental parts must be considered as an equally significant issue. In smaller venues, the acoustic instruments require little or no amplification, and in this case the positioning of loudspeakers close to the performers with a small amount of amplification of the instruments, can be sufficient to forge a connection with the amplified, electronic sounds. In larger venues, the greater issue is often the disconnect between the perceived position of the amplified, instrumental sound (which often collapses to the nearest PA loudspeaker), and the visual position of the performer on stage. If this issue is not addressed (perhaps through the positioning of performers close to the loudspeakers which are being used to amplify their particular instruments), then a connection between the acousmatic and instrumental sound may be achieved but a connection to the visual performer may not.

10 Spatial Music Composition

In this Chapter, a number of original works of acoustic, electronic and mixed-media spatial music will be discussed and analyzed. These compositions are presented chronologically, beginning with *Discordianism*, the piece which as discussed previously, provided much of the impetus for this thesis.

10.1 Case Study I: Discordianism

Discordianism is the earliest piece of spatial music written by the author which is included in this thesis and as we shall see, the use of amplitude panning and a quadraphonic array proved to be highly problematic and a relatively inexperienced choice. This quadraphonic tape piece was not composed with Denis Smalley's theory of spectromorphology in mind, however, this theory is a useful way of classifying different types of material and is referenced extensively in the following discussion.

The primary source material consists of freely improvised material produced using an electric guitar and a wide assortment of effects pedals. Various traditional instrumental gestures such as plucked harmonics were extracted from the original guitar recording (the first order surrogacy material using Smalley's terminology). A sample of a single note on a prepared piano which is introduced later on in the piece also provides a similar degree of first-order surrogacy. This straightforward instrumental material is slightly abstracted using various processes such as multi-channel granulation. This material, although abstracted, still retains aspects of the original instrumental gesture and so would be classified as second order surrogacy. However, much of the original recording is also second-order as this material was not produced directly with the guitar but instead by manually controlling the effects pedals to produce bursts of feedback and other sustained tones. For example, a feedback pedal (which essentially loops the output of an effects pedal(s) back into its input) allowed some highly unusual effect such as controlling the pitch of feedback using the tone control on the guitar. The resulting sounds therefore contain a degree of human gestural activity, but the spectromorphological profile is not particularly identifiable as a musical instrument. The granulated material fulfils a similar function as the temporal reordering and pitch shifting produce bubbling, irregular textures which still retain a degree of human gesture. Finally, granular time-stretching and resonant filters are used to create a variety of sustained drone including a heavily distorted and sustained drone taken untreated from the original performance. The various drones represent the most abstract material, and hence remote surrogacy.

10.1.1 Analysis

The title of this piece is a reference to the modern, chaos-based religion founded in either 1958 or 1959 which has been described as both an elaborate joke disguised as a religion, and as a religion disguised as an elaborate joke. The number five is highly important in this parody religion and *Discordianism* is structured accordingly. It is in 5/4 time, it is composed of five, twenty-five bar sections, and it is approximately five minutes long. It also uses elements of chaos theory in its construction, specifically in the granular synthesis algorithm which incorporates the classic population-growth model, and this chaotic process, combined with the "*preciso e meccanico*" rhythmical activity is intended to reflect the irreverent philosophy referred to in the title.

The opening section of this work is marked by repeated sequences (in multiples of five) of a plucked guitar harmonic which cycle around the four channels and builds in dynamic to a sudden decrescendo. This first-order surrogacy gesture interacts with the other material in various ways, sometimes triggering a short burst of feedback, sometimes effecting the activity or density of the bubbling, spatial granulated texture operating in the background, or sometimes provoking simply silence. Gradually, the density and activity of the granular texture increases until at the one minute mark, a burst of feedback triggers a heavy, distorted feedback drone in both front channels. The sustained nature of this drone extends the second-order surrogacy even further, although occasional human gestural activity from the original performance remains. The short feedback bursts now function in a structural sense, marking out five bar intervals and triggering the introduction of new material such rapid granulated tremoloed notes and waves of white noise produced by resonant filters. These additional drones slowly drift around the four channels, in contrast to the static distorted drone locked to the front channels. The dynamic and density of the material steadily increases before a brief decrescendo which is immediately followed by a sudden, highly dynamic reintroduction of all the material, triggered again by a short burst of feedback.

This highly dramatic point signals the start of the third and final section of this work and the gradual introduction of four layers of percussive material proceeding at different tempos from the four corners the array. It also forcefully reintroduces both first-order and remote surrogacy and finally confirms the connection between them in the form of spatially distributed layers of clearly instrumental material, operating at different tempi, and a granulated version of the distorted drone from the second section. This new, heavily distorted drone now moves dynamically in space with dynamic and irregular lateral movements created by the stereo spreading of the granulation process and a rapid back and forth movement between the front and rear channels using amplitude panning. The final section steadily increases in dynamic and intensity as more layers of rhythmic and granulated material are added building to a sudden crescendo and rapid cut-off before a brief, swinging spatial movement of the distorted drone which is finally, cut off by one last burst of feedback.

10.1.2 Critical Evaluation

This early work was written for a quadraphonic tape concert, and then later remixed to two-channel stereo for a CD release. As such, it provides an excellent opportunity to assess various issues related to the performance and indeed the function of space in music.

During performances of this piece, the opening section was relatively robust in terms of the spatialization. However, this is perhaps to be expected as in this section each channel is primarily just used as a single point source. The sustained drones in the second section also seem to function relatively well, although the degree of envelopment was certainly reduced in comparison to other eight channel works which were performed at the same concert. The amplitude panned drone which is highly prominent in the third and final section appeared to be the most affected by the change in listening environment. The spatial movement of this material was relatively clear in a studio environment but the increased reverberation, greater inter-loudspeaker distance, and distributed listening positions significantly distorted the perceived trajectory to the point where no clear direction of front-back movement could be distinguished in the performance venue. These, admittedly highly subjective experiences, would seem to agree with the results of more formal evaluations discussed earlier in this thesis.

In this piece, spatialization is primarily used to clarify different rhythmical material and to a less extent, to articulate the various gestures. As noted earlier, this work was composed during the very early stages of this thesis and the lack of an underlying rationale as to the approach to spatial movement is perhaps all too clear. Clear gestural relationships between the repeated plucked harmonic and granulated drone textures are created, however these relationships could have been further heightened through the dynamic movement of the drone material in response to the triggering plucked harmonic. The rhythmical approach adopted here was influenced more by György Ligeti than Henry Brant, yet the use of spatial distribution to clarify and distinguish different layers of electronic drones and rhythmic material is highly effective.

The two channel stereo reduction also provides an interesting point of comparison as to the benefit of the spatialization process. The stereo version, which was necessary for a CD release, was professionally mastered and its overall dynamic range is therefore reduced in comparison to the four channel mix used for performances. Upon listening to the four channel mix after the stereo version, the first obvious difference is that the different layers of material are much easier to distinguish when spatially distributed among a greater number of channels. The degree of intelligibility in the stereo mix could perhaps be increased using equalization to carve out individual spectral space for each layer. Yet much of the interest in this piece rests in the similarities between the spectromorphological profiles of the different surrogate layers and this would presumably be affected by the spectral adjustments necessary. Most of all, the dramatic and expressive benefits of spatialization are clearly demonstrated by the dynamically moving, heavily distorted drone in the final section which is significantly static in the two channel reduction.

10.2 Case Study II: Sea Swell (for four SATB choirs)

The initial inspiration for this work came from some experiments carried out by the author with granulated white noise. When a very large grain duration was employed the granulation algorithm produced a rolling noise texture which rose and fell in a somewhat irregular fashion and was highly reminiscent of the sound of breaking waves. This noise texture was originally intended to accompany the choir, however in practice this proved to be unnecessary as the choir could easily produce hissing, noise textures which achieved the same effect.

The choir is divided into four SATB groups positioned symmetrically around the audience and facing a central conductor. The piece utilizes many of Brant's ideas, albeit with a much less atonal harmonic language. In particular, spatial separation is used to clarify and distinguish individual lines within the extremely dense, sixteen part harmony, which is shown in Figure 10.1. If this chord was produced from a single spatial location, it would collapse into a dense and largely static timbre. However, this problem is significantly reduced when the individual voices are spatially distributed. Certain individual voices are further highlighted through offset rhythmic pulses and sustained notes, as shown in the score excerpt in Figure 10.2.

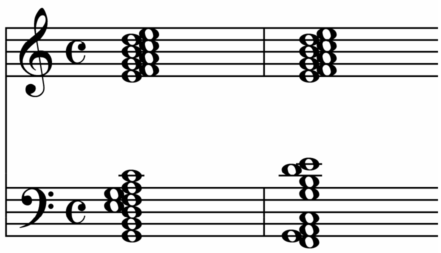

Fig. 10.1 *Sea Swell* – harmonic structure

Exact rhythmic coordination is difficult to achieve when the musicians are spatially separated and hence a slow, regular tempo is used throughout. The beat is further delineated through the use of alliteration in the text as the unavoidable sibilant *s* sounds mark the rhythmic pulse (see Figure 10.2). These sibilant sounds are also used in isolation at the beginning of the piece, to create a wash of noise-like timbres which mimic the sounds of crashing waves suggested by the title. Each singer produces a continuous unpitched "sssssssssss" sound which follows the dynamic pattern shown in Figure 10.3. The angled lines indicate the dynamics which are divided into three levels, below the bar line indicating fading to or from silence, above the bar line indicating maximum loudness and on the bar line indicating a medium volume. The note durations are provided to indicate durations for the dynamic changes and do not indicate pitch. Instead a continuous "ssssss" sound is produced for each entire phrase.

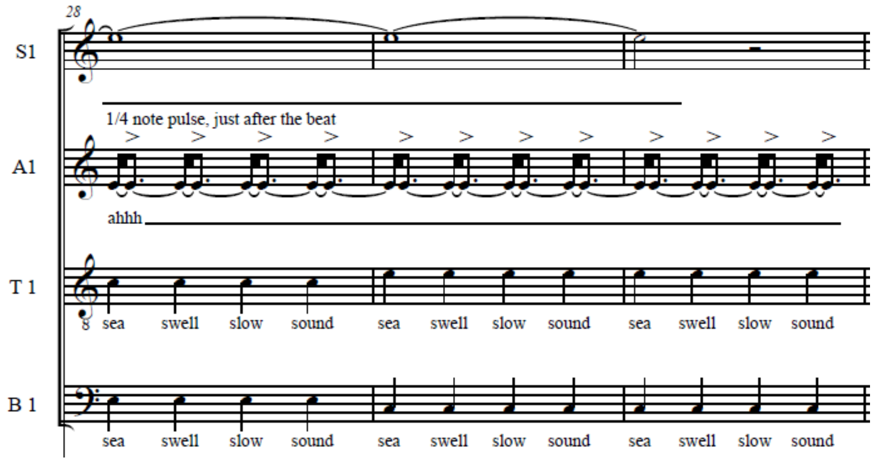

Fig. 10.2 *Sea Swell* – rhythmic alliteration

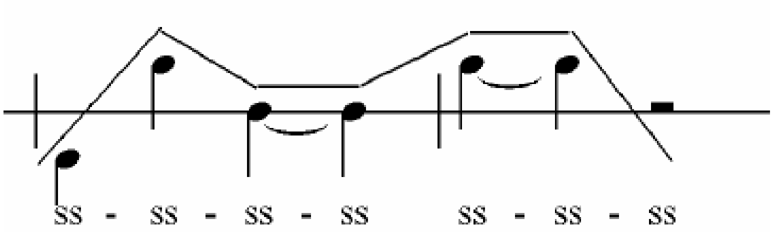

Fig. 10.3 *Sea Swell* – notation example for opening sibilant section

10.2.1 Analysis

Brant used the term *spill* to describe the effect of overlapping material produced by spatially distributed groups of musicians. However, Brant was also aware of the perceptual limitations of this effect and he noted that precise spatial trajectories became harder to determine as the complexity and density of the spatial scene increases. The opening of this piece will contain a significant degree of spill due to the common tonality of each part and will thus be perceived as a very large and diffuse texture. Large numbers of these sibilant phrases are overlapped in an irregular fashion to deliberately disguise the individual voices and create a very large, spatially complex timbre of rolling noise-like waves. This opening section can be seen in full in Figure 10.4.

The loose ABA structure of this work ends with a shortened refrain of the opening section discussed earlier. The B section consists of a lyrical and highly melodic progression of layered cannons. The lyrics consist of some fragments of poetry by Pablo Neruda as well as some original lines by the author. Three different lines are cannoned between the sopranos, altos and tenors in a form of spatial and melodic counterpoint, while the basses reintroduce the sibilant sounds from the introduction.

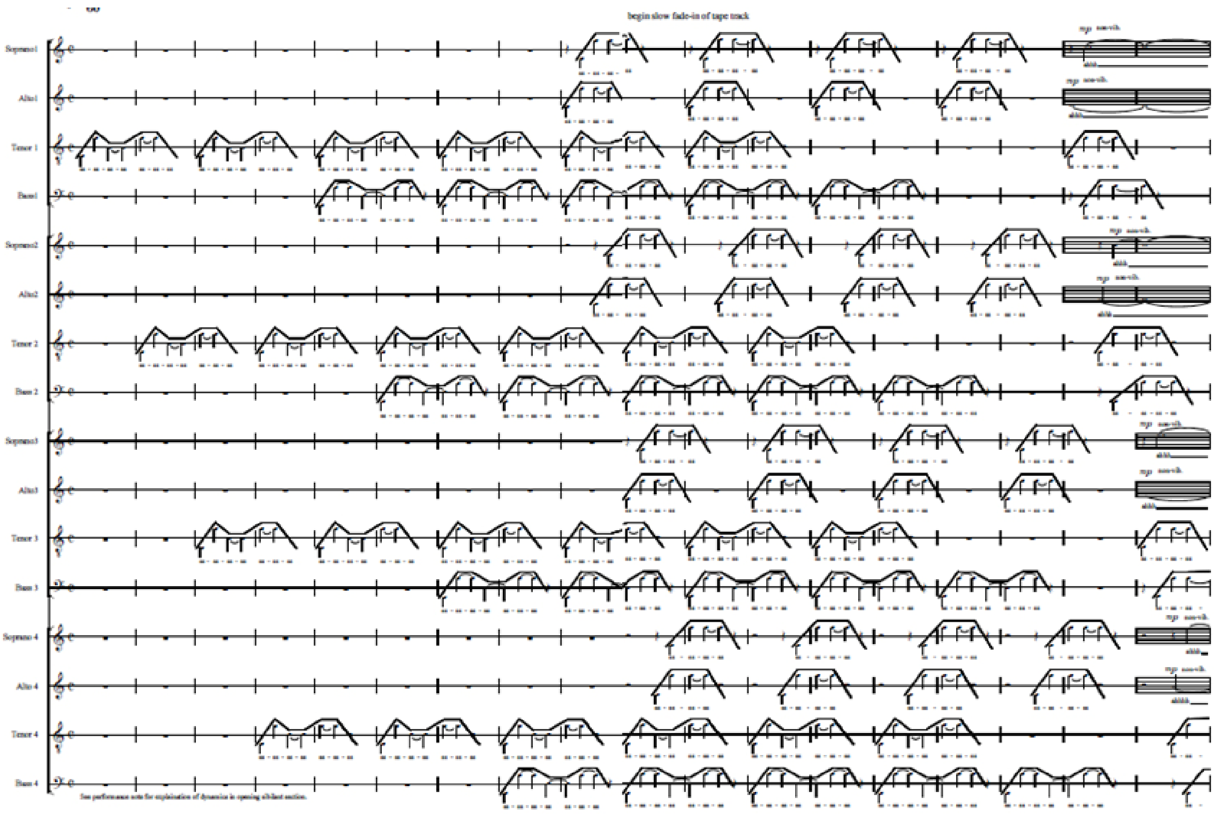

Fig. 10.4 *Sea Swell* – introductory section

10.2.2 Critical Evaluation

The opening *hiss* section of this piece has worked extremely well in practice, particularly once all of the sixteen voices have entered. The significant amount of reverberation and the similar timbre of each part is such that the individual voices largely disappear into a very diffuse and spatially dynamic noise texture which is highly enveloping and quite evocative. The use of alliterative ‘s’ sounds is also quite effective as a rhythmical device as these sibilants are quite pronounced and prominent. While this is highly useful rhythmically, care must be taken when hissing sounds are used in conjunction with normal vocal parts (e.g. bars 45-66) to ensure that the level of the hiss sounds does not dominate. The spatial separation of the dense harmonies and multiple canons is also quite beneficial as this separation really helps to clarify and separate the multiple, layered cannons.

10.3 Case Study III: String Quartet No. I

The first part of this piece to be composed was the granulated and time-stretched recordings of whole-tone trills used in the electronic part. The spatially and spectrally dynamic drone possessed a clear timbral relationship to the acoustic sound of a quartet, and the dynamic spatialization of this part could clearly enhance the existing dynamic of acoustic quartet. However, this scenario left little room to freely compose a string quartet and this approach indeed proved to be exceedingly difficult in practice as it took a period of six months before any progress was made with the instrumental part.

The issues raised by Smalley and discussed in the previous section will clearly be of concern in a work of electroacoustic spatial music written for string quartet. A string quartet will provide a strong, frontally biased visual focus for the audience which can only be avoided if the musicians are distributed around the audience. In this work, the amplified string quartet (using the front loudspeakers, either the house PA or the front loudspeakers of the surround array) remains on stage while the spatialized electronic part is used to extend and widen the acoustic sound of the quartet from the stage and out into the hall. However, due to the unavoidable frontal bias of this layout, the electronic part only rarely moves entirely to the rear, and functions more as a dynamic, spatial and timbral extension of the acoustic sound of the quartet.

10.3.1 Analysis

The primary concept behind this piece is of an endless ascent, inspired by the optical paradox of the Penrose stairs (see Figure 10.9) or its auditory equivalent, the shepard tone. This concept is implemented using an ascending dominant seventh scale on G, which is equivalent to the Mixolydian mode. This mode uses the same series of intervals as the major scale (C major in this case), except the fifth is taken as the tonic. Although this rather straightforward progression is the basis of the entire work, it is generally disguised and distorted in different ways. At the beginning of the piece, the cello begins with the tonic G, the first note in the progression, however, the other three parts respectively start on the fifth, sixth and seventh notes in the series. Similarly, different rhythmic groupings are used for each part and this is particularly evident in the opening ascent which is shown in Figure 10.10. The cello holds each note for twelve crotchets, the viola for ten, the second violin for eight and the first violin for six, and in addition, the higher three parts begin after a single crotchet rest, in order to reduce the number of times more than one part changes note simultaneously and disguise the exact harmonic root of the progression.

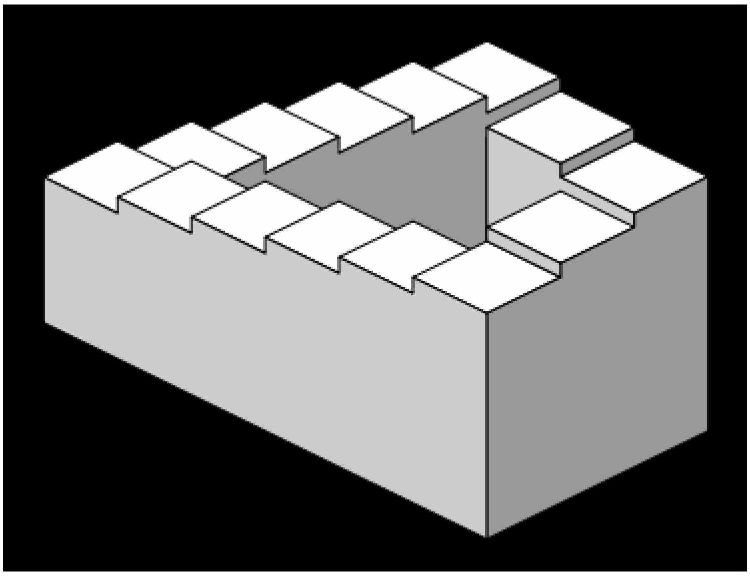

Fig. 10.5 The Penrose stairs

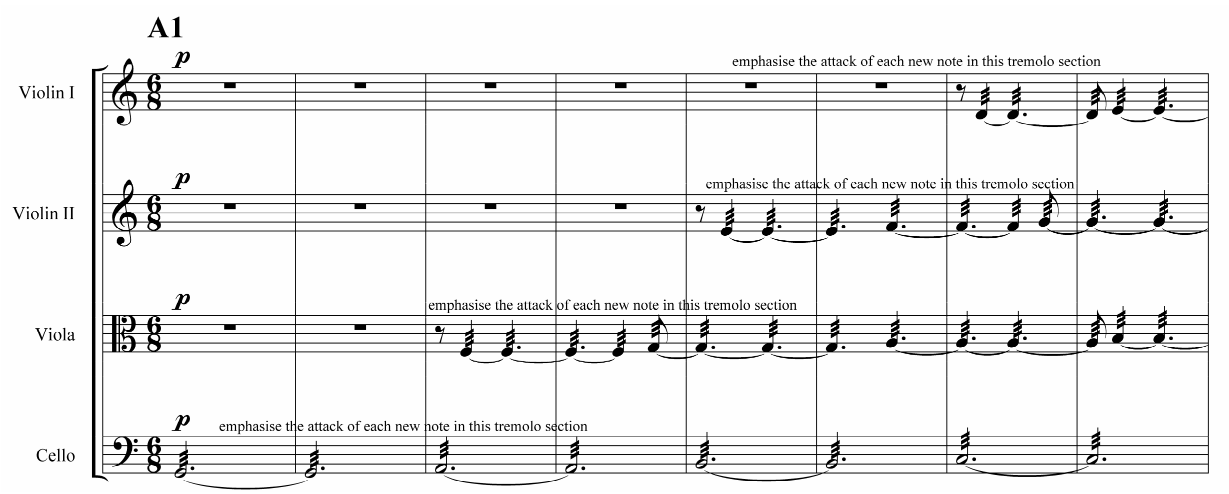

Fig. 10.6 Opening eight bars of *String Quartet No. I*

The electronic part is constructed from granulated and time-stretched recordings of whole-tone trills played on the individual instruments of the quartet. The granulation process transformed each monophonic recording into a wide, stereophonic drone and the time-stretching process lengthened the original rapid trill to a slow and irregular glissando. Composers such as Barry Truax [Truax, 1990] and Curtis Roads [Roads, 2004] have noted that when sounds are granulated using longer grain durations, the original timbre is much more apparent, and this approach was adopted here. As a result, the granulated drone has a similar timbre to the original recordings, and therefore to the acoustic quartet. In addition, the bowing action which is clearly audible in the original recordings is also apparent in the granulated drone. Using Smalley's terminology, the drone retains aspects of the spectromorphological profile of the original instrumental gestures, which will in turn bond with the live instrumental gestures of the quartet. Multiple layers of these drones were finally superimposed to create a very dense drone which nonetheless retains the spectromorphology of the original instrumental gesture.

The large scale structure of this piece is delineated by a series of six crescendos, which is clearly illustrated in Figure 10.11. At first, the electronic part is positioned to the front, with the quartet, and is harmonically restricted to the same dominant seventh scale. As the dynamic increases, the electronic drone becomes increasingly chromatic as more notes, not present in the original scale are introduced, and by the first crescendo, all twelve pitches are present in the electronic part. At this point, the electronics move from the front to a wider, 90^{o} spread. As the quartet begin the following syncopated section (beginning at bar 45) the electronic part mimics the short, stabbing rhythms of the instruments and extends them spatially. The descending progression which ends at the third crescendo is reminiscent of the opening ascent, but now the tone intervals of the original progression are filled, resulting in a more chromatic descent by the quartet which reflects the "corrupting" influence of the electronic part. The spatialization of the electronics matches the descending progression of the quartet as it collapses from a wide spatial distribution to a narrow, frontal position matching the position of the quartet.

The second section (beginning at bar 98) opens with a single, granulated cello trill on the root G, diffused to the rear of the hall. This is the first time that the tonic is clearly revealed, however, exact harmonic root is still slightly uncertain due to the slow and irregular glissandoing of the time-stretched trill. The quartet explores this idea with a series of slow, descending dominant seventh progressions which are continually interrupted by glissandos and the introduction of chromatic notes by the quartet and the electronics. By the fourth, and lowest crescendo, the quartet has resolved to a somewhat stable harmonic centre, but this quickly disintegrates into the penultimate ascent which is primarily driven by a two and a half octave upward glissando on the cello. The fifth and final ascent is essentially a refrain of the initial progression with a significantly increased dynamic. The final eight bars again reveal the slow glissando in the electronic part which is finally echoed and resolved by the cello and viola.

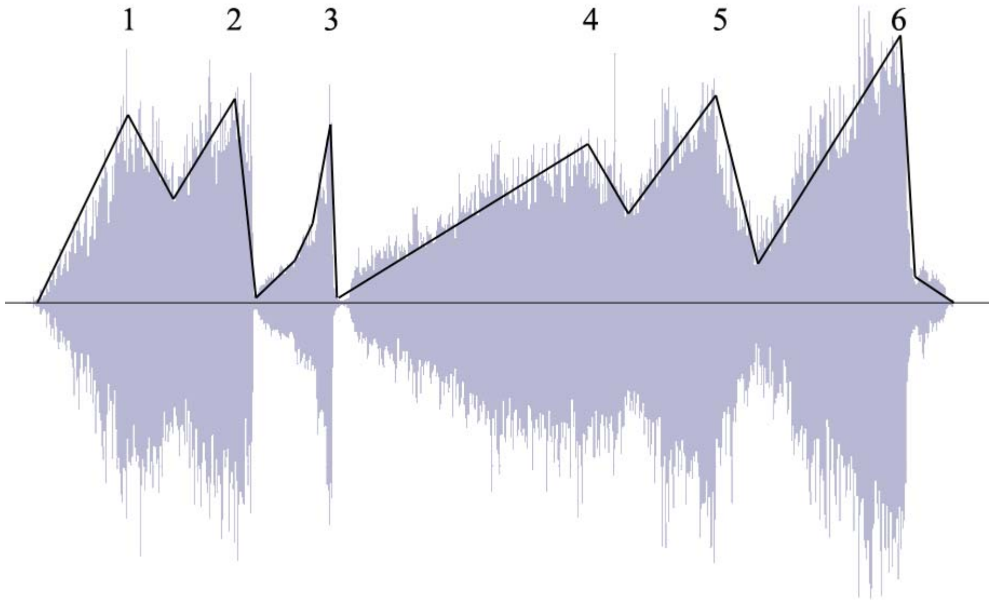

Fig. 10.7 Large scale structure of *String Quartet No. I*

Fig. 10.8 Diffusion score example

10.3.2 The Spatialization Process

The electronic part in this quartet is organized in a number of stereo files which are cued and triggered by the diffusionist using a Max/MSP patch. As the electronic parts are non-rhythmical, no click track is required, so the diffusionist must manually trigger the various sound files as indicated in the score. The patch contains two controls which sequentially trigger the audio files associated with the stereo channels A and B. The cues for the electronic part are indicated on a single line stave with a capital letter A or B and the associated file number, as shown in Figure 10.12. The azimuth control is indicated by a straight line arrow, above the stave indicates a narrow frontal azimuth (Figure 10.13 (b)), at the stave indicates a wide azimuth (Figure 10.13 (c)) while below the stave indicates a rear orientated azimuth (Figure

10.13 (d)). The example in Figure 10.12 indicates that the azimuth should move from approx 10-20° to 90° over the course of ten bars.

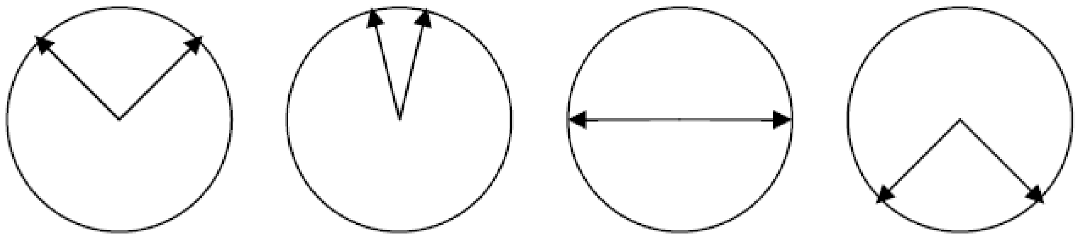

Fig. 10.9 Diffusion ranges in *String Quartet No. I*

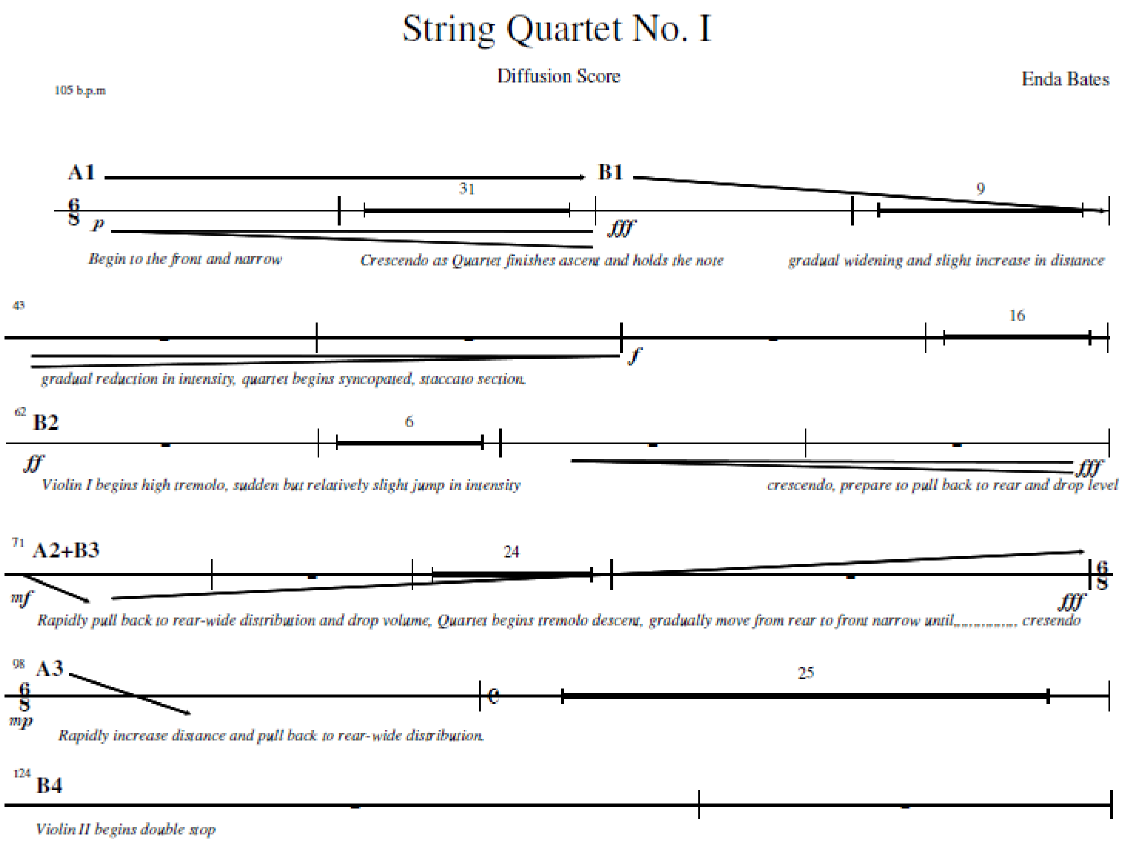

Fig. 10.10 *String Quartet No. I* - diffusion score

The two channels can be spatialized directly from the patch to a multichannel array using higher order Ambisonics, or alternatively, a stereo feed can be routed directly from the patch into a mixing desk for manual diffusion to a loudspeaker orchestra. If the spatialization is to be implemented using Ambisonics, then a MIDI controller is also required to adjust the distance and two azimuth controls during the performance. This controller (or the computer keyboard) is also used to trigger playback of the various audio files.

10.3.3 Critical Evaluation

In practice, the overall dynamic of the electronic part was restrained so that it coloured and shaded the quartet, without dominating, however, the dynamics of the electronic part should be pushed for dramatic effect at the main crescendo points in the piece. The electronic part primarily acts to extend and widen the quartet sound, and so is never entirely spatially separated from the frontal quartet. It is for this reason that the patch is arranged with stereo channels which are adjusted using the distance and azimuth controls. Azimuth in this case indicates the stereo spread and so acts as the primary spatial control for each channel, where as distance acts as the primary amplitude control. A suggested diffusion strategy is indicated in the diffusion score shown in Figure 10.14.

At the time of writing, this piece has been performed three times with the author as diffusionist in each case. The first two performances took place in quite large and reverberant acoustics (namely SS Michael and John's Hall, Dublin, Ireland and Muziekgebouw aan 't IJ, Amsterdam, Netherlands) using an eight channel array and Ambisonics, while the third performance took place in a medium sized theatre venue (the Beckett Theatre, Dublin, Ireland) using manual diffusion to three pairs of loudspeakers at varying distances from the audience. Perhaps surprisingly, for this piece the more diffuse sound of the larger venues and Ambisonics was preferable as this enhanced the blending of the synthetic and acoustic sounds which was one of the primary goals of this piece. In the smaller and less reverberant theatre, the loudspeaker positions (particularly the lateral loudspeakers) became much more apparent and again Ambisonics was beneficial in overcoming this problem. Performing this piece also revealed the significant skill and experience required to successfully and transparently diffuse sounds to a diverse range of loudspeakers. While faders on a mixing desk provide much more sensitive control than MIDI faders and ambisonic panning, this physical benefit is more than offset by the skill required to smoothly move sounds using fader movements alone.

10.4 Case Study IV: Auto Harp

Fig. 10.11 A “Rosen” autoharp

Auto Harp is an eight channel acousmatic work constructed from recordings of strummed arpeggios and individual plucked strings on the autoharp shown in Figure 10.5. This particular type of autoharp was manufactured in the former East Germany in the 1950s and 60s and second-hand instruments are now widely and cheaply available. East German factories were not allowed to put their own factory/family name on instruments during this time and so no identifying marks appear on these instruments apart from the distinctive rose decal. Hence these instruments have come to be collectively known as “Rosen” autoharps, from the German word for roses [White, 2008]. Autoharps are played by strumming with a pick or finger. Pressing one of a number of levers (six in this case) causes certain strings to be muted, leaving a chord. This results in an instrument which is very easy to play, but is also quite limited, as picking and muting individual notes is difficult due to the tight spacing of adjacent strings. The autoharp has therefore been predominantly used as an accompaniment instrument, particularly in American folk music.

The acousmatic piece, *Auto Harp*, is an attempt to extend the musicality of this simple instrument beyond its physical limitations. Smalley’s ideas of gestural surrogacy therefore play an important part in this work as at various times the focus shifts from clearly instrumental gestures (first-order surrogacy) to static environmental textures (second-order and remote surrogacy) created from the complex, sustained resonance which results from a strummed arpeggio through all twenty-four undampened strings of the autoharp. Various strummed arpeggios and picked sequences were recorded using both monophonic, stereo and Soundfield microphone arrangements, and these arpeggios form the basic building blocks of the entire work. These clearly instrumental passages, presented in static spatial locations, represent the first order surrogacy material, which is then abstracted to second-order and remote surrogacy using a variety of techniques. Each individual string was also recorded in isolation, and these samples are used with various algorithmic techniques to generate new sequences. Although these algorithmic passages are still clearly instrumental in origin, the rapid spatial distribution and complicated picked sequences produced are somewhat removed from the basic first order gestures produced initially from static spatial locations. More remote levels of surrogacy are attained through the use of granulation time-stretching techniques which emphasize and sustain the complex resonance of the vibrating strings. The different material used in this piece is arranged in terms of first-order, second-order, and remote surrogacy in Figure 10.6. The only material which is entirely synthetic, and hence displays remote surrogacy, is the layered sine waves used in the two Franssen sections that open and close the piece. The title of this section comes from the psychoacoustical illusion, the Franssen Effect, which is used throughout the opening section of this work.

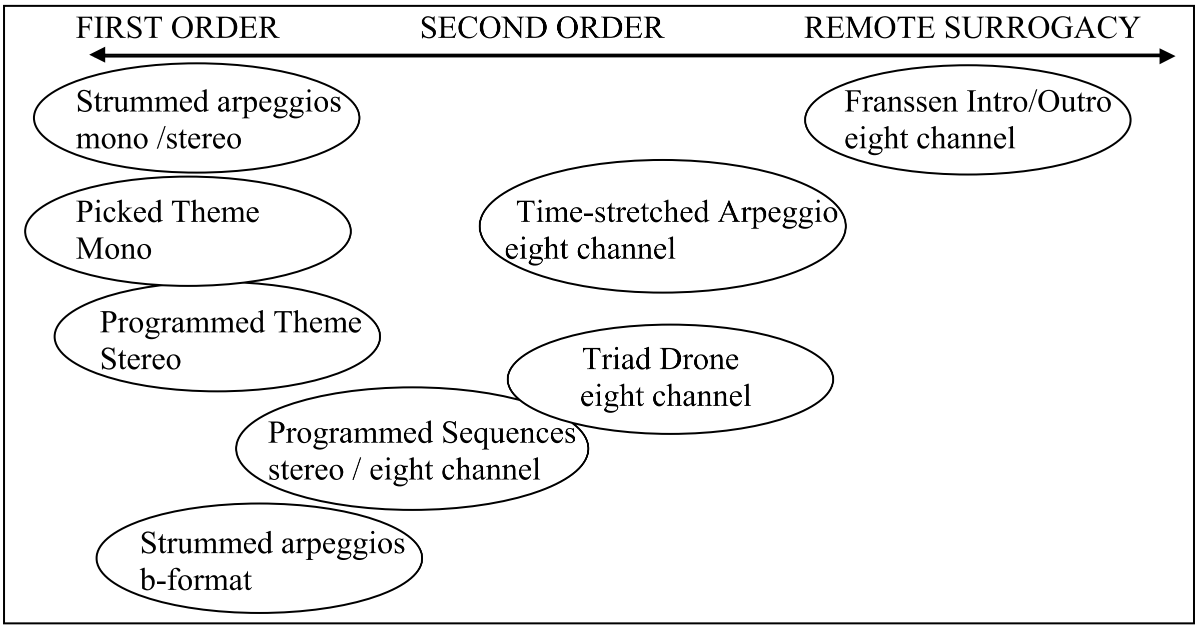

Fig. 10.12 *Auto Harp* source material, grouped according to order of surrogacy

The Franssen effect is an auditory illusion which produces large and unpredictable localization errors in human listeners [Franssen, 1962]. This illusion was first reported by N. V. Franssen in his 1962 thesis on stereophonic localization. In the initial experiment, two loudspeakers were arranged symmetrically in front of a single listener. A sine tone with a sharp onset transient is routed to one loudspeaker and immediately begins to exponentially decay over a fixed time period, *t,* while concurrently, the same signal exponentially rises in the other loudspeaker, also over the same period, *t,* as shown in Figure 10.7. Franssen found that listeners consistently localized the source signal to the initial loudspeaker, even after the signal had entirely moved to the other loudspeaker, and this illusion has since been referred to as the Franssen effect. This illusion was investigated further by Hartmann and Rakerd in 1989 in a series of detailed tests which revealed the nature and limitations of this unusual effect [Hartmann *et al*, 1989]. They discovered that the Franssen effect is entirely based on the inability of a listener to accurately localize a steady-state sine tone in a reverberant space. The source is first perceived at the location of the initial onset transient, and as the following steady-state sine wave provides no further reliable localization cues, the perceived source location remains unchanged, even after the real physical location of the source has moved to the other loudspeaker. Hartmann's experiments revealed that the Franssen effect is highly dependent on various factors such as the source signal, the transition time, *t,* and the listening environment [Hartmann *et al*, 1989]. Firstly, Hartmann reported that the illusion fails entirely for broadband noise signals, as these spectrally rich signals produce viable localization cues even in the absence of a sharp onset transient. Secondly, the illusion fails entirely under anechoic conditions as in this case the steady-state localizations are sufficient for accurate localization.

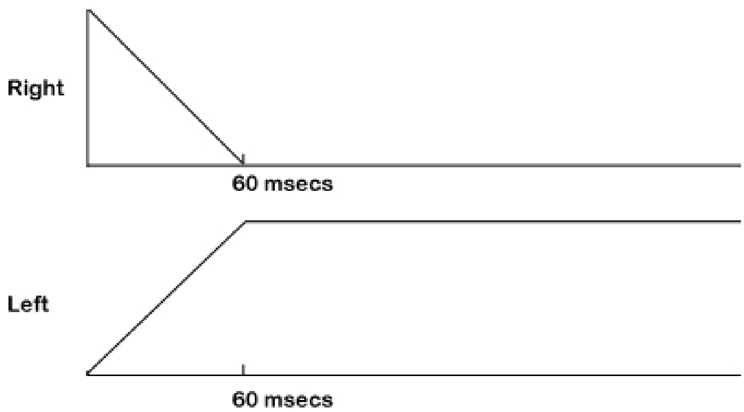

Fig. 10.13 Loudspeaker envelopes in *Auto Harp* Franssen section

String No.	Harmonic	Freq
1	1	261. 6
2	1	261. 6
3	3	392. 4
4	7	457.8
5	2	523. 3
6	9	294.3
7	5	327.0
8	11	359.7
9	3	392. 4
10	7	457.8
11	2	523. 3
12	9	294.3
13	5	327.0
14	21	343. 4
15	3	392. 4
16	7	457.8
17	15	490. 5
18	17	278.0
19	19	310.7
20	5	327.0
21	23	376.1
22	3	392. 4
23	7	457.8
24	2	523. 3

Table 10.1 *Auto Harp* tuning scheme

10.4.1 Analysis

The introductory section of *Auto Harp* uses the loudspeaker amplitude envelope shown in Figure 10.7 in conjunction with twenty-four sine tones which correspond to the twenty-four strings of the autoharp. Each string is tuned to a different harmonic (octave adjusted) of the lowest note (C 261.6Hz) as shown in Table 8.1. Each harmonic is adjusted so it falls within a two octave range. This spectral tuning produces a complex resonance with very little beating due to the close harmonic relationship between each string. Each sine tone onset is presented in sequence from the four loudspeakers to the right, moving from front to back, and is shifted to the matching loudspeaker on the left using a transition time *t* of 60ms. After all twenty-four sine tones have been introduced, reversed recordings of each individual string are gradually introduced at the same real spatial locations as the associated sine tone, and the entire sequence builds to a sharp crescendo. The overall effect of this sequence is to delay and disguise the movement of the layered sine tones from right to left, as the initial onsets are clearly localized to the right, while the final reversed autoharps recording clearly crescendo on the left side. Localization during the middle steady-state section will therefore vary depending on the acoustical environment, the individual listener and the associated strength of the Franssen effect, introducing an element of variability in a composition for fixed media.

The overall structure of *Auto Harp* is marked on the waveform of a monophonic reduction of the eight-channel work in Figure 10.8. The crescendo which marks the end of the initial Franssen section is articulated with a spatial sweep from front to back on the left side of the array. This is immediately responded to by a sweep through each individual note in a corresponding spatial motion from back right to front right. The entire Play section which follows uses this type of spatial dialogue between static arpeggio recordings and dynamic, algorithmic sequences. In each case, a recorded arpeggio (mono, stereo or b-format) produces an algorithmic response, which is constructed from the individual string recordings, pitch shifted up an octave. The spectromorphological profile of each recorded arpeggio was carefully applied to the algorithmic response by matching the duration of each sequence. Karlheinz Essl's Real-Time Composition Library for Max/MSP was used to create these algorithmic responses [Essl, 1995]. The *series* object was used to create a random sweep through all twenty-four notes, while the *sneak* object randomly selects notes but only chooses notes adjacent to the currently selected note. This object was used to create the distinctive and rapid sequences of notes which are prominent in this section.

The next section is introduced by a mono recording of the plucked sequence which functions as the main melodic theme. The theme is presented four times in this section, initially as the performed sequence in the original mono recording. This is followed by a sequenced recreation of the theme using the individually recorded strings which are spread across the front pair of loudspeakers. The final two iterations of the theme use both the original progression, and various sequenced progressions which are distributed across each of the four loudspeaker pairs. In this case, the sequenced progressions are also transposed to harmonize with the initial progression. A granulated time-stretched version of this progression is also introduced in this section, shifting the focus away from the highly gestural material which has dominated up to this point. The spatial distribution and dynamic of this drone alters in response to the instrumental gestures and at times the drone itself displays strong spatial gestures. This is evident in the crescendo that ends this section, and also at

3’48, when the diffuse drones builds and then suddenly collapses to low frequency drone statically positioned in the front loudspeaker pair.

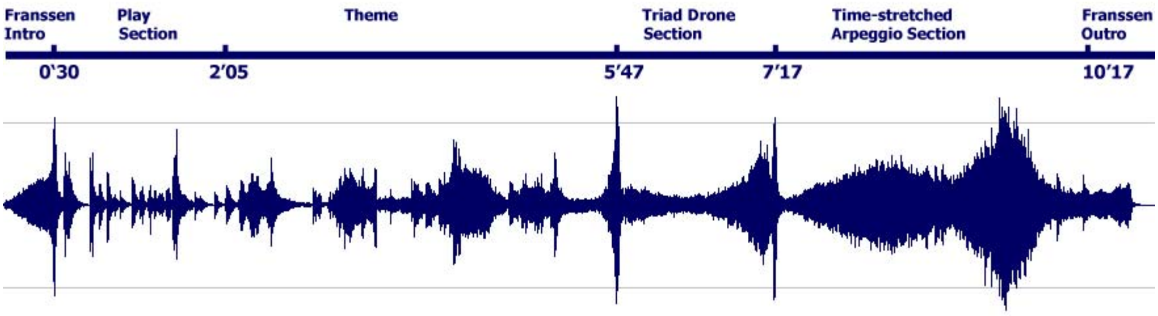

Fig. 10.14 Monophonic reduction of *Auto Harp* indicating section durations

The *Triad Drone* section is the first to move away completely from the highly gestural arpeggios and sequences which dominated the opening half of this piece. The first harmonies which are not contained within the tuning system shown earlier are also introduced in this section. Multiple stereo channels of a spectrally rich drone were constructed from a recorded arpeggio which was time-stretched using the *granule* opcode in the Csound processing language (Csound is discussed in more detail in chapter 12.3). Two of these decorrelated stereo channels quickly begin a slow upward glissando before eventually settling at two pitches to form the triad drone which gives this section its name. The two pitch-shifted drones were tuned by ear to produce a relatively consonant harmony. The tuning of these two new drones corresponds approximately to the fifth and ninth harmonics of the original root note.

The penultimate section is, as its name suggests, primarily constructed from a time-stretched arpeggio constructed using a granulation algorithm in Max/MSP. The resulting drone quickly separates into a dense cloud of clicks created from the initial plucking action and a thick, mid-frequency drone created from the resonance of the instrument. This spatially diffuse sequence decays slowly and then builds again to a final crescendo which is punctuated by occasional unprocessed arpeggios. The piece ends with a reversal of the initial Franssen section, as the drone fades into associated sine tones which collapse spatially in the same way as they appeared.

10.4.2 Critical Evaluation

The use of the Franssen effect in the opening section of this work has proven to be highly effective, although the precise effect perceived is highly dependent on both the acoustic of the performance space and the position of the listener within the array. The onset of each tone is generally well localized to the right of the array, and the perceived source direction then gradually shifts from right to left before collapsing to a spatial arpeggio down the left hand side of the array. However, due to the inherent difficulty in localizing sine tones in a reverberant environment, the speed at which the sine tones move from right to left in this section is quite unpredictable, which introduces a nice element of variability into this fixed media piece.

The autoharp is in many respects an ideal source for a work of spatial music as each plucked note contains a sharp, percussive transient that is highly localizable, followed by a more diffuse, sustained pitch. In this way, each note can play the role of either an instrumental gesture (for example, the programmed spatial arpeggios) or environmental texture (the time-stretched, sustained drones). Furthermore, the instrumental gestures can be divided into first order surrogacy, recorded instrumental gestures, or second order surrogacy, programmed gestures, and moving between these different levels of surrogacy provides much of the impetus for this piece. The inherent and quite dramatic nature of the basic gesture, i.e. the plucked arpeggio, is also clearly related to the programmed arpeggios, while the sustained pitch of the initial gesture also blends very well with the time-stretched, environmental like drones. Finally, the use of both Ambisonics and amplitude panning is a highly effective means of spatially segregating the different layers of material. This is particularly true of the close-miked, soundfield microphone recordings which are spatially quite distinct from the surrounding and immersive drone material.

11 Spatial Music Instruments

The preceding chapter explored how different composers have attempted to connect the musical gestures of an instrumental performer to the spatial gestures of the electronic part. While some composers have turned to synchronised tape parts or real-time electronic processing to achieve this connection, others have attempted to design new instruments which incorporate the spatialization processes. While various new musical interfaces have been developed which feature some form of spatial control, the idiosyncratic nature of these devices mean they are unlikely to be widely adopted. Augmented instruments, i.e. traditional musical instruments featuring additional hardware or sensors, represent a possible solution to this problem, as they can potentially combine existing and sophisticated instrumental practice with spatial or timbral processing algorithms. The mapping of musical gestures to electronic processes is a critical issue in the design of any augmented instrument. The composer David Wessel suggests that "musical control intimacy and virtuosity require both spatial and temporal precision in the sensing of gestures (control intimacy refers to a tight connection between a body movement and change in an auditory feature)" [Wessel, 2006]. One form of augmentation specific to stringed instruments is the use of polyphonic pickups which produce a separate audio signal for each string. The discrete multi-channel output of these instruments would seem to be very suitable for spatialization to a multi-channel loudspeaker array. By linking the spatial location to the choice of string, the spatialization process can be synchronized to the physical performance of the instrument. In addition, spatialization algorithms and other processes can be applied to each individual string as required. This signal processing approach to instrument augmentation has the advantage that the performer does not need to learn any new gestures or instrumental techniques. In addition, the necessary hardware has become widely available, particularly for electric guitar, and can often be retrofitted non-destructively to an existing instrument. The electronic violin developed by Max Mathews [Boulanger, 1986] and the Hyperinstruments developed by Todd Machover [Machover, 1992] are two examples of augmented stringed instruments which incorporate polyphonic pickup technology. The polyphonic output of these instruments was used extensively by the composers who wrote for them. As many manufacturers now produce electric instruments with polyphonic outputs, this approach represents a viable and generalised solution to stringed instrument augmentation that is not tied to specific hardware.

11.1 Augmented Strings

The electronic violin developed by Max Mathews in 1986 (see Figure 11.1) was an attempt to use the considerable dynamic, expressive timbral range and highly evolved instrumental technique of the violin as the musical source, modifier, and controller of various real-time performance networks [Boulanger, 1986]. The electronic violin used four contact microphones inserted into the bridge to pickup the sound of the strings. Mathews noticed that when the outputs from the four strings were combined and passed through a single amplifier and speaker, nonlinearities resulted in unpleasant combination tones. To eliminate this problem, Matthews used a separate amplifier and loudspeaker for each string. When the composer Richard Boulanger later came to compose a piece for the instrument, he commented that the "discrete four-channel output of the instrument significantly directed the composition in a number of ways" [Boulanger, 1986], specifically toward a composed spatial component. In the resulting piece *Three Chapters from the book of Dreams* Boulanger supports and contrasts the linear counterpoint and compound melody with a concurrent spatial counterpoint. He comments "By assigning the output of each string to a separate speaker, the audience is given the unique sense that they are seated within the violin body. This spatial component is based on the inherent design of the instrument and its antiphonal treatment is at the same time quite old and quite new. " [Boulanger, 1986]

The Hyperinstrument group at MIT Media Lab have been researching and developing augmented instruments since the late eighties [Machover, 1992]. The first of these, the Hypercello (shown in Figure 11.2), was completed in 1991 and combined an electroacoustic cello with additional sensors to provide data on various performance parameters such as bow position, placement and pressure, string and finger position and pitch tracking. The composer Tod Machover worked with the renowned cellist Yo-Yo Ma to create *Begin Again Again*, an interactive composition in which different playing techniques such as tremolo, bow bounce, pizzicato and legato are mapped to various electronic processes including spatial movement.

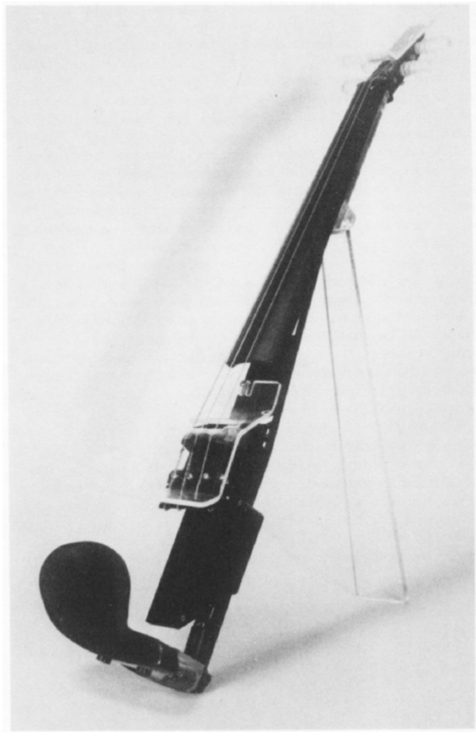

Fig. 11.1 The Max Mathews augmented violin

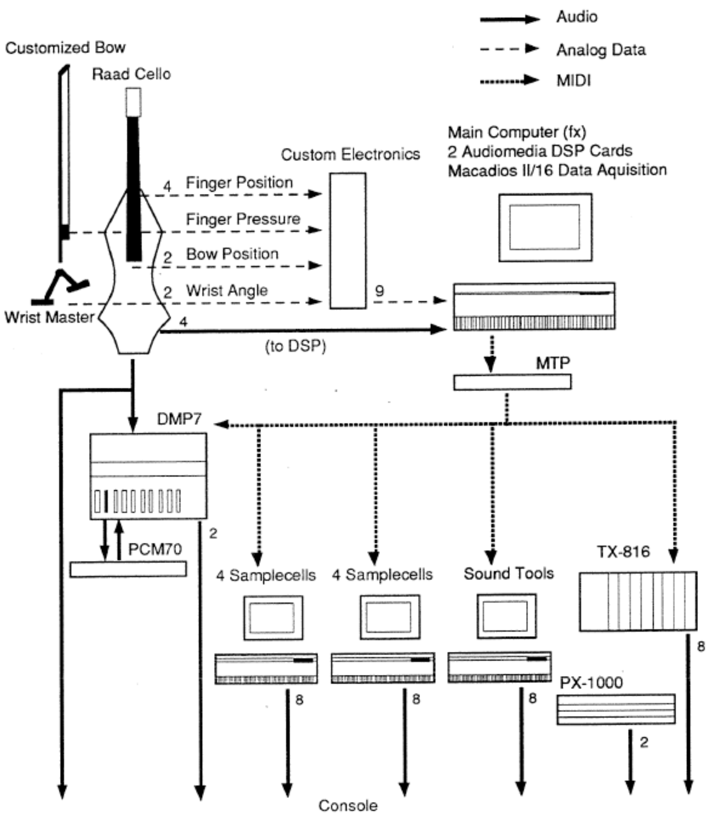

Fig. 11.2 The Hypercello system

11.2 The Hexaphonic Guitar

Polyphonic pickups (also called divided or split pickups) have been widely used over the past three decades to detect and convert the pitch coming from individual guitar strings into MIDI messages. The majority of these systems however do not provide split signals for external processing, preferring instead to multiplex the signals into a single cable which can then be easily connected to a MIDI converter. This emphasis on MIDI capability is changing however and some dedicated systems such as the Gibson HD. 6X-Pro have been developed specifically for polyphonic processing. The initial prototyping of the instrument was carried out by Adrian Freed and the Guitar Innovation Group at UC Berkeley's Centre for New Music and Audio Technologies (CNMAT) [Yeung, 2004]. Further research has since been carried out at CNMAT on polyphonic guitar effects based on vocal-tract modeling, frequency localized distortion and coordinated equalization [Jehan *et al*, 1999]. The co-director of CNMAT, David Wessel has also conducted research on augmented instrument design and interactive computer music with a particular focus on the live performance of improvised computer music. Wessel had used polyphonic pickups in performances such as *Situated Trio*, an interactive live performance for a hexaphonic guitarist and two computer musicians with expressive controllers [Wessel *et al*, 2002]. In this piece, the polyphonic guitar signal is processed by the two computer musicians using various algorithms such as granulation, looping, cross-synthesis and spatialization. The polyphonic guitar signal is also converted to MIDI to provide a high level discrete event representation of the guitarist's performance for triggering automated computer based processes.

The adaptation of existing MIDI guitar systems is an alternative and cost effective method for deriving a polyphonic signal from an electric guitar. Popular MIDI guitar systems by Roland, RMC and AXON use a 13-pin connector which carries the six individual signals from the hexaphonic transducer, a mono audio feed from the guitar's normal pickups, and some controls specific to the Roland system. The wiring diagram shown in Figure 11.3 indicates the individual string signals from the 13-pin connector which can be accessed simply by wiring the appropriate pins to several 1/4 inch jack connectors. A schematic for such a breakout box is shown in Figure 11.4 [Berg, 2007].

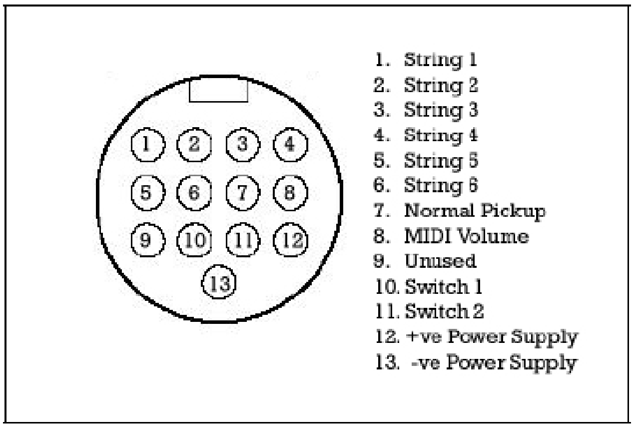

Fig. 11.3 Roland 13-pin wiring diagram

Fig. 11.4 13-pin breakout box schematic

11.3 Case Study V: Etude No. 2 for Hexaphonic Guitar

Writing for a hexaphonic guitar in which each string is individually spatialized requires the composer to carefully consider which strings will be used to produce the chosen pitches. In this scenario, the routing and particular spatialization method used will inform the composition in a very real and tangible way. In addition, the relationship between the visible and spatially static guitarist and the spatially distributed sounds must be carefully considered. In the beginning of this piece, the spatial distribution of the guitar strings is disguised and this helps to forge a connection between the visible performer and the diffused sound. Once this connection is made, each distinct spatial location is then gradually introduced and layered to eventually create a large, enveloping drone which gradually overcomes the live performer and closes the piece.

11.3.1 Analysis

In this etude for hexaphonic guitar the strings are consecutively routed in pairs through three instances of Ambience, the freeware reverb VST plugin by Smartelectronix (see Figure 11.5). Various plucked harmonics and intervals are sustained using the reverb hold function which is triggered with a MIDI foot controller. After each group of three harmonics/intervals have been layered, the combined drone is then dynamically routed to a spatialization algorithm (2^{nd} order Ambisonics using the ICST externals for Max/MSP) which chaotically pans the drone around the entire array. This entire process is then repeated three times with increasing intensity, culminating in a very loud drone which dominates the instrumental part. This process is highlighted during the final section as the performer remains motionless while the vast and spatially dynamic drone continues, culminating in a loud crescendo.

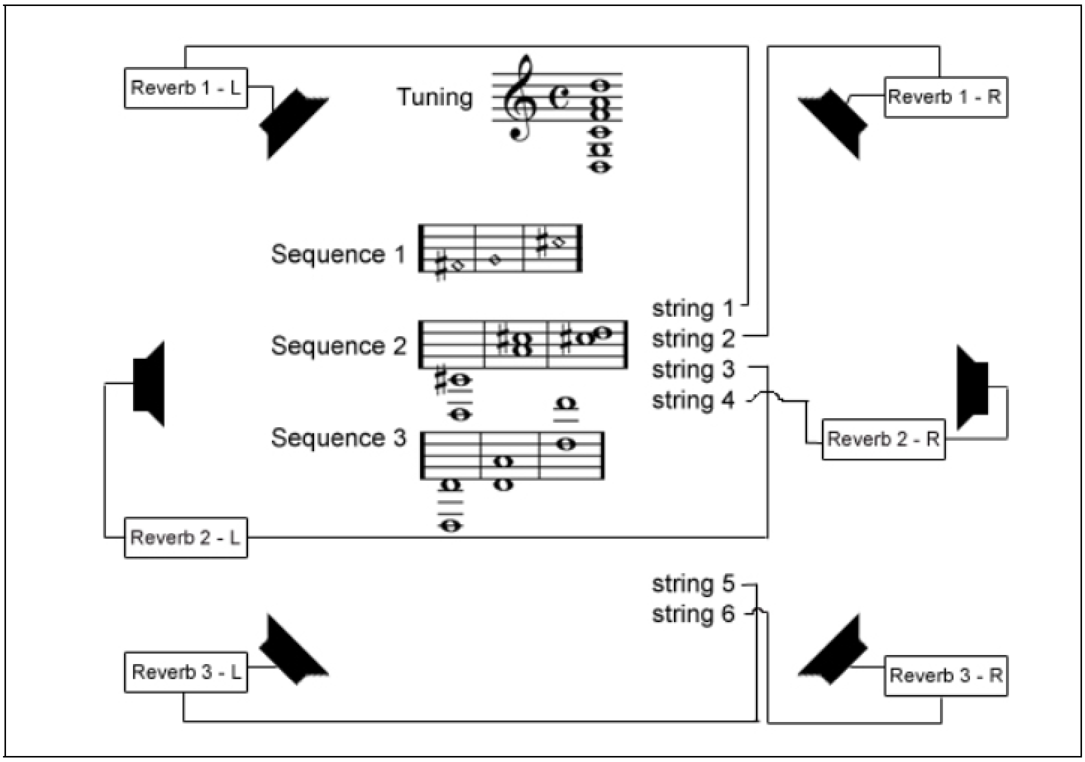

Fig. 11.5 *Etude No. 3* tuning, spatialization and interval sequence

11.3.2 Critical Evaluation

The Ambience VST plugin produces a highly synthetic reverberation that is much less realistic than many other reverberation algorithms, however, artificial reverberation is used in this piece more as an effect than as a realistic simulation of an acoustic space. This becomes particularly obvious when the sustain function of this reverb effect is repeatedly engaged and the various drones are reproduced using each of the three pairs of loudspeakers.

The percussive sounds at the start of the piece are produced by striking the six, muted strings with the heel of the hand behind the bridge and the resulting sound is therefore distributed relatively equally among the six loudspeakers. The staccato, rhythmical nature of the material and the enveloping, nondirectional nature of the spatial distribution helps to create an initial connection between the instrumental player and the spatialized audio. The following introduction of single, sustained notes is perhaps the first time that the spatial distribution of the guitar strings is made clear and this gradual introduction of the spatialization scheme helps to reduce the contrast between the location of the performer and the resulting sounds. As this process is repeated in different ways and with an increasing dynamic, the layered drones gradually build to a vast, spatially dynamic noise drone which becomes increasingly remote from the actions of the performer. In the very final section, this separation is completed and the performer rests, motionless while the spatial drone plays for approximately 20-30 seconds, before triggering the end of the piece at whatever point they choose.

12 Behavioural Space

Modern computer technology has now advanced to such a degree that an ordinary laptop computer and suitable audio interface is capable of processing and spatializing numerous channels of audio in real-time. These developments have allowed composers to utilise more advanced algorithms to control the spatialization of sounds. Flocking algorithms have proved to be particularly popular in this regard and conceptually, the notion of flocking, or swarming, contains obvious connotations in terms of spatial motion. Flocking algorithms such as Craig Reynold's Boids [Reynolds, 1987], model the overall movement of groups of autonomous agents, such as a flock of birds or a shoal of fish, and have been widely used in computer graphics and animation. Recently, these algorithms have also been adapted for musical applications including spatialization.

12.1 The BOIDS Flocking Algorithm

In his investigation into the motion of flocks of birds, Reynolds discovered that the aggregate motion of the flock as a whole can be simulated through the interaction of autonomous individual agents following a number of simple rules [Reynolds, 1987]. He identified the three primary factors of separation, alignment and cohesion which govern the aggregate motion of the flock. These factors are represented as a set of behavioural rules for each individual agent, whereby each agent avoids colliding with nearby agents, attempts to match the velocity and heading of nearby agents, and attempts to move toward the centre of the flock.

The 3D implementation of Boids for Max/MSP/Jitter [Singer, 2008] (shown in Figure 12.1) generates a visual representation of the flock and outputs the current and previous position of each individual agent at a specified period as Cartesian coordinates. The behaviour of the flock is modified using controls such as speed range, strength of centring instinct, neighbour speed matching, avoidance and preferred distance and the strength of attraction to a specified attract point, which allows the user to direct the overall movement. Other implementations of the Reynolds flocking algorithm include a *predator* which the flock will fly around and avoid, as shown in Figure 12.2 [Martin, 2005].

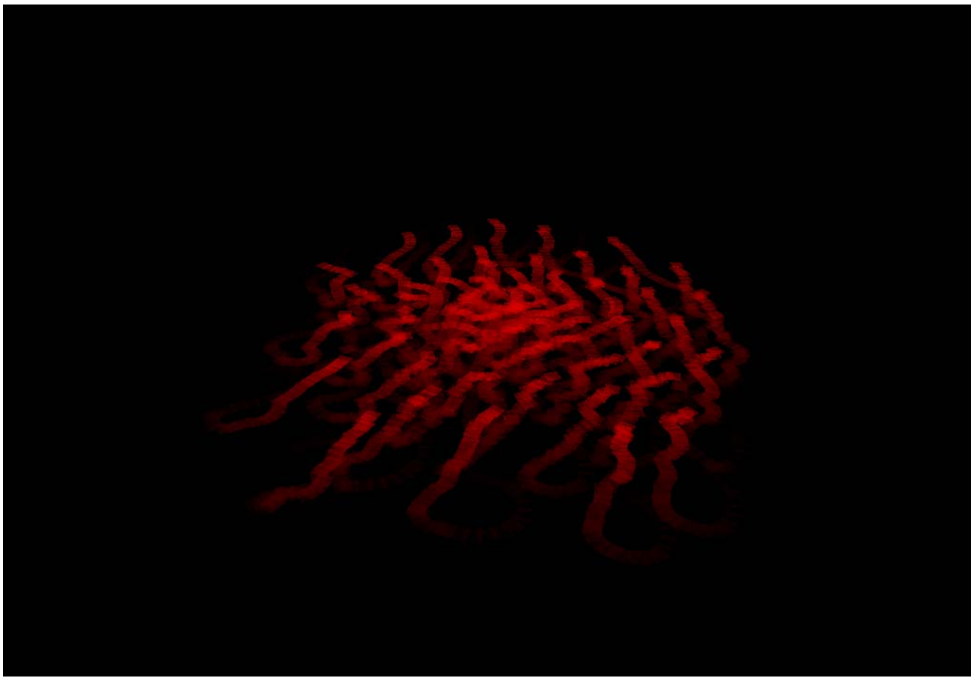

Fig. 12.1 Eric Singer's Boids for Max/MSP/Jitter

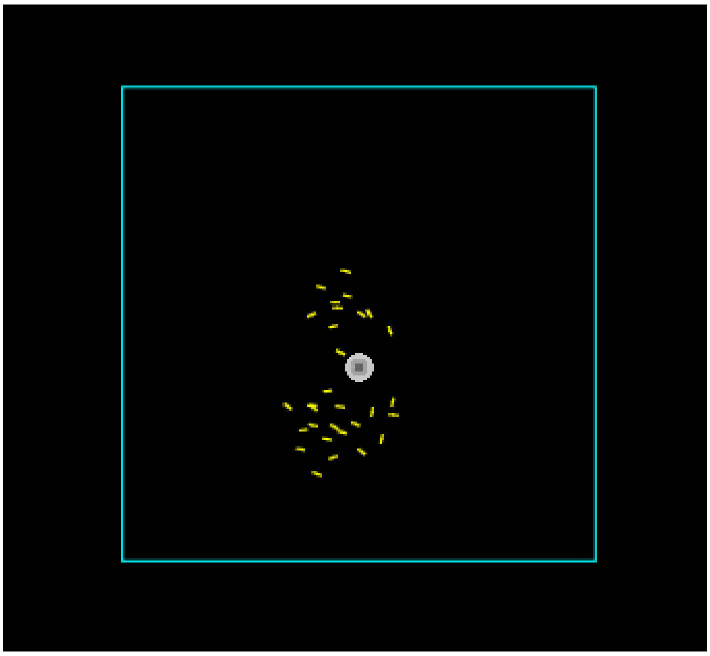

Fig. 12.2 Illustration of a flock avoiding a predator

Flocking algorithms such as Craig Reynold's Boids [Reynolds] have been adapted for various music applications such as interactive audio installations [Unemi *et al*, 2005], real-time music improvisation by a computer [Blackwell *et al*, 2004] and real-time algorithmic composition [Uozumi, 2007]. Flocking algorithms have also been used in conjunction with granular synthesis and both techniques would seem to share many characteristics. Both use large numbers of individual agents or grains, both produce macro-level complexity through relatively simple micro-level structures, and both are indirectly controlled through various micro and macro level parameters. Flocking algorithms would therefore appear to be highly suitable as a control mechanism for granular synthesis and also suggests an obvious spatialization scheme where each grain is positioned in space according to the location of a single agent in the algorithmically generated flock.

12.2 Case Study VI: Flock

The acousmatic composition *Flock* was constructed in the Max/MSP environment using the following processes;

- Eric Singer's implementation of Craig Reynolds' flocking algorithm, Boids
- Nobuyasu Sakonda's granulation algorithm, granular 2.5
- ICST Higher Order Ambisonics Externals
- IRCAM's Le Spatialisateur

These processes were combined so that each grain corresponds to a single element in each flocking algorithm. The flock of individual *grainboids* are then spatialized using one of the two spatialisation algorithms.

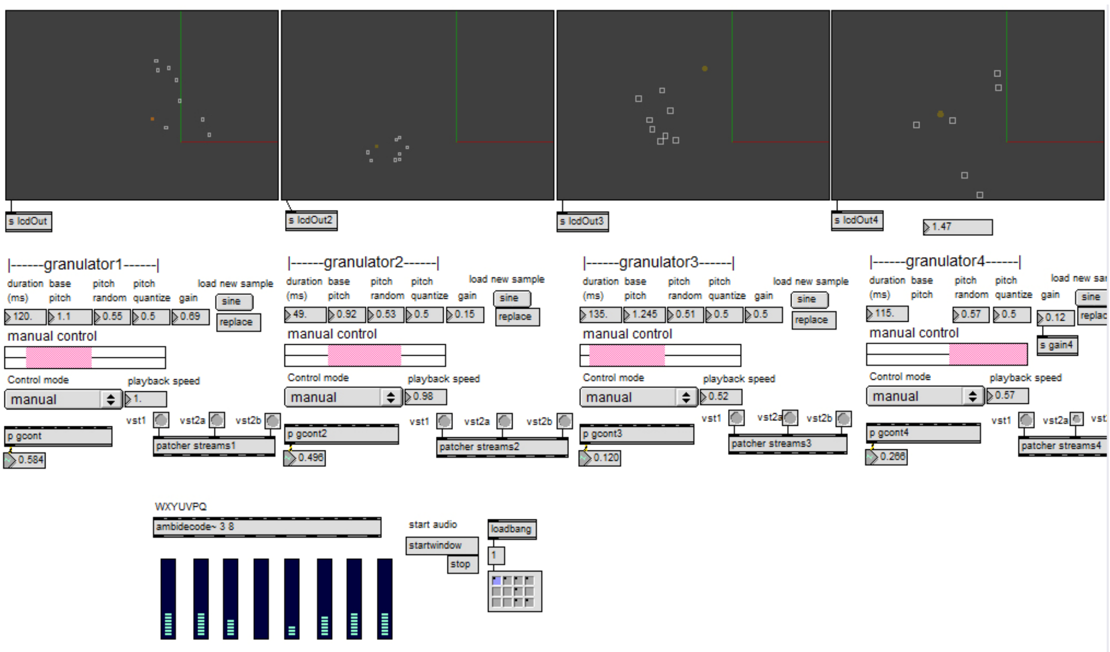

Fig. 12.3 Max/MSP Patch used to construct the composition *Flock*

The source material used in this piece consists of noise/click samples and pre-recorded tremoloed notes on viola and cello. The overall movement of each flock can be controlled with a user controlled *attract point.* While the individual spatial movement of each *grainboid* is controlled by the flocking algorithm, the overall movement of each flock can be adjusted by the user in a number of ways. Firstly a bounding box can be created which will contain the flock within a certain specified area. Alternatively, the strength of the flocks attraction to the user-specified *attract point* can be used to direct the flock to follow a particular point in space. In this piece, noise/click samples and pre-recorded tremoloed notes on viola and cello are associated with a particular flock of eight *grainboids* using the Max/MSP pitch shown at the start of this section in Figure 12.3. Each flock follows a slightly different overall trajectory, using the associated *attract point* with a medium attraction strength setting so that the specified path is only loosely followed.

12.2.1 Analysis

The nature of the source signal has a large bearing on the effectiveness of the perceived spatialization (see Chapter two) and the source material for this piece was chosen with that in mind. Localization accuracy has been found to improve when the source signal contains plenty of onset transients, and when broadband signals are used. The initial testing of this system was carried out using a noise/click sample for this reason and this material was then also used in the final composition. Tremoloed string samples were chosen as they provide pitched material which also contains lots of irregular transient information due to the rapid bowing action. In addition, the fast irregular bowing motion recorded in the original tremolo string samples is well matched to the granulation process.

Two spatialisation methods were used in order to investigate the perceptual difference between these two techniques, namely amplitude panning and Higher Order Ambisonics. A subtle yet distinct difference was perceived between these two spatialization schemes, however, this difference is certainly also influenced by the artificial reverberation and early reflections added by IRCAM's Le Spatialisateur program. This perceptual difference, although relatively subtle, does help to spatially distinguish the various flocks and heightens the spatial micro-complexity of the piece. The Doppler Effect was utilized throughout the work, perhaps surprisingly considering the presence of pitched material. However, the view was taken that the pitch variations introduced by the spatial movement and resulting Doppler shift was aesthetically pleasing and fitted the micro-variations in pitch resulting from the rapid bowing action of the original samples, as well as supporting the perceived movement of each flock.

12.2.2 Critical Evaluation

The flocking algorithm spatialization scheme used in this piece produced a complex spatial motion which is rich in micro-level detail. The precise motion of each individual *grainboid* is not clearly perceptible, however the overall spatial movement is highly coherent and very evocative. The complex spatial motion produced by the flocking algorithm also has a significant effect on other parameters such as timbre and particularly, dynamics, as this parameter is closely linked to the distance of each flock from the centre point. The precise harmonies are therefore somewhat indeterminate, as the flocking algorithms control the precise timing of the movement of each flock through the centre point. Overall, the piece follows a loose ABA structure which is delineated by two low register pedal tones underneath the changing harmonies produced by the other pitched flocks. The main harmonic change at the start of the B section occurs at the approximate halfway point (5’15) while the return to the final A section occurs gradually as the new pedal tone flock is gradually replaced by the original.

The loose, almost drone-like aesthetic utilized in this composition was greatly influenced by the real-time operation of the synthesis and spatialization methods. While real-time control of these processes is highly suitable for performance or experimentation, it is perhaps less suitable for the creation of longer works due to the very large number of parameters involved. In addition, the real-time operation of this system placed significant limitations on the amount of material which can be generated concurrently. It was found that a maximum of thirty-two grainboids in four flocks (as shown in Figure 12.3) could be realised using Ambisonics encoding, while IRCAM’s Le Spatialisateur processor could at most realise a single flock of only eight grainboids (based on a 2GHz Pentium Processor with 2GHz of RAM). The finished work therefore had to be constructed from individual layers of pre-composed material which were then later sequenced together. The Boids algorithm is not itself computationally expensive and flocks of many hundreds of agents can be generated, controlled and visualized in real-time. However, when combined with advanced audio synthesis and spatialization techniques, computational limitations restrict what can be achieved in real-time. Some of these issues have been investigated by the Interactive Swarm Orchestra (ISO) project at the Institute for Computer Music and Sound Technology (ICST) in Zurich (the ICST has also released a popular set of Higher Order Ambisonics externals for Max/MSP). The ISO project similarly employs flocking algorithms to control real-time computer sound synthesis and 3D sound positioning, using a collection of generic open-source C++ libraries [Bisig *et al*, 2008]. Some encouraging preliminary work has been carried out at the ICST on Ambisonics Equivalent Panning (AEP), a computationally efficient implementation of Ambisonics suitable for large numbers of sources [Neukon *et al*, 2008]. However, the addition of distance effects in the form of artificial reverberation is still too computationally demanding to achieve in real-time for a large number of sources. In addition, this real-time, performance-oriented approach also inevitably restricts the complexity and sophistication of compositions realized with these systems. While the various real-time implementations of this technique are highly suitable for performance applications, an offline approach is in many ways much more suitable for the creation of longer works and provides a much greater degree of control over the synthesized material.

12.3 Offline Spatialization with the Boids Algorithm

Csound [Boulanger, 2000] is a well established synthesis language, widely used in computer music composition, which supports a variety of granular synthesis techniques [Roads, 2004]. In Csound, a synthesis method, defined in an instrument in the orchestra file, produces macro-level sound events according to note statements in an associated score file. In the case of granular synthesis, each note statement corresponds to a micro-level event (although higher level granular synthesis opcodes are also available). This approach trades off instrument complexity for score complexity and provides a great deal of control over the synthesis, however, it also requires a very large score file. Granular event generator programs such as CMask [Bartetzki, 2008] can be used to create these micro-level score files using various macro-level probability and stochastic functions. For this piece, a number of Csound granular event generators and instruments, entitled AmbiBoid were developed by the author to implement various granular synthesis techniques such as granulation, glisson synthesis and grainlet additive synthesis. Each grain is then spatialized using the Boids algorithm and second order Ambisonics with distance modelling. [2]

The score file generators read spatial co-ordinates from text files generated by a Max/MSP patch containing Singer's Boids object. Due consideration was given to coding the Boids algorithm within either Csound or C. However, generating the spatial trajectories in real-time has a number of benefits, namely;

- it provides a clear visual representation of the spatial movement during the composition stage.
- it provides immediate feedback on the effect of changing flocking parameters.
- it reduces Csound rendering time.

In addition, this approach allows the same recorded trajectories to be applied to different synthesis methods, at different speeds, and with different dynamically changing parameters, so they can be used to generate musical building blocks, at a macro level. The Max/MSP patch converts the Cartesian coordinate pairs (previous and current value) of each agent to azimuth, elevation and distance values which are then written to three ASCII text files. The patch currently records a maximum of 128 agents but its modular design can be readily extended to larger numbers. The Boids object generates new values at a regular user-defined interval. In real-time applications this interval is often set equal to the grain duration so that the Boids and grain parameters are only updated when the grain envelope is at a zero point. However, with this approach a single grain duration must be used for all grain streams, which rules out many of the more sophisticated granular synthesis algorithms. The Ambiboid score generators therefore map each grain stream to an individual agent in the flock and linear interpolation to ensure that streams with different grain durations read through the spatial data at the same rate, preserving the flocking behaviour. The overall rate at which spatial coordinates are read is set by the minimum grain duration. Therefore, if the overall tempo of the spatial movement generated in Max/MSP is to be preserved, the boids object's clock period should be set equal to the intended minimum grain duration.

[2] The Ambiboib source code, along with various audio examples, can be downloaded at www.endabates.net/ambiboid.

Once the desired spatial trajectories have been performed and recorded in Max/MSP, the AmbiBoid utilities can be used to generate a Csound score file. Each utility/instrument pair implements a specific form of granular synthesis with shared parameters such as overall duration, number of grain streams, global reverb, etc. The density is controlled via the inter-grain spacing which sets the amount of silence between each consecutive grain as a multiple of that stream's grain duration. The density can be either fixed or dynamically varying between specified values at five hit-points equally distributed across the total duration. The grain duration can also be fixed or dynamically varying in a similar fashion. Various windowing functions such as Gaussian, Hamming and Hanning windows can also be chosen. Three different granular synthesis techniques were implemented, namely;

12.3.1 Granulation

The AmbiBoidgranule score generator and instrument performs straight forward granulation of a monophonic sample. The grain durations can either be a single fixed value for all streams, a single dynamically varying value for all streams, or a randomly chosen duration for each stream varying between a given maximum and minimum value. The playback position within the file for each grain is randomly chosen and the playback rate of the audio sample is defined in the score file.

12.3.2 Grainlet Additive Synthesis

Grainlet synthesis is a combination of granular synthesis and wavelet synthesis developed by Curtis Roads [Roads, 2004]. In this synthesis method, the duration of each grain is related to the frequency of the source signal within that grain. In this additive synthesis version, each grain stream corresponds to a single partial whose grain duration is related to the frequency of that partial. The user specifies the fundamental frequency and the number of grain streams/partials. The grain duration consists of one hundred cycles of the signal, which results in a grain duration of one second for the fundamental, 31ms for the 32nd harmonic, 15ms for the 64th harmonic, etc. The source signal consists of either a sine wave or a set of harmonically related sine tones.

12.3.3 Glisson Synthesis

In glisson synthesis, each grain, or glisson, has an independent frequency trajectory, an ascending or descending glissando [Roads, 2004]. The perceived relationship between spectral space and physical space from such terms as low and high pitches, or ascending or descending melodies. The mapping of spatial elevation to source frequency would therefore seem to be a natural and intuitive choice for this combination of glisson synthesis and the three dimensional spatial movement of the flocking algorithm. In this implementation, the source signal frequency can be mapped to spatial elevation in two ways. The maximum and minimum elevation angles (±90^{o}) can be linearly mapped to a specified frequency range, or alternatively mapped as octaves so that each glisson contains an octave glissando whose direction depends on the spatial movement. The usual density and grain durations controls are also available.

Each instrument uses standard Csound opcodes to generate the window function and source signal but most of the instrument code is concerned with the spatialization algorithm. The current and previous spatial coordinates are changed into a dynamic k-rate variable which is used to spatialize each individual grain with distance effects, early reflections and global reverberation. These distance effects are modelled using Chowning's algorithm for the simulation of moving sources, discussed previously in Chapter Three. In this model, the intensity of the direct sound reaching the listener falls off more sharply with distance than the reverberant sound. The loss of low intensity frequency components of a sound with increasing distance from the listener is modelled by mapping distance to the cut-off frequency of a low pass filter. The Doppler Effect is implemented using Csound code developed by Christopher Dobrian [Dobrian, 2008]. Each grain is sent to an interpolating delay line whose delay time is mapped to the distance of the source. The changing distance and delay time produces a pitch shift in the audio signal which mimics the Doppler Effect. Early reflections and global reverberation are applied to each grain using Csound code developed by Jan Jacob Hofmann [Hofmann, 2008] for an ambisonic spatialization instrument. Specular and diffuse reflections are generated according to the distance from the source to the listener, and to the specified dimensions of the virtual space. A global reverb is implemented using an eight delay line feedback delay network reverb with a feedback matrix based upon the physical modelling scattering junction of eight lossless waveguides of equal characteristic impedance [Hofmann, 2008; Smith, 1985]. Finally, all the audio signals are encoded using the Furse-Malham-Set (FMH) of encoding equations [Malham, 1999] for Second Order Ambisonics. The final output can be decoded for various loudspeaker configurations such as stereo, quadraphonic, octagon or a cube, using various ambisonic decoding schemes.

12.4 Case Study VII - *Rise*

An 8-channel composition, Rise, was realized using the three synthesis methods discussed in the preceding section. A number of spatial trajectories were recorded in Max/MSP and applied to the different synthesis techniques in various ways. These spatially dynamic textures were then contrasted with static decorrelated, monophonic signals played directly through the eight-channel array. Glisson synthesis is used to mimic birdsong and, indeed, the entire piece can be thought of as a sort of virtual soundscape with echoes of natural sounds and spaces.

12.4.1 Analysis

Two distinct trajectories were used throughout this piece;

- a fast spiralling trajectory with lots of vertical movement around the centre point
- a slow front-back-front trajectory passing through the centre point

The first trajectory was created specifically for use with glisson synthesis as it contains significant vertical movement. Elevation values are mapped to the frequency and direction of the glissando with the glissando start frequency restricted to linear harmonic intervals, producing a more melodic texture. This can be clearly heard in isolation at the very beginning of the piece. The second section begins with synthesized tones created using the grainlet additive synthesis instrument and the slow front-back trajectory. These synthesized tones provide the underlying harmony which is derived from a two chord guitar progression (a granulated version of which can be faintly heard later on). These strongly tonal textures are contrasted with wideband noise textures which also follow the same front-back trajectory. These rolling waves of noise were created using the granulation instrument and a heavily distorted version of the original guitar progression. The underlying tonality of these noise bursts becomes more apparent as the piece progresses and the grain duration increases. These contrasting textures are accompanied by reiterations of the opening glisson movement. As the glisson texture becomes increasingly dense and unifies as a high frequency tone it leads into a noisy texture presented as eight decorrelated signals played directly from the loudspeaker array. These static sources provide a sense of spatial perspective as their static spatial location contrast with the approaching and receding tonal and noise signals. The piece ends with a variation of the initial glisson texture, this time decreasing in elevation and density.

12.4.2 Critical Evaluation

While each specific granulation algorithm is quite distinct in terms of its sonic output, the spatialization algorithm results in some behaviour which is common to all of the synthesis methods. Dynamic variation in the density of the granular texture has a significant influence on the perceived timbre. At high densities the flocking behaviour manifests largely as variations in the spatial, spectral and amplitude characteristics of the overall texture, while at lower densities the individual voices and their spatial motion become more readily apparent. In other words, at higher densities a more unified tone is perceived while at lower densities, the signal tends to split into multiple distinct voices. The spatial distribution of the flock as a whole also has a significant influence on the perceived timbre. A flock with a strong centring instinct and a low minimum separation distance will be tightly bunched together spatially and when this is combined with a high density it results in more unified tone. The distance of the flock from the centre point is another highly important parameter with even very dense and tightly gathered flocks separating into individual voices as they pass through the central listening position.

13 Summary & Conclusion

> *"The trouble with space is that it's the whole piece. It's the sounds and everything. The impressions of space are created through the types of sounds and their temporal experience. Space is the whole thing. It is not usually something that people perceive as separate from the sounds themselves, although the composer might consider space separately. For the listener, they're all moulded into one. That's why we end up talking about the piece as a whole, because the whole is the space or spaces of the piece"*
> Denis Smalley [Smalley *et al*, 2000].

The above quote illustrates the vast range of issues related to the production, composition and performance of spatial music which this thesis has attempted to address. A composer of spatial music must consider specific aspects of the work such as the types of sounds and the particular spatialization process involved, while also remembering that many of these specific aspects may not be perceived directly by the listener. As Smalley states the piece as a whole must be looked at, because "the whole is the space or spaces of the piece". The success of any work of spatial music can therefore only be considered in terms of the overall compositional strategy which describes the relationship between space and every other musical parameter.

This thesis has attempted to examine all of these issues in terms of auditory perception as ultimately, the success of any work of spatial music depends upon the perception of the listener. A number of different spatialization schemes were analysed and a meta-analysis of the results of a large number of other listening tests and simulations were presented. The results indicate that each spatialization scheme has particular strengths and weaknesses, with the result that the most applicable technique in any situation is dependent on the particular spatial effect required.

13.1 Spatialization Results & Recommendations

The review of various spatialization schemes presented earlier in this thesis indicated the particular capabilities of each technique and that the most suitable spatialization scheme in any situation is entirely dependent on the particular context. What may be the most suitable approach in one case may not be the best approach in another, and this applies to the different types of sounds and spatial movements within a single composition as much as it applies to different musical aesthetics and performance practices.

The practice of stereo diffusion to a loudspeaker orchestra is clearly quite distinct from other approaches to spatialization. The real-time and manual nature of the diffusion process indicates the intuitive nature of this approach which is as much concerned with adapting the work to the acoustic of the particular performance space as it is with generating spatial gestures or locations. Technically speaking, the approach in this case is therefore relatively clear and well understood, and the technical and artistic practice of diffusion has already been covered in detail by composers and writers such as Jonty Harrison [Harrison, 1999], Denis Smalley [Smalley *et al*, 2000] and others [MacDonald, 1995].

The relatively new technique of wavefield synthesis would seem to also be quite distinct from multichannel techniques such as Ambisonics or VBAP, or the practice of diffusion. The spatial aliasing frequency is crucial in any WFS system and the results of a number of tests [Start, 1997; Huber, 2002; Wittek, 2007] illustrate the degradation in localization accuracy and timbre that occurs when this parameter is too low. However, due to the direct relationship between the spatial aliasing frequency and the physical gap between each loudspeaker, a very large number of loudspeakers will be required to cover a listening area suitable for a musical performance. One of the most often cited innovations of WFS is the ability to position virtual sources both behind and in front of the loudspeaker array. Numerous simulations have illustrated the correct wavefront curvature produced by such a source but the results of listening tests [Kerber *et al*, 2004; Wittek, 2007] suggest that in reality this effect does not provide a strong perceptual cue. This is particularly true if the listener's position is fixed or if the performance space contains significant early reflections and reverberation. The only scenario where the wavefront curvature produced by a WFS virtual source does seem to support the perception of distance is when the listener is free to move around the listening area. In this case the movement of the listener produces a sense of changing perspective or motion parallax due to correct curvature of the virtual source over an extended area, and this has been shown to support the perception of distance. While WFS is potentially highly suitable for certain applications, clearly more research is required to overcome these significant technical and logistical issues if WFS is to become more prevalent in performances of spatial music.

Multichannel spatialization schemes such as Ambisonics and stereophony would appear to be the most flexible approach, even though the performance of these techniques is significantly affected when the listening area is expanded to surround a group of listeners, or when significant early reflections and reverberation are present. Although the majority of tests carried out with these systems have focussed on a single listener in an acoustically treated space, these results do provide some indication of the minimum technical requirements of any performance system. Experience with the quadraphonic format revealed that a minimum of six loudspeakers is required to produce a consistent localization with amplitude panning [Thiele *et al*, 1976] and the results of various other tests indicate that a hexagonal loudspeaker arrangement is also generally preferred for Ambisonics [Benjamin *et al*, 2006; Guastavino *et al*, 2004]. If a minimum of six loudspeakers is required for a single listener then it is reasonable to assume that a greater number will be required for an expanded listening area. Eight loudspeakers would therefore seem to be a suitable minimum requirement and indeed, this particular arrangement has increasingly become the standard for performances of electroacoustic music in recent years. An octagonal array is particularly suitable as it allows for up to third order Ambisonics, can be readily implemented with most digital audio hardware, and provides eight distinct monophonic sources. Although this arrangement has become increasingly common, no clear preference has emerged for either of the two possible configurations shown in Figure 12.1. An octagonal layout with a centre-front channel (Figure 12.1 left) is perhaps preferable in the case of a rectangular venue, as potentially only channels four and eight would need to be repositioned. In addition, this layout can be more readily adapted for other schemes such as quad or 5.1. However, the alternative configuration (Figure 12.1 right) is perhaps more suitable for the presentation of multiple two-channel stereo tracks.

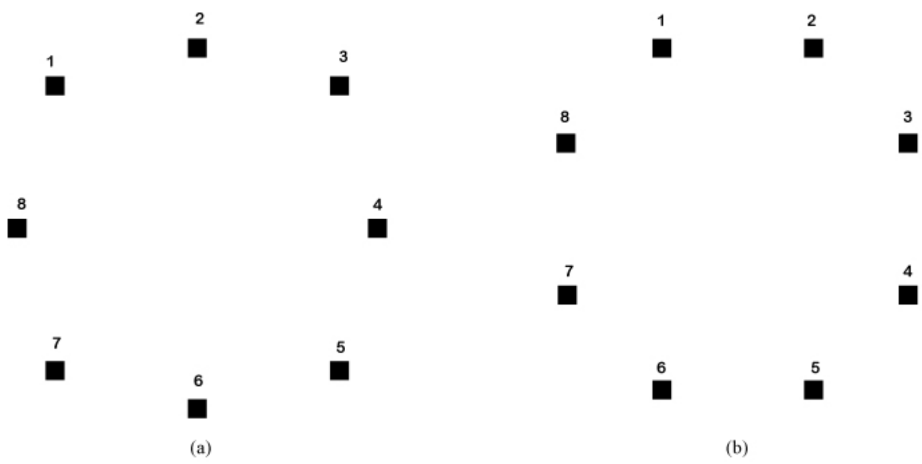

Fig. 13.1 Octagonal loudspeaker layout with centre front (a) and without (b)

Even with this increase in the number of loudspeakers, reliable localization is difficult to achieve for off-centre listeners with either stereophony or Ambisonics [Bates *et al*, 2007b]. As amplitude panning techniques will only ever use a maximum of two loudspeakers to produce a virtual source, this approach would seem to be preferable if horizontal localization is a critical parameter. Higher order Ambisonic decoding schemes which are optimized for extended listening areas (such as max-r_E) significantly reduce the number of contributing loudspeakers, however, stereophonic virtual sources can readily collapse to a single loudspeaker, which will provide the highest degree of localization accuracy possible. This feature is highly beneficial when the source position is static but introduces significant artefacts when the source is in motion as the number of contributing loudspeakers changes with the position of the virtual source. A number of different tests [Pulkki *et al*, 2005; Martin *et al*, 1999; Dickins *et al*, 1999] have shown that spatialization schemes which utilize a consistent number of loudspeakers, irrespective of source position, provide a much smoother trajectory than straight forward stereophonic panning. In addition to Ambisonics, a number of alternative amplitude panning techniques have been developed (VBAP for example) which can control the number of contributing loudspeakers independently of the source azimuth. This is clearly very similar to the Ambisonics approach of optimizing the energy vector r_E for all directions, at the cost of reducing the maximum localization accuracy that could be achieved at the loudspeaker positions. These techniques may be beneficial if a source must change from a static position at a single loudspeaker to a dynamic trajectory around the array, and are easier to implement for certain non-symmetrical loudspeaker arrays. Indeed amplitude panning algorithms such as these are in many ways analogous to a local ambisonic velocity decode whereby only the loudspeakers closest to the source direction are used, and the requirement to optimize the velocity component ($r_V = 1$) is dropped. However, in general, Ambisonics would seem to be more flexible, particularly in terms of loudspeaker layout, and there is some evidence to suggest that Ambisonics is still preferred for dynamically moving sources [Pernaux *et al*, 1998]. Other tests found that stereophony provided precise localization and good readability but a lack of immersion and envelopment, while Ambisonics provides an improved sense of immersion and envelopment but poor localization accuracy and readability of the scene [Gaustavino *et al*, 2007]. Consequently it would appear that stereophony is preferable when the directionality and focus of the virtual source is paramount, while

Ambisonics is preferable if a more diffuse enveloping sound field is required or to minimize panning artefacts in the case of dynamic sources. Additional distance cues such as the direct to reverberant ratio, air absorption and the Doppler effect can also improve dynamic trajectories and are relatively straightforward to implement in both stereophonic and Ambisonics systems. There is evidence to suggest that the addition of spatially accurate early reflections will improve the perception of distance but their effect on horizontal localization is not entirely clear, particularly for off-centre listeners. The results of tests carried out by Neher indicate that while a spatial early reflection pattern is important, the precise angle of each early refection is not hugely relevant [Neher, 2004].

The results presented in this section suggest an approach to spatialization which is optimized in terms of the perceptual capabilities of these different techniques. If a source must smoothly change from a static, tightly focussed position to a dynamic and smooth trajectory then an extended amplitude panning technique is perhaps preferable. However, if static and dynamic sources are to be contrasted than a combination of mono or stereophony (for the static sources) and Ambisonics (for dynamic sources) is preferable as the perceptible differences between these techniques will support the contrast between the two different types of sources.

13.2 Perception and Spatial Music Aesthetics

In the latter half of this thesis, a number of landmark works of spatial music were discussed and analysed in terms of the perceptual validity of their approach to spatialization. Although many differences exist between these different compositions, some general trends are evident. For example, a number of different composers used spatial distribution to improve the intelligibility of different layers of independent and potentially dissonant musical material; the "co-existence of dissimilars" described by John Cage. It is entirely possible that this spatial distribution of the performers was first implemented merely to facilitate the performance of overlapping yet unrelated musical layers at different tempi or metres. However, there is now a significant amount of scientific research to show that a listener's ability to detect and understand the content of multiple signals is indeed improved when the signals are spatially separated [Bregman, 1990; Shinn-Cunningham, 2003; Best, 2004]. Composers such as Charles Ives, Henry Brant and Karlheinz Stockhausen were clearly well aware of this fact, and they made extensive use of spatial distribution in their work for this reason.

The presentation of any work of spatial music to a distributed audience will necessarily result in certain members of the audience being situated outside the sweet spot. This will result in an unavoidable distortion in the spatial trajectory which is perceived by each member of the audience, and this distortion depends upon the position of the listener within the array. Moore suggested that the differences in perception among listeners is analogous to perspective distortion in photography or cinematography [Moore, 1983]. Although this is perhaps unsurprising, it raises significant questions about spatial music compositions which attempt to create and relate recognizable "sound shapes", such as in certain works by Iannis Xennakis for example. This form of spatial counterpoint, which McNabb describes as the motivic use of space assumes that spatial locations and trajectories are clearly and unambiguously perceived by every listener [McNabb, 1986]. However the results of the listening tests presented in this thesis show that this is extremely difficult to achieve. Even in the case of point sources which are clearly localized, each listener will be orientated differently with regards to the loudspeaker array, and so will have a different perspective on the spatial layout. In this case it is very hard to see how spatial motifs can be clearly and unambiguously perceived unless they are restricted to very rudimentary movements. In the words of the composer Henry Brant,

> *"Ideas of that kind seem to me more an expression of hope than of reality…. . It is hard enough to make the sounds do what you want "in sound" without saying that the sound should be shaped like a pretzel or something like that"* [Brant, 1967]

It is difficult now to reliably gauge the perceptual effectiveness of vast multimedia extravaganzas such as the Phillips Pavilion at Brussels in 1958 or the Japanese Pavilion at Osaka in 1970. This form of audiovisual spatial music, referred to by Xenakis as the polytype, contains a visual element which may support the auditory spatial trajectory. Although there is very little perceptual evidence to suggest that the abstract geometrical designs suggested by Varèse and Xenakis can be reliably achieved using sounds alone, this may be possible using a multimodal presentation such as the polytype.

Stockhausen's serialization of angular direction in *Kontakte* is an often cited example of an approach to spatialization which is not perceptually justified. While it is certainly true that the carefully calculated directions worked out by Stockhausen will not be precisely perceived by most, if not all the members of the audience, this is perhaps not as critical as it first appears. Spatial movement in *Kontakte* is controlled serially in the same way as rhythm and pitch. The precise angular location is not intended to be accurately perceived by each member of the audience, instead, a change in spatial location is used to indicate a certain temporal duration. So even though each listener will perceive a slightly different spatial movement, they will all perceive the same spatio-temporal movement. The abstract designs of these serialist composers differ therefore from the abstract graphical designs discussed earlier in that the specific trajectories and movements are not intended to be perceived by the audience. The following quote from Pierre Boulez represents a very different conception of spatial music to the highly visual form suggested by Xennakis and Varèse.

> *"However, I do not want to suggest that one thereby plays as in tennis or car racing, back and forth or in a cycle with everything following exactly. For me, sound mixtures are far more interesting; one feels that they rotate somehow, but one can't have the picture of an exactly comprehensible movement. With these different rates and with these different times, one gets an impression, which is much richer for me than a more normal circle or a loop to the left or right. For me, this is too anecdotal. "* [Boulez *et al*, 1988]

Abstract and complex spatial designs can perhaps therefore be effective when used indirectly in this fashion. This approach is somewhat reminiscent of other abstract processes which have been used to indirectly create and control complex textures in orchestral music. Consider this quote by the composer Gyorgy Ligeti from an interview in 1978.

> *"Technically speaking, I have always approached musical texture through part-writing. Both Atmospheres and Lontano have a dense canonic structure. But you cannot actually hear the polyphony, the canon. You hear a kind of impenetrable texture, something like a very densely woven cobweb. The polyphonic structure does not come through, you cannot hear it, it remains hidden in a microscopic, underwater world, to us inaudible* [Bernard, 1987]. "

The dynamic spatial textures created using granular synthesis and flocking algorithms would seem to function in the same way, as in this case only the overall motion and the motion of sounds relative to each other will be perceived, while the individual trajectories followed by each *grainboid* are not generally clearly perceptible. In this case, it is the overall motion, rather than the specific location or direction, which is important and hence, Ambisonics would seem to be a suitable spatialization scheme for this purpose.

In general, two distinct approaches to the use of space as a musical parameter have developed from the aesthetics of *Musique Concrète* and *Elektronische Musik* which were prevalent in the mid-twentieth century. This dichotomy of an abstract syntax or an abstracted syntax [Emmerson, 1986], an organic structure or an architectonic structure [Harrison, 1999], has been discussed extensively by composers and theorists for the past fifty years, and is also clearly evident in much of the spatial music discussed in this thesis. In terms of spatial music, an abstract syntax involves the application of predefined spatial effects to different sources which may have been created specifically for this purpose, such as the serialist compositions of Stockhausen for example. In music which follows an abstracted syntax, such as diffusion concerts and Smalley's spectromorphological compositions, a spatial strategy is derived from the sonic attributes (timbre, spectromorphological profile, etc) of the source signal. This form of spatial music would seem to engage much more directly with the perceptual difficulties that arise in presentations of spatial music to multiple listeners. Although diffusion is in many respects quite limited in what it can achieve spatially, it is admirably focussed on adapting each work for the particular performance space. Smalley's theory of spectromorphology has its origins in the practice of diffusion, yet this theory can equally be applied to other types of performances and aesthetics and a gestural approach would seem to be highly applicable when live instrumental performers and spatial audio are combined. The use of augmented instruments which map the actions of the performer to a spatialization algorithm would also seem to be very suitable for spatial electroacoustic performances, however, the necessity for specialized and expensive hardware is a significant obstacle. Multi-transducer devices such as the hexaphonic pickup are one possible solution to this problem as they may be non-destructively adapted for existing instruments.

13.3 Conclusion

This thesis has attempted to examine the various interrelated factors which influence the musical use of space in a performance context. It has been shown that the optimum spatialization technique in any situation is highly dependent on other factors, such as the overall compositional aesthetic. Various landmark works of spatial music have been assessed in terms of the perceptibility of the particular spatialization scheme adopted for that work, and also in terms of how the various spatial attributes involved were incorporated into the overall compositional framework. One strategy which has been adopted by many, quite different, composers is the use of spatial distribution to improve the intelligibility of multiple layers of independent musical material. In this instance, space is undoubtedly a fundamental aspect of the composition which is crucial if the work is to be perceived correctly. Although this is clearly an important aspect of spatial music, it is also quite rudimentary and straightforward in terms of the actual spatialization scheme adopted. Compositions which make use of more elaborate spatialization schemes are more difficult to assess, however, one feature which is common to many of the compositions discussed in this thesis is the way in which space is absorbed into the overall compositional aesthetic. This stands in marked contrast to surround mixes of popular music, an area which has received increased attention in recent years due to the emergence of the DVD and 5.1 surround sound as a consumer format. Despite the level of activity in this area, there is still little consensus as to why certain instruments or sounds should be positioned at certain locations. While the spatial distribution of parts will help to separate and clarify the different parts, it will also significantly alter the rhythmic and harmonic relationship between them, and this may or may not be of benefit, depending on the musical context.

The same cannot be said of such well regarded compositions such as *the Unanswered Question, Kontakte, Empty Vessels* or *Répons* as in these works, spatial effects are tightly integrated within the overall compositional strategy and are not simply applied for their own sake. This implies that if space is to be successfully used as a musical parameter, it must be supported by other aspects of the work. This conception of space is difficult to reconcile with suggestions that space can become "the primary carrier of meaning" [Henriksen, 2002] in a musical composition or that spatial structures can become more significant than the actual sounds themselves. The composer Johannes Goebel suggests in the following quote that space can function in a similar fashion to dynamic differences in volume which seems to agree with many of conclusions of this thesis.

> *"In my opinion the spatial placement of sounds, whether instrumental or electronically, has about the same potential for aesthetic differentiation as loudness. Compared to pitch and timbre, localization yields far less potential for aesthetic differentiation, but on the other hand no one would deny that in quite a few pieces loudness is an important, highly sophisticated, composed part of music. And the same could be said for the distribution of sound in space. "*
> [Goebel, 2001]

The artistic control of volume can undoubtedly add expressiveness to a musical performance and the importance of dynamics has long been an essential part of music composition. In the opinion of the author, this is equally true for the musical use of space, and this is particularly true in the case of electronic music. Performances of electronic music often lack the visual component of traditional instrumental music performances. During an instrumental performance visible actions by the performer produces an audible result and it is this multi-modal feedback process that makes a performance “live”. The dynamic spatialization of electronic sounds, particularly the gestural/environmental structuring principle suggested by Denis Smalley, can potentially replace or substitute for this missing visual component and provide a necessary and potentially thrilling physicality.

14 List of Publications

P1 Bates, E. & Furlong, D., *"Score File Generators for Boids Based Granular Synthesis in Csound"*, Proceedings of the 126th Audio Engineering Society Convention, Munich, Germany, May 2009.

P2 Bates, E., Furlong, D. & Dennehy, D. *"Adapting Polyphonic Pickup Technology for Spatial Music Performance",* Proceedings of the International Computer Music Conference, Belfast, Northern Ireland, August 2008.

P3 Bates, E., Kearney, G., Furlong, D. & Boland, F. *"Localization Accuracy of Advanced Spatialization Techniques in Small Concert Halls"*, Proceedings of the 153rd Meeting of the Acoustical Society of America, Salt Lake City, USA, June 2007.

P4 Bates, E., Kearney, G., Furlong, D. & Boland, F. *"Monophonic Source Localization for a Distributed Audience in a Small Concert Hall",* Proceedings of the 10th International Conference on Digital Audio Effects, Bordeaux, France, September 2007.

15 Summary of Published Articles

P1 In this paper we present a set of score file generators and granular synthesis instruments for the Csound language. The applications use spatial data generated by the Boids flocking algorithm along with various user-defined values to generate score files for grainlet additive synthesis, granulation and glisson synthesis instruments. Spatialization is accomplished using Higher Order Ambisonics and distance effects are modelled using the Doppler Effect, early reflections and global reverberation. The sonic quality of each synthesis method is assessed and an original composition by the author is presented.

P2 This paper describes how polyphonic pickup technology can be adapted for the spatialization of electric stringed instruments such as the violin, cello and guitar. It is proposed that mapping the individual strings to different spatial locations integrates the spatial diffusion process with the standard musical gestures of the performer. The development of polyphonic guitar processing is discussed and a method of adapting MIDI guitar technology for this purpose is presented. The compositional and technical strategies used with various augmented instruments is presented along with an analysis of three compositions by the author for spatialized hexaphonic guitar.

P3 A comparison of spatialization schemes is presented in terms of their localization accuracy under the non-ideal listening conditions found in small concert halls. Of interest is the effect of real reverberant conditions, non-central listening positions and non-circular speaker arrays on source localization. The data is presented by comparison of empirical binaural measurements to perceptual listening tests carried out using Ambisonics, Vector Base Amplitude Panning (VBAP), Spat (with B-format encoding) and Delta Stereophony (DSS) systems. The listening tests are conducted by comparing the localization of phantom sources generated by the spatialization systems, to monophonic sources generated by reference loudspeakers. The reference and phantom sources are presented at front, side and back locations about a 9 listener audience, and the systems are tested in a random order with a calibrated 16 loudspeaker array situated around the audience area. The binaural recordings are compared to the subjective measurements of localization accuracy through the inter-aural time difference (ITD) cues at each listener position.

P4 The transfer of multichannel spatialization schemes from the studio to the concert hall presents numerous challenges to the contemporary spatial music composer or engineer. The presence of a reverberant listening environment coupled with a distributed audience are significant factors in the presentation of multichannel spatial music. This paper presents a review of the existing research on the localization performance of various spatialization techniques and their ability to cater for a distributed audience. As the first step in a major comparative study of such techniques, the results of listening tests for monophonic source localization for a distributed audience in a reverberant space are presented. These results provide a measure of the best possible performance that can be expected from any spatialization technique under similar conditions.

Bibliography

[Ahrens *et al*, 2009] Ahrens, J. and Spors, S., *"Spatial Encoding and Decoding of Focussed Virtual Sound Sources"*, Graz Ambisonics Symposium, 2009.

[Ambisonics Standards, 2009] *"Analysis of Normalisation Schemes: A supporting document for the "Universal Ambisonic Standard"*, Ambisonics Google Groups, http://docs.google.com/Doc?id=df4dtw69_66d4khc7fb&hl=en

[Asano *et al*, 1990] Asano, F., Suzuki, Y. and Sone, T., *"Role of spectral cues in median plane localization",* Journal of the Acoustical Society of America, Vol. 88, pp. 159-168, 1990.

[Baalman, 2006] Baalman, M., "*swonder3Dq: Auralisation of 3D objects with Wave Field Synthesis",* Proceedings of the Linux Audio Conference, 2006.

[Ballan *et al*, 1994] Ballan, O., Mozzoni, L. and Rocchesso, D., *"Sound spatialization in real time by first reflection simulation",* Proceedings of the International Computer Music Conference, pp. 475-476, 1994.

[Ballas *et al*, 2001] Ballas, J. A., Brock, D., Stroup, J. and Fouad, H., "*The Effect of Auditory Rendering on Perceived Movement: Loudspeaker Density and HRTF",* Proceedings of the 2001 International Conference on Auditory Display, 2001.

[Bamford, 1995], Bamford, J. S. "*An Analysis of Ambisonic Sound Systems of First and Second Order"*, Doctoral Thesis, University of Waterloo, Canada, 1995.

[Barrett, 1990] Barrett, N., *"Spatio-musical Composition Stratagies"*, Organized Sound, Vol. 7(3), 2003.

[Barrett *et al*, 2007] Barrett, N. and Otondo, F. *"Creating Sonic Spaces: an interview with Natasha Barrett"*, Computer Music Journal, Vol. 31(2), pp. 10-19, 2007.

[Barron *et al*, 1981] Barron, M. and Marshall, A. H., "*Spatial Impression due to*

Early Lateral Reflections in Concert Halls: The Derivation of a Physical Measure", Journal of Sound and Vibration, Vol. 77(2), pp. 211-232, 1981.

[Bartetzki, 2008] Bartetzki, A., *"Csound score generation and granular synthesis with cmask"* www.kgw.tuberlin.de/abart/cmaskpaper/cmaskarticle.html, retrieved on 07-07-2008.

[Bates *et al*, 2007a] Bates, E., Kearney, G., Boland, F. and Furlong, D., "*Monophonic Source Localization for a Distributed Audience in a Small Concert Hall*", 10th International Conference on Digital Audio Effects (DAFX-07), 2007.

[Bates *et al*, 2007b] Bates, E., Kearney, G., Boland, F. and Furlong, D., *"Localization Accuracy of Advanced Spatialization Techniques in Small Concert Halls",* 153rd Meeting of the Acoustical Society of America, 2007.

[Bauer, 1962] Bauer, B. B., *"Phasor Analysis of Some Stereophonic Phenomena"*, The Journal of the Acoustical Society of America, Vol. 33(11), pp. 1536-1962.

[Begault, 1992] Begault, D., *"Perceptual Effects of Synthetic Reverberation on Three-Dimensional Audio Systems"*, Journal of the Audio Engineering Society, Vol. 40(11), 1992.

[Begault, 1994] Begault, D. R., "*3-D Sound for Virtual Reality and Multimedia*", Academic Press Inc, pp. 45-46; pp. 51-56, 1994.

[Benjamin *et al*, 2006] Benjamin, E. M., Lee, R. and Heller, A. J., *"Localization in Horizontal-Only Ambisonic Systems"*, Proceedings of the 121st Convention of the Audio Engineering Society, 2006.

[Beranek, 1996], Beranek, L., *"Concert and opera halls. How they sound",* A.I.P. Press, 1996.

[Berg, 2007] Berg, J., *"Midi breakout box",* www.unfretted.com, 2007, retrieved on 01/01/2009.

[Berkhout, 1998] Berkhout, A. J., *"A Wavefield Approach to Multichannel Sound"*, Proceedings of the 104th Convention of the Audio Engineering Society, Preprint 4749, 1998.

[Bernard, 1987] Bernard, J. W., *"Inaudible Structures, Audible Music: Ligeti's Problem, and His Solution"*, Music Analysis, Vol. 6(3), pp. 207-236, 1987.

[Bertet et al, 2009] Bertet, S., Daniel, J., Parizet, E. and Warusfel, O, *"Influence of Microphone and Loudspeaker Setup on Perceived Higher Order Ambisonics Reproduced Sound Field",* Proceedings of the Graz Ambisonics Symposium, 2009.

[Best, 2004] Best, V. A., *"Spatial Hearing with Simultaneous Sound Sources: A Psychophysical Investigation",* Doctoral Thesis, University of Sydney, 2004.

[Bisig *et al*, 2008] Bisig, D., Neukom, M. and Flury, J. *"Interactive Swarm Orchestra - An Artificial Life Approach to Computer Music"*, Proceedings of the International Computer Music Conference, 2008.

[Blackwell *et al*, 2004] Blackwell, T. and Young, M., *"Self-organised music"*, Organized Sound, Vol. 9, pp. 123–136, 2004.

[Blauert, 1997] Blauert, J., *"Spatial Hearing: The Psychophysics of Human Sound Localisation"*, pp. 3-4, pp. 38-39; pp. 47-48 pp. 118-119, pp. 140-154, pp173-190;pp. 348-358; MIT Press; 1997.

[Boone *et al,* 1995] Boone, M. M., Verheijen, E. N. G. and van Tol, P. F., *"Spatial sound field reproduction by wave field synthesis"*, Journal of the Audio Engineering Society, Vol. 43(12), pp. 1003-1012, 1995.

[Boone *et al*, 1999] Boone, M. M., Horbach, U. and de Bruijn, W. P. J., *"Virtual Surround Speakers with Wave Field Synthesis"*, Proceedings of the 106th Convention of the Audio Engineering Society, Preprint 4928, 1999.

[Boone *et al*, 2003] Boone, M. M. and de Bruijn, W., *"Improved speech intelligibility in teleconferencing using Wave Field Synthesis"*, Proceedings of the 114th Convention of the Audio Engineering Society, Preprint 5800, 2003.

[Boulanger, 1986] Boulanger, R., *"Toward a new age of performance: Reading the book of dreams with the Mathews electronic violin"*, Perspectives of New Music, Vol. 24(2), pp. 130–155, 1986.

[Boulanger, 2000] Boulanger, R., *"The Csound book"*, MIT Press, 2000.

[Boulez *et al*, 1988] Boulez, P. and Gerzso, A., *"Computers in Music"*, Scientific American, Vol. 258(4), 1988.

[Brant, 1967] Brant, H., "*Space as an essential aspect of musical composition",* Contemporary Composers on Contemporary Music, Holt, Rinehart and Winston, pp. 221-242, 1967.

[Bregman, 1990] Bregman, A. S., *"Auditory Scene Analysis"*, 196-200, pp. 204-250; pp. 455-490, MIT Press; 1990.

[Bromberger, 2007] Bromberger, E., *"Berlioz 'Requiem' - San Diego Symphony Orchestra Program Notes"*, 2007, www.sandiegosymphony.com/uploads/pdfs/pdfmw060713.pdf, retrieved on 06/06/2009.

[Bronkhurst, 2002] Bronkhorst, A. W., *"Modeling auditory distance perception in rooms"*, Proceedings of the EAA Forum Acusticum Sevilla, 2002.

[Brungart *et al*, 1999] Brungart, D. S. and Rabinowitz, W. M., *"Auditory localization of nearby sources. Head-related transfer functions"*, Journal of the Acoustical Society of America, Vol. 106(3), 1999.

[Bryant, 1981] Bryant, D., "*The 'cori spezzati' of St Mark's: myth and reality"*, Early Music History, Vol. 1, pp. 165-186, 1981.

[Cage, 1957] Cage, J., *"Experimental Music",* Address given to the convention of Music Teachers National Association, Chicago, 1957. Reprinted in Silence, MIT Press, 1961.

[Capra *et al*, 2007] Capra, A., Fontana, S., Adriaensen, F., Farina, A. and Grenier, Y., *"Listening Tests of the Localization Performance of Stereodipole and Ambisonic Systems",* Proceedings of the 123rd Convention of the Audio Engineering Society, Preprint 7187, 2007.

[Caulkins *et al*, 2003] Caulkins, T., Corteel, E. and Warusfel, O., *"Wave Field Synthesis Interaction with the Listening Environment, Improvements in the Reproduction of Virtual Sources Situated inside the Listening Room"*, Proceedings of the 6th International Conference on Digital Audio Effects (DAFX-03), 2003.

[Chowning, 1971], Chowning, J., *"The simulation of moving sound sources",* Journal of the Audio Engineering Society, Vol. 19, pp. 2–6, 1971.

[Clark *et al*, 1958] Clark, H. A. M., Dutton, G. F., Vanderlyn, P. B., *"The 'Stereosonic' Recording and Reproducing System: A Two-Channel Systems for Domestic Tape Records"*, Journal of the Audio Engineering Society, Vol. 6(2), pp. 102-117, 1958.

[Corteel *et al*, 2003] Corteel, E. and Nicol, R., *"Listening room compensation for Wave Field Synthesis. What can be done?"*, Proceedings of the 23rd International Conference of the Audio Engineering Society, 2003.

[Corteel *et al*, 2004] Corteel E., Caulkins T., Warusfel O., Pellegrini R., Rosenthal M. and Kuhn C., "*Sound Scene Creation using Wave Field Synthesis"*, Actes Journées du Design Sonore, 2004.

[Corteel *et al*, 2008] Corteel, E., Pellegrini, R. and Kuhn-Rahloff, C., *"Wave Field Synthesis with increased aliasing frequency",* Proceedings of the 124th Convention of the Audio Engineering Society, Preprint 7362, 2008.

[Cott, 1973], Cott, J., "Stockhausen; conversations with the composer", Simon and Schuster, 1973.

[Daniel *et al*, 1998] Daniel, J., Rault, J. and Polack, J. D., *"Ambisonics Encoding of Other Audio Formats for Multiple Listening Conditions"*, Proceedings of the 105th Convention of the Audio Engineering Society, Preprint 4795, 1998.

[Daniel *et al,* 1999], Daniel, J., Nicol, R. and Moreau, S, *"Further Investigations of High Order Ambisonics and Wavefield Synthesis for Holophonic Sound Imaging",* Proceedings of the 114th Convention of the Audio Engineering Society, Preprint 5788, 1999.

[Daniel, 2000] Daniel, J., *"Représentation de champs acoustiques, application à la transmission et à la reproduction de scènes sonores complexes dans un contexte multimédia",* Doctoral thesis, Université Paris, 2000.

[Daniel, 2003] Daniel, J., *"Spatial Sound Encoding Including Near Field Effect: Introducing Distance Coding Filters and a Viable, New Ambisonic Format"*, Proceedings of the 23rd International Conference of the Audio Engineering Society, 2003.

[de Vries, 1995] de Vries, D., *"Sound enhancement by wave field synthesis: adaption of the synthesis operator to the loudspeaker directivity characteristics"*, Proceedings of the 98th Convention of the Audio Engineering Society, Preprint 3971, 1995.

[Dickins *et al*, 1999] Dickins, D., Flax, M., Mckeag, A. and Mcgrath, D., *"Optimal 3D speaker panning"*, Proceedings of the 16th International Convention of the Audio Engineering Society, 1999.

[Dobrian, 2008] Dobrian, C., *"Doppler shift in a stereo Csound instrument*", http://music.arts.uci.edu/dobrian/dsp2001/csound/dopplertheory.htm, retrieved on 07/07/2008.

[Elfner *et al*, 1968] Elfner, L. F. and Tomsic, R. T., *“Temporal and intensive factors in binaural lateralization of auditory transients”,* Journal of the Acoustical Society of America, Vol. 43, pp. 747-751, 1968.

[Emmerson, 1986] Emmerson, S., *“The relation of language to materials”*, The Language of Electroacoustic Music, pp. 20-24, Macmillan Press, 1986.

[Essl, 1995] Essl, K., *“Lexikon-Sonate. An Interactive Realtime Composition for Computer-Controlled Piano”,* Proceedings of the 2nd Brazilian Symposium on Computer Music, 1995.

[Farina *et al*, 1998] Farina A. and Ugolotti E., *“Software Implementation of B-Format Encoding and Decoding”*, Proceedings of the 104th International AES Convention, 1998.

[Farina *et al*, 2000] Farina, A., Zanolin, M. and Crema, E., *“Measurement of sound scattering properties of diffusing panels through the Wave Field Synthesis approach”,* Proceedings of the 108th International Convention of the Audio Engineering Society, 2000.

[Felder, 1977] Felder, D., “*An interview with Karlheinz Stockhausen”*, Perspectives of New Music, Vol. 16(1), pp. 85–101, 1977.

[Fellget, 1975] Fellgett, P., *“Ambisonics. Part One: General System Description”*, Studio Sound, Vol.17, pp.20–22;40, 1975.

[Fels, 1996] Fels, P., *“20 Years Delta Stereophony System - High Quality Sound Design”*, Proceedings of the 100th Convention of the Audio Engineering Society, Preprint 4188, 1996.

[Franco *et al*, 2004] Franco, A. F., Merchel, S., Pesqueux, L., Rouaud, M. and Soerensen, M. O., *"Sound Reproduction by Wave Field Synthesis"*, Project Report, Aalborg University, 2004.

[Frank *et al,* 2008] Frank, M., Zotter, F. and Sontacchi, A., *"Localization Experiments Using Different 2D Ambisonics Decoders"*, 25th Tonmeistertagung – VDT International Convention, 2008.

[Franssen, 1962] Franssen, N. V., *"Stereophony",* Philips Technical Library, The Netherlands, English translation (1964).

[Furlong *et al*, 1992] Furlong, D. J. and MacCabe, C. J., *"Loudspeakers, Listening Rooms, and Effective Synthetic Auralization"*, Proceedings of the 93rd Convention of the Audio Engineering Society, Preprint 3445, 1992.

[Furse, 2000] Furse, R., *"First and Second Order Ambisonic Decoding Equations",* http://www.muse.demon.co.uk/ref/speakers.html, retrieved on 03/06/2009.

[Gardner, 1962] Gardner, M. B., *"Binaural Detection of Single- Frequency Signals in Presence of Noise"*, Journal of the Acoustical Society of America, Vol. 34, pp. 1824-1830, 1962.

[Gayou, 2007] Gayou, E., *"The GRM: landmarks on a historic route",* Organised Sound, Vol. 12(3), pp. 203–211, Cambridge University Press, 2007.

[Guastavino *et al*, 2004] Guastavino, C and Katz, B. F. G., *"Perceptual evaluation of multi-dimensional spatial audio reproduction"*, Journal of the Acoustical Society of America, Vol. 116(2), 2004.

[Guastavino *et al*, 2007] Guastavino, C., Larcher, V., Catusseau, G. and Boussard, P., "*Spatial Audio Quality Evaluation: Comparing Transaural, Ambisonics and Stereo*", Proceedings of the 13th International Conference on Auditory Display, 2007.

[Gerzon, 1974a] Gerzon, M. A., *"Sound Reproduction Systems",* Patent No. 1494751.

[Gerzon, 1974b] Gerzon, M. A., *"Surround Sound Psychoacoustics"*, Wireless World, Vol. 80, pp. 483-486, 1974.

[Gerzon, 1992a], Gerzon, M. A., *“General Methatheory of Auditory Localisation”*, Proceedings of the 92nd International Audio Engineering Society, Preprint 3306, 1992.

[Gerzon, 1992b] Gerzon, M. A., *"The Design of Distance Panpots",* Proceedings of the 92nd Audio Engineering Society Convention, Preprint 3308, 1992.

[Gerzon, 1992c], Gerzon M. A., *“Psychoacoustic Decoders for Multispeaker Stereo and Surround Sound”*, Proceedings of the 93rd International Audio Engineering Society Convention, Preprint 3406, 1992.

[Goebel, 2001] Goebel, J., “Letters”, Computer Music Journal, Vol. 25(4), pp. 5-11, MIT Press, 2001.

[Griesinger, 2009], Griesinger, D., *“The importance of the direct to reverberant ratio in the perception of distance, localization, clarity, and envelopment (auditory engagement)”*, Proceedings of the 126th Audio Engineering Society Convention, 2009.

[Gröhn , 2002] Gröhn, M., *“Localization of a Moving Virtual Sound Source in a Virtual Room, the Effect of a Distracting Auditory Stimulus”,* Proceedings of the 2002 International Conference on Auditory Display, 2002.

[Harley, 1997] Harley, M. A., *“An American in Space: Henry Brant's "Spatial Music"*, American Music, Vol. 15(1), pp. 70-92, 1997.

[Harley, 1998a] Harley, M. A., *“Spatiality of Sound and Stream Segregation in 20th-Century Instrumental Music”,* Organized Sound, Vol. 3(2), pp. 147-166, 1998.

[Harley, 1998b] Harley, M. A., *“Music of Sound and Light: Xenakis's Polytopes”,* Leonardo, The MIT Press, Vol. 31(1), pp. 55-65, 1998.

[Harris *et al*, 1998] Harris, N. and Flanagan, S., *"Stereophonic Localization in Rooms, Comparing Conventional and Distributed-Mode Loudspeakers",* Proceedings of the 105th Convention of the Audio Engineering Society, Preprint 4794, 1998.

[Harrison, 1999] Harrison, J., *"Sound, space, sculpture: some thoughts on the 'what', 'how' and 'why' of sound diffusion"*, Organised Sound, Vol. 3(2), pp. 117–27, Cambridge University Press, 1999.

[Hartmann, 1983] Hartmann, W. M., "*Localization of Sound in Rooms"*, Journal of the Acoustical Society of America, Vol. 74, pp. 1380-1391, 1983.

[Hartmann *et al*, 1985] Hartmann, W. M. and Rackerd, B., "*Localization of Sound in Rooms II: The Effects of a Single Reflecting Surface"*, Journal of the Acoustical Society of America, Vol. 78, pp524-533, 1985.

[Hartmann *et al*, 1989] Hartmann, W. M. and Rackerd, B., *"Localization of sound in rooms IV: The Franssen effect"*, Journal of the Acoustical Society of America, Vol. 86(4), pp. 1366-1373, 1989.

[Häusler *et al*, 1985] Häusler, J. and Boulez, P., *"Pierre Boulez on Répons",* Teilton, Schriftenreihe, der Heinrich-Strobel-Stiftung des Südwestfunks, Vol. 4, pp. 7-14, Kassel, Bärenreiter, translated from the German by Clemens Merkel and Christina Unger, 1985.

[Henriksen, 2002] Henriksen, F. E., *"Space in Electroacoustic Music: Composition, Performance and Perception of Musical Space",* Doctoral thesis, City University, London, 2002.

[Hodgkinson, 2001] Hodgkinson, T., *"An Interview with Pierre Schaeffer*", The Book of Music & Nature, Wesleyan University Press, pp. 34-44, 2001.

[Hoffman, 2001] Hoffmann, P., *"Iannis Xenakis"*, The New Grove Dictionary of Music and Musicians, Vol. 27, pp. 607-608, Macmillan Press, 2001.

[Hofmann, 2008] Hofmann, J. J., *"Ambisonics tutorial"*, http://www.sonicarchitecture.de, retrieved on 07/07/2008.

[Howard *et al*, 1996] Howard, D. M. and Angus, J., *"Acoustics and Psychoacoustics",* pp. 257-262, Focal Press, 1996.

[Huber, 2002] Huber, T.*, "Zur Lokalisation akustischer Objekte bei Wellenfeldsynthese"*, M.Phil thesis, IRT/ TU Munich, 2002, www.hauptmikrofon.de/wfs.html, retrieved on 06/06/2009.

[Hubel *et al*, 1959] Hubel, D. H. and Wiesel, T. N., *"Receptive fields of single neurons in the cats striate cortex"*, Journal of Physiology, Vol. 148, pp. 547-579, 1959.

[Jehan *et al*, 1999] Jehan, A., Freed, T. and Dudas, R., *"Musical applications of new filter extensions to max/msp"*, Proceedings of the International Computer Music Conference, pp. 504–507, 1999.

[Johnson, 2002] Johnson, M. E., *"Charles Ives's (Utopian, Pragmatist, Nostalgic, Progressive, Romantic, Modernist) Yankee Realism",* American Music, Vol. 20(2), pp. 188-233, University of Illinois Press, 2002.

[Jot, 1999] Jot, J., *"Real-time spatial processing of sounds for music, multimedia and interactive human-computer interfaces"*, Multimedia Systems, Vol. 7(1), pp. 55-69, 1999.

[Jot *et al*, 1999] Jot, J., Larcher, V. and Pernaux, J. M., *"A Comparative Study of 3-D Audio Encoding and Rendering Techniques",* Proceedings of the 16th International Conference of the Audio Engineering Society, 1999.

[Kendall *et al*, 1984] Kendall, G. S. and Martens, W. L., *"Simulating the Cues of Spatial Hearing in Natural Environments"*, Proceedings of the 1984 International Computer Music Conference, 1984.

[Kerber *et al*, 2004] Kerber, S., Wittek, H., Fastl, H. and Theile, G., *"Experimental investigations into the distance perception of nearby sound sources: Real vs. WFS virtual nearby sources"*, Joint Meeting of the German and the French Acoustical Societies (CFA/DAGA), 2004.

[Komiyama *et al,* 1991] Komiyama, S., Morita, A., Kurozumi, K. and Nakabayashi, K., *"Distance control system for a sound image''*, Proceedings of the 9th International Conference of the Audio Engineering Society, pp. 233–239, 1991.

[Kratschmer *et al*, 2009] Kratschmer, M. and Rabenstein, R., *"Implementing Ambisonics on a 48 Channel Circular Loudspeaker Array",* Proceedings of the Graz Ambisonics Symposium, 2009.

[Ligeti, 1958] Ligeti, G., *"Metamorphoses of Musical Form"*, Die Reihe, Vol. 7, pp. 5-19, 1958.

[Litovsky *et al*, 1999], Litovsky, R.Y., Colburn, H. S, Yost, W. A., Guzman, S. J., "*The Precedence Effect"*, Journal of the Acoustical Society of America, Vol. 106(4), pp. 1633-1654, 1999.

[Lund, 2000] Lund, T., *"Enhanced Localization in* 5.1 *Production",* Proceedings of the 105th Convention of the Audio Engineering Society, Preprint 5243, 2000.

[MacDonald, 1995] MacDonald, A., *"Performance Practice in the Presentation of Electroacoustic Music",* Computer Music Journal, Vol. 19(4), pp. 88-92, 1995.

[Machover, 1992] Machover, T., *"Hyperinstruments - MIT Media Lab Research Report"*, 1992.

[MacPherson, 2002] MacPherson, E. A. and Middlebrooks, J. C., "*Listener weighting of cues for lateral angle: The Duplex Theory of sound localization revisited*", Journal of the Acoustical Society of America, Vol. 111, pp. 2219–2236, May 2002.

[Makita, 1962], *"On the Directional Localization of Sound in the Stereophonic Sound Field"*, E. B. U. Review, part A no. 73, pp. 102-108, 1962.

[Malham, 1992] Malham, D. G., *"Experience with large area 3-D Ambisonic sound systems"*, Proceedings of the Institute of Acoustics, Autumn Conference on Reproduced Sound, Vol. 8, 1992.

[Malham *et al,* 1995], Malham, D. G. and Myatt, A. *"3-D Sound Spatialization using Ambisonic Techniques",* Computer Music Journal, Vol. 19(4), pp. 58-70, 1995.

[Malham, 1998] Malham, D. G. *"Sound Spatialisation",* Proceedings of the First COST-G6 Workshop on Digital Audio Effects (DAFX98), pp. 62-70, 1998.

[Malham, 1999] Malham, D. G., *"Higher order Ambisonic systems"*, abstracted from "Space in Music - Music in Space", M. Phil Thesis, University of York, 2003.

[Marentakis *et al*, 2008] Marentakis, G., Corteel, E. and McAdams, S., *"Wave Field Synthesis Evaluation using the Minimum Audible Angle in a Concert Hall",* Proceedings of the 124th Convention of the Audio Engineering Society, Preprint 7360, 2008.

[Martin *et al*, 1999a] Martin, G., Woszczyk, W., Corey, J. and Quesnel, R., *"Sound source localization in a five channel surround sound reproduction system"*, Proceedings of the 104th Convention of the Audio Engineering Society, Preprint 4994, 1999.

[Martin *et al*, 1999b] Martin, G., Woszczyk, W., Corey, J. and Quesnel, R., *"Controlling Phantom Image Focus in a Multichannel Reproduction System"*, Proceedings of the 107th Conference of the Audio Engineering Society, Preprint 4996, 1999.

[Martin *et al*, 2001] Martin, G., Depalle, P., Woszczyk, W., Corey, J. and Quesnel, R., *"A Hybrid Model for Simulating Diffused First Reflections in Two-Dimensional*

Synthetic Acoustic Environments", Proceedings of the 19th Conference of the Audio Engineering Society, 2001.

[Martin, 2005] Martin, A., *"Algorithms for Sonification: Smoke, Boids and the Game of Life"*, M. Phil Thesis, Trinity College Dublin, 2005.

[McFarlane, 2001] Mc Farlane, M. W., *"The Development of Acousmatics in Montréal"*, Journal of the Canadian Electroacoustic Community, eContact!, Vol. 6(2), 2001.

[McNabb, 1986] McNabb, M. *"Computer Music: Some Aesthetic Considerations"*, The Language of ElectroAcoustic Music, pp. 146-149, Macmillan Press, 1986.

[Goodale *et al*, 1992] Goodale, M. A. and Milner, A. D. *"Separate visual pathways for perception and action"*, Trends in Neurosciences, Vol. 15(1), pp. 20-25, 1992.

[Mershon *et al*, 1975] Mershon, D. H. and King, L. E. *"Intensity and reverberation as factors in the auditory perception of egocentric distance"*, Perception & Psychophysics, Vol. 18, pp. 409-415, 1975.

[Mershon, 1997] Mershon, D. H. *"Phenomenal geometry and the measurement of perceived auditory distance"*, Chapter 13 in *Binaural and Spatial Hearing in Real and Virtual Environments*, Lawrence Erlbaum Assoc. pp. 257-274, 1997.

[Meyer-Eppler, 1949] Meyer-Eppler, W., *"Elektronische Klangerzeugung: Elektronische Musik und synthetische Sprache"*, Bonn: Ferdinand Dümmlers, 1949.

[Michelsen *et al*, 1997] Michelsen, J. and Rubak, P., *"Parameters of Distance Perception in Stereo Loudspeaker Scenario"*, Proceedings of the 102nd Convention of the Audio Engineering Society, Preprint 4472, 1997.

[Misch *et al*, 1998] Misch, L., Hentschel, F. and Kohl, J. *"On the Serial Shaping of Stockhausen's Gruppen für drei Orchester"*, Perspectives of New Music, Vol. 36(1), pp. 143-187, 1998.

[Moore, 1983], Moore, F. R., *"A general model for spatial processing of sounds"*, Computer Music Journal, Vol. 7, pp. 6–15, 1983.

[Moore *et al*, 2007] Moore, J. D. and Wakefield, J. P., *"Surround Sound for Large Audiences: What are the Problems?"*, Annual University of Huddersfield Researchers' Conference. 2007.

[Moorer, 1979] Moorer, J. A., *"About This Reverberation Business"*, Computer Music Journal, Vol. 3(2), pp. 13-28, 1979.

[Moreau *et al*, 2006] Moreau, S., Daniel, J. and Bertet, S., *"3D Sound Field Recording with Higher Order Ambisonics – Objective Measurements and Validation of a 4th Order Spherical Microphone"*, Proceedings of the 120th Convention of the Audio Engineering Society, 2006.

[Morgan, 1978] Morgan, R. P., *"Ives and Mahler: Mutual Responses at the End of an Era"*, 19th-Century Music, Vol. 2(1), pp. 72-81, University of California Press, 1978.

[Morgan, 1991] Morgan, R. P., *"Stockhausen's Writings on Music"*, The Musical Quarterly, Vol. 75(4), pp. 194-206, Oxford University Press, 1991.

[Moritz, 2002] Moritz, A., *"Stockhausen: Essays on the Works"*, http://home.earthlink.net/~almoritz/, retrieved on 01/05/2009.

[Mortenson, 1987] Mortenson, G. C., *"Father to Son: The Education of Charles Ives"*, Music Educators Journal, Vol. 73(7), pp. 33-37, 1987.

[Neher, 2004] Neher, T., "*Towards A Spatial Ear Trainer*", Doctoral Thesis, University of Surrey, pp. 126-164, 2004.

[Neukon *et al*, 2008] Neukom, M. and Schacher, J. C., *"Ambisonics Equivalent Panning"*, Proceedings of the International Computer Music Conference, 2008.

[Nicol *et al,* 1999], Nicol, R. and Emerit, M., "*3-D-sound reproduction over an extensive listening area: A hybrid method derived from holophony and ambisonic*," Proceedings of the 16th International Conference of the Audio Engineering Society, 1999.
[Nielsen, 1993] Nielsen, S. H., *"Auditory distance perception in different rooms",* Journal of the Audio Engineering Society, Vol. 41, pp. 755-770, 1993.

[Noguès *et al*, 2003] Noguès M., Corteel E. and Warusfel O., *"Monitoring Distance Effect with Wave Field Synthesis",* Proceedings of the 6th International Conference on Digital Audio Effects (DAFX-03), 2003.

[Ono *et al,* 1997] Ono, K., Komiyama, S. and Nakabayashi, K., *''A method of reproducing concert hall sounds by loudspeaker walls",* Proceedings of the 102nd Convention of the Audio Engineering Society, Preprint 4490, 1997.

[Palombini, 1993] Palombini, C., *"Machine Songs V: Pierre Schaeffer: From Research into Noises to Experimental Music",* Computer Music Journal, Vol. 17(3), pp. 14-19, MIT Press, 1993.

[Pernaux *et al*, 1998] Pernaux, J., Boussard, P. and Jot, J., *"Virtual Sound Source Positioning and Mixing in* 5.1 *Implementation on the Real-Time System Genesis",* Proceedings of the International Conference on Digital Audio Effects (DAFx-98), 1998.

[Plenge *et al*, 1975], Plenge, G., Lehmann, P., Wettschurek, R. and Wikens, H., *"New methods in architectural investigations to evaluate the acoustic qualities of concert halls"*, Journal of the Acoustical Society of America, Vol. 57, pp. 1292-1299, 1975.

[Poletti, 1996] Poletti, M., *"The Design of Encoding Functions for Stereophonic and Polyphonic Sound Systems"*, Journal of the Audio Engineering Society, Vol. 44(11), pp. 948-963, 1996.

[Pulkki, 1997] Pulkki, V., *“Virtual sound source positioning using Vector Base Amplitude Panning”,* Journal of the Audio Engineering Society, Vol. 45, pp. 456–466, 1997.

[Pulkki, 1999] Pulkki, V., "*Uniform spreading of amplitude panned virtual sources"*, Proceedings of the 1999 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, Mohonk Mountain House, 1999.

[Pulkki *et al*, 2001] Pullki, V. and Karjalainen, M. *“Directional Quality of 3-D Amplitude-Panned Virtual Sources”*, Proceedings of the 7th International Conference on Auditory Display (ICAD2001), pp. 239-244, 2001.

[Pulkki, 2002] Pulkki, V., *“Compensating displacement of amplitude-panned virtual sources”*, Proceedings of the 22nd International Conference of the Audio Engineering Society, pp. 186-195, 2002.

[Pulkki *et al*, 2005] Pulkki, V. and Hirvonen, T., *“Localization of Virtual Sources in Multichannel Audio Reproduction”,* IEEE Transactions on Speech and Audio Processing, Vol. 13(1), 2005.

[Ratliff, 1974] Ratliff, P. A., *"Properties of Hearing Related to Quadrophonic Reproduction",* BBC Research Department Report, BBC RD 38, 1974.

[Reynolds, 1987] Reynolds, C. W., *“Flocks, herds, and schools: A distributed behavioral model”*, Computer Graphics, Vol. 21(4), pp. 25–34, 1987.

[Roads, 1996] Roads, C., *“The Computer Music Tutorial*”, pp. 470-472, MIT Press, 1996.

[Roads, 2004] Roads, C., “*Microsound”,* MIT Press, 2004.

[Ross, 2009] Ross, A., *“The Rest is Noise”,* Farrar, Straus and Giroux, 2007.

[Rumsey, 1998], Rumsey, F. *"Subjective assessment of the spatial attributes of reproduced sound,"* Proceedings of the Audio Engineering Society 15th International Conference, pp. 122-135, 1998.

[Rumsey, 2001] Rumsey, F., *"Spatial Audio"*, pp. 10-12, pp. 21-39, pp. 52-60, pp. 111-116, pp. 204-208, Focal Press, 2001.
[Saberi *et al,* 1990] Saberi, K. and Perrott, D. R., *"Minimum audible movement angles as a function of sound source trajectory"* Journal of the Acoustical Society of America, Vol. 88(6), 1990.

[Sanson *et al*, 2008] Sanson, J., Corteel, E. and Warusfel, O., *"Objective and subjective analysis of localization accuracy in Wave Field Synthesis",* Proceedings of the 124th Convention of the Audio Engineering Society, Preprint 7361, 2008.

[Schacher *et al,* 2006] Schacher, J. C. and Kocher, P. "*Ambisonics Spatialization Tools for Max/MSP"*, Proceedings of the International Conference on Computer Music, 2006.

[Schroeder, 1962] Schroeder, M. R., *"Natural Sounding Artificial Reverberation"*, Journal of the Audio Engineering Society, Vol. 10, pp. 219-223, 1962.

[Schroeder *et al*, 1974] Schroeder, M. R., Gottlob, D. and Siebrasse, K. F., *"Comparative study of European concert halls: Correlation of subjective preference with geometric and acoustic parameters"*, Journal of the Acoustical Society of America, Vol. 56, pp. 1195-1201, 1974

[Sheeline, 1983], Sheeline, C. W., *"An investigation of the effects of direct and reverberant signal interaction on auditory distance perception"*, Doctoral Thesis, Stanford University, 1983.

[Shinn-Cunningham, 2003] Shinn-Cunningham, B. G., *"Spatial hearing advantages in everyday environments"*, Proceedings of ONR workshop on Attention, Perception, and Modelling for Complex Displays, 2003.

[Simms, 1986] Simms, B. R., *“Music of the Twentieth Century: Style and Structure*”, Schirmer Books, 1986.

[Singer, 2008] Singer, E., “Boids for Max MSP/Jitter”, http://www. ericsinger. com, retrieved on 07/07/2008.

[Smalley, 1996] Smalley, D., *“The Listening Imagination: Listening in the Electro-acoustic Era”*, Contemporary Music Review, Vol. 13(2), pp. 77-107, Harwood Academic Publishers, 1996.

[Smalley, 1997] Smalley, D., *“Spectromorphology : Explaining Sound-Shapes”*, Organised Sound, Vol. 2(2), pp. 107-126, Cambridge University Press, 1997.

[Smalley, 2007] Smalley, D., *“Space-Form and the Acousmatic Image”*, Organised Sound, Vol. 12(1), pp. 35-58, Cambridge University Press, 2007.

[Smalley J., 2000] Smalley, J. *“Gesang der Jünglinge: History and Analysis”,* Masterpieces of 20th-Century Electronic Music: A Multimedia Perspective, Columbia University Computer Music Centre, www.music.columbia.edu/masterpieces/notes/stockhausen, retrieved 02/07/2009.

[Smalley *et al*, 2000] Smalley, D. and Austin, L., *“Sound Diffusion in Composition and Performance: An Interview with Denis Smalley”*, Computer Music Journal, Vol. 24(2), pp. 10-21, MIT Press, 2000.

[Smith, 1985] Smith III, J. O., *“A new approach to digital reverberation using closed waveguide networks”*, Proceedings of the International Computer Music Conference, pp. 47–53, 1985.

[Sonke, 2000] Sonke, J. J., "*Variable Acoustics by wave field synthesis*", Doctoral Thesis, Delft University of Technology, 2000.

[Sonoloco, 2007], *"Stockhausen Edition no. 91 (Cosmic Pulses)"*, Sonoloco Record Reviews, http://home.swipnet.se/sonoloco25/stockhausen/91.html, retrieved 02/07/2009.

[Spikofski, 2001] Spikofski, G. and Fruhmann, M., "*Optimisation of Binaural Room Scanning (BRS): Considering inter-individual HRTF-characteristics*", Proceedings of the 19th International Conference of the Audio Engineering Society, June 2001.

[Spors *et al*, 2003] Spors, S., Kuntz, A. and Rabenstein, R., *"An Approach to Listening Room Compensation with Wave Field Synthesis"*, Proceedings of the 24th International Conference of the Audio Engineering Society, 2003.

[Start, 1997] Start,E., *"Direct Sound Enhancement by Wave Field Synthesis"*, Doctoral Thesis, Delft University of Technology, 1997.

[Steinberg *et al*, 1934] Steinberg, J. and Snow, W., *"Auditory Perspective – Physical Factors"*, Electrical Engineering, Vol. 53(1), 1934, Reprinted in *"An Anthology of articles on stereophonic techniques"*, Audio Engineering Society, 1986.

[Stockhausen, 1959] Stockhausen, K., ". . . *How Time Passes . . ",* English translation by Cornelius Cardew, Die Reihe, Vol. 3, pp. 10-40, Universal Edition Publishing, 1959.

[Stockhausen, 1962] Stockhausen, K., *"The Concept of Unity in Electronic Music"*, English translation by Elaine Barkin, Perspectives of New Music, Vol. 1(1), pp. 39-48, 1962.

[Stockhausen, 1975a] Stockhausen, K.,*"Electronic and Instrumental Music",* English translation by Ruth Koenig, Die Reihe, Vol. 5, pp. 59-66, Universal Edition Publishing, 1975.

[Stockhausen, 1975b] Stockhausen, K.,*"Music in Space",* die Reihe, Vol. 5, pp. 67-82, English translation by Ruth Koenig, Universal Edition Publishing, 1975.

[Theile *et al*, 1976], Theile, G. and Plenge, G., *"Localization of lateral phantom-sources".* Journal of the Audio Engineering Society, Vol. 25, pp, 196-200, 1976.

[Theile, 1980] Theile, G., "*On the Localisation in the Superimposed Soundfield",* Doctoral Thesis, Berlin University of Technology, 1980.

[Theile, 1991] Theile, G. *"On the Naturalness of Two-Channel Stereo Sound",* Journal of the Audio Engineering Society, Vol. 39, 1991.

[Theile *et al,* 2003] Theile, G., Wittek, H. and Reisinger, M., *"Potential Wave Field Synthesis Applications in the Multichannel Stereophonic World"*, Proceedings of the 24th International Conference of the Audio Engineering Society, 2003.

[Trochimczyk, 2001] Trochimczyk, M., *"From Circles to Nets: On the Signification of Spatial Sound Imagery in New Music",* Computer Music Journal, Vol. 25(4), pp. 39-56, MIT Press, 2001.

[Truax, 1990] Truax, B., *"Composing with Real-Time Granular Sound",* Perspectives of New Music, Vol. 28(2), 1990.

[Truax, 1996] Truax, B., *"Soundscape, Acoustic Communication & Environmental Sound Composition"*, Contemporary Music Review, Vol. 15(1), pp. 49-65, 1996.

[Truax, 1999] Truax, B., *"Composition and diffusion: space in sound in space",* Organised Sound, Vol. 3(2), pp. 141-6, 1999.

[Truax, 2008] Truax, B., *"Soundscape Composition as Global Music: Electroacoustic Music as Soundscape",* Organised Sound, Vol. 13(2), pp. 103-109, 2008.

[Unemi *et al*, 2005] Unemi, T. and Bisig, D., *"Music by interaction among two flocking species and human",* Proceedings of the 3rd International Conference on Generative Systems in Electronic Arts, pp. 171–179, 2005.

[Uozumi, 2007] Uozumi, Y., *"Gismo2: An application for agent-based composition"*, Proceedings of the 2007 EvoWorkshops, pp. 609–616, 2007.

[Usher *et al*, 2004] Usher, J., Martens, W. L. and Woszczyk, W., *"The influence of the presence of multiple sources on auditory spatial imagery as indicated by a graphical response technique",* Proceedings of the 18th International Congress on Acoustics (ICA 2004), 2004.

[Varèse, 1998] Varèse, E., *"The Liberation of Sound",* Contemporary Composers on Contemporary Music*,* pp. 197, Perseus Publishing, 1998.
[Verheijen, 1998] Verheijen, E., "*Sound reproduction by Wave Field Synthesis",* Doctoral Thesis, Delft University of Technology, 1998.

[Vogel, 1993] Vogel, P., "*Application of wave field synthesis in room acoustics"*, Doctoral Thesis, Delft University of Technology, 1993.

[Wallach *et al*, 1949] Wallach, H., Newman, E. B. and Rosenzweig, M. R., *"The precedence effect in sound localization"*, American Journal of Psychology, Vol. 57, pp. 315-336, 1949.

[Wessel *et al*, 2002] Wessel, M., Wright, D. and Schott, J., *"Situated Trio - an interactive live performance for a hexaphonic guitarist and two computer musicians,"* Proceedings of the 2002 Conference on New Instruments for Musical Expression (NIME-02), pp. 171-173, 2002.

[Wessel, 2006] Wessel, D., *"An Enactive Approach to Computer Music Performance",* Le Feedback dans la Creation Musical, pp. 93-98, 2006.

[White, 2008] White, N. S., *"The Rosen' - a pretty thing, but is it a 'serious' autoharp?"*, Autoharp Notes, www.ukautoharps.org.uk, retrieved on 01/05/2009.

[Wiggins, 2004] Wiggins, B., "*An Investigation into the Real-time Manipulation and Control of Three-dimensional Sound Fields*", Doctoral Thesis; University of Derby, 2004.

[Wishart, 1985] Wishart, T., *"On Sonic Art"*, Imagineering Press, 1985.

[Wittek, 2002] Wittek, H., *"Optimised Phantom Source Imaging (OPSI) of the high frequency content of virtual sources in Wave Field Synthesis",* www.hauptmikrofon.de/wfs.html, retrieved 03/06/2009.

[Wittek, 2003] Wittek, H., *"Perception of Spatially Synthesized Sound Fields"*, 2003, www.hauptmikrofon.de/wfs.html, retrieved 03/06/2009.

[Wittek, 2007] Wittek, H., *"Perceptual differences between wavefield synthesis and stereophony"*, Doctoral Thesis, University of Surrey, 2007.

[Xenakis, 2008] Xenakis, I., "*Music and Architecture",* Architectural Projects, Texts, and Realizations Compiled, translated and commented by Sharon Kanach, pp. 146, Pendragon, 2008.

[Yadegari *et al,* 2002] Yadegari, S., Moore, F. R., Castle, H., Burr, A. and Apel, T., *"Real-Time Implementation of a General Model for Spatial Processing of Sounds"*, Proceedings of the International Computer Music Conference, pp. 244-247, 2002.

[Yeung, 2004] Yeung, B., "*Guitar dreams, an interview with Adrian Freed,"* www.sfweekly.com, 2004, retrieved on 01/04/2007.

[Zacharov, 1998] Zacharov, N., "*Subjective Appraisal of Loudspeaker Directivity for Multichannel Reproduction",* Journal of the Audio Engineering Society, Vol. 46(4), 1998.

[Zvonar, 2004] Zvonar, R., *"A History of Spatial Music"*, eContact, Vol. 7(4), Canadian Electroacoustic Community, 2004.

Appendix A – Scores