Spatial Composition Techniques and Sound Spatialisation Technologies

MARIJE A.J. BAALMAN

Pibemalaan 16, 8749 GS, Pingjum, The Netherlands
E-mails: marije@nescivi.nl, nescivi@gmail.com

In the context of current developments towards an exchange format for spatial audio, it is important to consider the interactions and tensions between spatial composition techniques and spatial audio technologies. This paper gives an overview of common compositional techniques and audio technologies used for spatial compositions, and discusses various forms of hybrid uses of audio technologies. A spatial composition created using certain audio technologies may lose integrity when transferred to another audio technology, when the compositional intent is not taken into account. Tools that are flexible with regard to the spatial audio reproduction technology applied are useful for the comparison of various spatialisation methods during the compositional process, but are also essential to enable a hybrid use of technologies within one composition.

1. INTRODUCTION: TECHNIQUES VERSUS TECHNOLOGIES

This paper aims to discuss the tension between compositional techniques to create ‘spatial music’, and audio technologies for spatialisation. While the former cannot be realised without using the latter, the choice of audio technology can have implications on the resulting effect of the composition, which may or may not be according to the composer’s intent. I think it is therefore important to reflect where technological limitations are influencing the choices in techniques used in spatial composition, and where composers are exploiting the advantages and limitations of spatial audio technologies. This is an especially important topic to consider, as the current development of a standardised format for spatial audio may have great impact on the production and performance practice of spatial sound compositions.

Spatial audio technologies attempt to find a technical solution to create a spatial sound image. The intent of the various methods is to represent a sound image that is somehow convincing or ‘realistic’. In most cases the motivation for the development of the technology is the desire to recreate an acoustic event; in other words, recordings are made of a specific acoustic event and the technology is used to give the listener a spatially correct reproduction of this acoustical event. The theory behind the operation of the technology is usually founded in the physical properties of sound propagation and/or the psycho-acoustic characteristics of our spatial hearing. The technology usually encompasses both prescriptions for speaker placement and algorithms for signal processing.

In contrast, spatial composition techniques embody the artistic approach to the use of space, and they are strongly related to or part of the artistic concept of the work. Considerations that are made in this context are dependent on what the composer wants to achieve with the composition, what kind of spatial image he or she wants to create. The choice of which technology to use to achieve the compositional concept is based on several factors, such as the desired accuracy of the spatial image, as well as the performance context.

To limit the scope of this paper, I am focusing on composition of electronic or electroacoustic music; in other words, music that is specifically composed for reproduction via loudspeakers. Usually this involves either synthesis of sounds or manipulation of recorded sounds, in the sense that the aim of the manipulation is to transform the recorded event, rather than reproduce it. I am not excluding sound installations which use loudspeakers as the sounding instruments from this discussion. However, it is beyond the scope of this paper to discuss the extensive body of work in the field of (mostly stereophonic) audio production methods, applied in both classical recordings and popular music. Similarly, the centuries-long history of spatial composition in vocal and orchestral music will not be addressed.

This paper will discuss various aspects of spatial composition, provide an overview of current spatial audio technologies, and discuss the caveats for the current development of a spatial audio format. The discussion is illustrated with an example of the author’s own compositional work.

2. SPATIAL COMPOSITIONAL TECHNIQUES

Common compositional techniques encompass the following:

- Creating trajectories; these trajectories will introduce choreography of sounds into the piece, and this choreography needs to have a certain meaning.
- Using location as a serial parameter (e.g. Stockhausen); this will also introduce choreography of sounds.

Organised Sound **15**(3): 209–218 © Cambridge University Press, 2010. doi:10.1017/S1355771810000245

- Diffusion, or (uniform) distribution of the sound energy; creating broad, or even enveloping, sound images.
- Simulation of acoustics; adding reverberation and echoes.
- Enhancing acoustics; tuned to the actual performance space, by exciting specific resonances of the space (e.g. Boulez in *Répons*).
- Alluding to spaces by using sounds that are reminiscent of specific spaces or environments (indoor/outdoor/small/large spaces) (Barrett 2002).

The spatial techniques usually interact with other elements of the compositional dimensions, both within music (e.g. the sound material, the synthesis processes and melody), or external to the music, when other media are involved in the work. Varèse was imagining the spatial locations and movements of sounds as fourth dimension in composition, before there even were technologies to achieve these concepts (Varèse 1936). He only got the chance to develop them when he was invited by Le Corbusier and supported by Xenakis to compose the *Poème électronique* in 1958 (Treib 1996).

Locations of sounds (both moving and static) can help with the stream segregation of sounds; that is, it is possible to distinguish different sonic streams from each other, because the sounds will be coming from a different position.

Spatial characteristics of the sounds will usually also change over the course of one composition, moving from one spatial impression to another (spatio-*temporality*). The effect of memory then also starts to play a role, as a spatial image may be perceived differently based on what was heard before (Barrett 2002).

Sounds can also be linked together in compositional groups, either because they are coming from the same original sound source (e.g. multichannel recordings of the same sonic event), or because they are connected to each other based on some internal compositional semantic. Then these groups can be subjected to spatial transformations, such as trajectories, simulated acoustics, and so on (see also Kendall 2007).

In some cases the artistic concept may also be to explore the 'artefacts' that occur when transgressing the limitations of specific spatial audio technologies. These may let the spatial image collapse all of a sudden, or create contradictory spatial cues for the listener (Barrett 2002). In a sense – within electroacoustic music – composers are interested in 'abusing' the technology; composers are not so much concerned with creating 'realistic' sound events, but rather interested in presenting the listener with spatial images which do not occur in nature.

The desired accuracy of trajectories or locations of sounds is also an important consideration: are precise locations of the sounds desired, or is it enough to think in terms of 'front-left', 'right', and so on? Is a clear localisation needed? Or is clarity of localisation also a parameter within the spatial composition (Baalman 2008, ch. 4)?

The choice of which audio technology to use within a composition is often a pragmatic one, based on availability but also on experimentation. During the composition process, the composer tries out various ways of realising his compositional ideas, and based on what gives the most satisfactory result that technology is chosen. It is therefore important to allow for flexible choice of technology. The same speaker setup may support various spatial audio technologies and the composer should thus not be restricted to using only one specific technology within one piece (e.g. using eight speakers of a circular ring as singular monophonic sources, as well as using them for other sections with stereophonic panning, or ambisonics). Even technologies which may not strictly be applicable to a given speaker setup still may be useful to achieve specific effects (e.g. using wave field synthesis on a sparse loudspeaker array). Composers may modify a given technology (i.e. alter the prescriptions of the technology and thus create a new form of spatial imagery), suitable to their compositional ideas.

The spatial concept of the piece may of course change in the process of composing, as experimentation with various techniques and their realisation with different audio technologies give the composer new ideas of where to go with the composition. Specific techniques or technologies may not work, either in a technical sense (i.e. they do not give the desired acoustic effect), or in an artistic sense (i.e. they do not result in the desired compositional effect).

The performance context also influences the choice of techniques and technologies to use. Considerations here are both what the available speaker setup will be, as well as how the audience will be experiencing the piece (i.e. will they have fixed seats, or can they move around to change their listening perspective?), the listening area (where does the audience go?), interactive components (does the audience or the performer have influence on the sound and spatialisation as the work is performed?), where the performance takes place (is it site-specific, or will it have to adapt to different setups for different performances?), and also the size of the performance space (small room sizes allow for different techniques/technologies from those that can be used in large spaces).

As in the development of spatial audio technologies, also within spatial composition, the characteristics of human aural spatial perception needs to be considered. Issues such as accuracy of localisation in the horizontal and vertical plane (e.g. Blauert 1997; Moore 2004), play a role, but so do psychological aspects such as intimacy vs. distance and comfort vs. discomfort, as

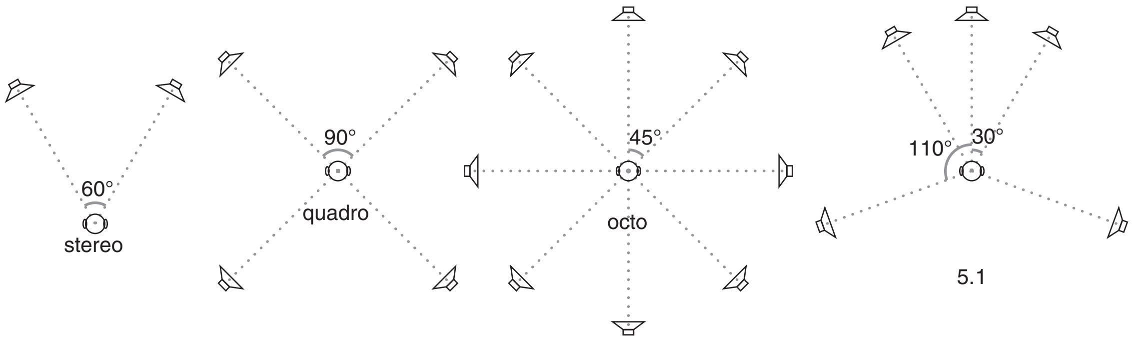

Figure 1. Stereophonic configurations, from left to right: stereophonic, quadrophonic, octophonic and 5.1 surround.

caused by spatial location (and culture). Also, when the aim is to confuse the listener as to spatial image, it is useful to know where our spatial perception behaves ambiguously.

3. SPATIAL AUDIO TECHNOLOGIES

The extent to which spatial concepts can be perceived by the listener – and thus play a role in the reception of a musical composition – is largely dependent on the technology available and used to achieve spatialisation. In the following an overview is given of common technologies and their advantages and disadvantages. The overview is limited to speaker-based technologies.

3.1. Reproduction technologies

3.1.1. Monophony and polyphony

Monophony means simply reproducing the sound from one loudspeaker; you can expand this to multichannel monophony, or polyphony, if you do not use any other technique than routing the audio signals to different single speakers – for example, if the source position (or speaker channel) is used as a serial parameter. Polyphony in this sense can also be useful to achieve stream segregation – in other words, achieving perceptually different melody lines (even when within the same frequency bands) as they are played back from different locations.

Some composers (e.g. G.M. König) have refrained from using stereophony for spatialisation as they feel dissatisfied with the result, because of the weaknesses of the technology. In a personal conversation with Alberto de Campo, Curtis Roads observes: 'If you want the sound to come from a specific place, put a loudspeaker there.'[1]

3.1.2. Stereophony

In the development of loudspeaker technology, there has been a development from mono to stereo, and to multichannel stereophony (quadrophony, octophony, 5.1 surround, 7.1 surround, etc.). In stereophony a spatial impression is created by level differences between the channels, and in some cases also timing differences. Stereophony works best if the listener is located in the 'sweet spot'; for two-channel stereophony this means the listener has to sit in the position where he or she faces both speakers at a 30° angle from the centre line. See figure 1 for an indication of the ideal speaker setups of several common stereophonic setups.

Within electroacoustic music, quadrophonic and octophonic layouts have become a fairly common concert setup. The main problem in using quadrophony (e.g. Kopp 2005) is not only that manufacturers did not agree upon a common format for distribution, but also that the angles between the four speakers (one in each corner of the hall) are actually too wide to achieve a good audible result: between all speaker pairs there is a 'hole' when the source is in the middle between the speakers. Moreover, the actual intended result is only heard properly in the centre of the four speakers, the so-called *sweet spot*. However, in many concert situations for electroacoustic music this position is reserved for the composer or engineer controlling the mixing desk and no audience member is actually seated in the sweet spot (Böhmer 1961). Octophonic setups (eight speakers equally spaced in a circle around the audience) remedy the situation only partially, as the angles between the speakers are made smaller, so the 'hole' in the middle between the speakers is not as serious. Still, there is a large dependency on the sweet spot.

There are some examples of works that use many more speakers, such as the multimedia work *Poème électronique* (1958) of Edgard Varèse, where over 300 loudspeakers were used to create spatial trajectories through the Philips Pavilion (by Le Corbusier and Xenakis), and the hemisphere spatialisation for the

---

[1] Some time around the year 2000, probably at Sound in Space Conference, Santa Barbara (source: email communication between the author and De Campo).

World Expo in Osaka (1970) using 50 loudspeaker groups (Gertich, Gerlach and Föllmer 1996).

Recently, there is a renewed interest in the hemisphere for spatialisation, for example the *Klangdom* in the ZKM, Karlsruhe (Ramakrishnan, Goßmann and Brümmer 2006). One of the reasons for this development could be that it is now much easier to control the spatialisation over such large systems, as the signals can be calculated by software on a commercially available computer (e.g. Bartetzki 2007), and are no longer dependent on custom-designed hardware for the purpose.

Simultaneously, the surround formats introduced by the film industry, such as 5.1 surround and 7.1 surround, are gaining some popularity in the field of electroacoustic music, since they provide a means for a large-scale distribution of the works, made available through the DVD format, as well as an interest on the part of radio stations to broadcast in these surround formats, and the use of movie theatres as concert halls. Nonetheless, these surround formats do pose some problems as they were primarily developed for a frontal presentation, and as a consequence there are more speakers located in the front of the listener (where the visual focus is), than in the rear, and in some of the surround sound audio codecs the rear channels are encoded with a lower resolution. Thus although the technology provides a means for distribution, the composer has to take into account that he or she has to compose for an unbalanced space, where the front is more important than the rear. It is hard, if not impossible, to place sounds at the sides, as the angles between the front-left and rear-left (and also front-right and rear-right) are too large for stereophonic panning.

3.1.3. Vector Base Amplitude Panning (VBAP)

Vector Base Amplitude Panning is a method for positioning virtual sound sources to arbitrary directions using a three-dimensional (3D) multichannel speaker setup (Pulkki 2001). Based on the desired position of the sound source and the positions of the loudspeakers, the audio signal is panned to one, two or three speakers at a time. Instead of pair-wise panning, as applied in stereophonic techniques where all speakers are placed in the horizontal plane, a triplet-wise panning paradigm is used. In the case of a two-dimensional (2D) loudspeaker setup, pair-wise panning is used. Based on the desired sound source direction, a weighting is made of the direction of the three speakers 'around' that direction to calculate three gain factors for distributing the sound source to these three speakers so that it will seem that the sound source comes from the desired direction. With more than eight speakers it is possible to create a fairly constant spatial impression to a large listening area. As with stereophonic techniques, problems of difference in spread can occur (localisation blur), depending on how many speakers are reproducing the sound.

VBAP is particularly suitable for hemisphere loudspeaker setups, where the actual distribution of the loudspeakers over the hemisphere may vary, but they are at the same distance from the centre of the sphere.

3.1.4. Distance based amplitude panning (DBAP)

Distance based amplitude panning extends the principle of equal intensity panning from a pair of speakers to an arbitrary number of speakers with no a priori assumptions about their positions (although the coordinates need to be known) in space or relations to each other (Lossius, Balthazar and De la Hogue 2009). The distance from the virtual sound source position to each speaker is calculated and it is assumed that for a source with unity amplitude the overall sound intensity of the source has to be constant, regardless of the position of the source. It is also assumed that all speakers are active at all times. By adding some spatial blur to the distance calculation, the method avoids the case where a sound source coincides with only one speaker, thus smoothing out variations in localisation blur. For sources outside of the field of the loudspeakers, the source position is projected onto the outer perimeter of all the speakers, and the overall intensity of the sound source is diminished based upon the distance of the actual desired source position to the projected position. The technology can be further extended by introducing weights for each speaker, which allows for the restriction of a sound source to a specific subset of speakers.

With this extension, DBAP is similar to the tools proposed by Rutz, which control the sound spatialisation based on speaker positions and source positions, where each speaker has an arbitrary (user-defined) function assigned to it determining how much of a source within the reach of this function is played back on a certain speaker (Rutz 2004). This approach can also be used to determine other parameters of sounds and a way to create a 2D control interface for multiple parameters (Momeni and Wessel 2003).

DBAP can be useful in an installation context, where loudspeakers are spread out over one or multiple physical spaces in which the listener can freely move around, and the choice of loudspeaker positions are motivated not solely by their acoustical purpose, but also by visual and architectural artistic concepts.

3.1.5. Ambisonics

The ambisonic sound system is a two-part technological solution to the problem of encoding sound directions and amplitudes, and reproducing them over practical loudspeaker systems so that listeners can perceive sounds located in a 3D space (Malham and Myatt 1995).

This can occur over a 360° horizontal-only sound stage (pantophonic system), or over the full sphere (periphonic system). The system encodes the signals in a so-called *B-format*; the first order version of this encodes the signal in three channels for pantophonic systems and a further channel for the periphonic; in other words, 'with-height' reproduction.

Essentially the system gives an approximation of a wave field by a plane wave decomposition of the sound field: in other words, a kind of 'quantisation' of the directions from which sounds are coming, at the listener's position. This approximation becomes more precise as the order of ambisonics is increased, which also means that more channels are needed for encoding, and that more speakers are needed for decoding. For very high orders, ambisonics can be equated to wave field synthesis (WFS) (Daniel, Nicol and Moreau 2003).

In recent years, ambisonic techniques have become more and more popular as a way of spatialising sound. First-order systems are most common, though higher-order systems are slowly gaining ground. The use of ambisonics allows for different spatialisation effects from those afforded by stereophonic techniques, though it is still dependent on a 'sweet spot' where the effect is optimal, especially in low-order ambisonics. Ambisonics does not encode the distance of sound sources explicitly (although in an ambisonics recording the information will be present), and the composer Jan Jacob Hofmann,[2] who has worked extensively with higher-order ambisonics in his work, commented after hearing a WFS demonstration that WFS was much better than ambisonics at reproducing spatial depth.

Ambisonics are particularly popular with artists working with field recordings due to the availability of microphones suitable for recording in the ambisonic format such as the *SoundField* microphone and the *TetraMic* (Adriaensen 2007).

3.1.6. Wave Field Synthesis

Wave Field Synthesis (WFS) is a holophonic approach to spatial sound-field reproduction based on the Huygens principle (Berkhout 1988). With WFS it is possible to create a physical reproduction of a wave field; this has the advantage over other techniques in that there is no *sweet spot* (i.e. at no point in the listening area is the reproduction better than at other places), but rather there is a rather large *sweet area*. Note, however, that this does not mean that the spatial impression of the listener is the same over this whole area. The spatial impression is dependent on the listener's position in space and relative to the virtual sound source positions. As the listener moves through space, his or her perspective on the sound scene will change, just as if there were real sound sources at the virtual source positions. This is in contrast with stereophonic techniques, where the virtual source position moves along with the listener until the sound coming from one speaker dominates and the sound 'collapses' onto that speaker.

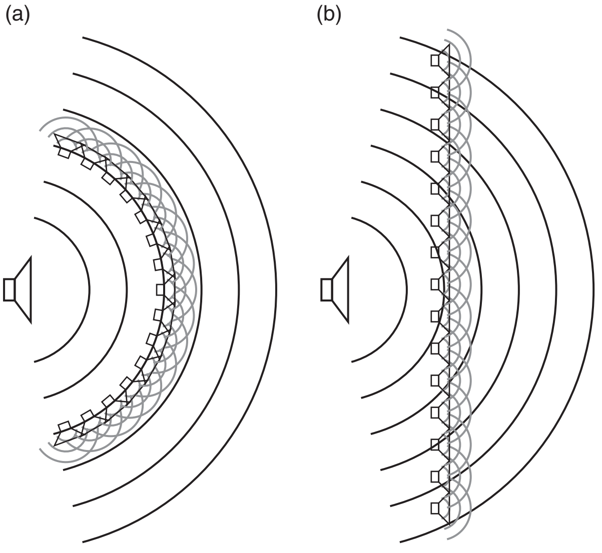

Figure 2. From the Huygens' Principle to Wave Field Synthesis: (a) The Huygens' Principle; (b) Wave Field Synthesis.

The principle of Huygens (see figure 2a) states that when you have a wave front, you can synthesise the next wave front by imagining on the wave front an infinite number of small sound sources, whose waves together will form the next wave front (Huygens 1690). A listener will then not be able to determine the difference between a situation where the wave front is real, or when it is synthesised. This principle can be translated to mathematical formulae using theories of Kirchhoff and Rayleigh and can then be applied for use with a linear array of loudspeakers (see figure 2b) (e.g. Baalman 2008); the algorithm consists of calculating appropriate delay times and attenuation factors for each sound source and speaker. The limitations of WFS are partly pragmatic as it needs a large number of loudspeakers and computation power. There is a theoretical limitation, resulting from the physical dimensions of loudspeakers, which causes a quantisation in the spatial domain, so that *spatial aliasing* will occur above certain frequencies. The extent to which this spatial aliasing (affecting the higher frequency bands of the sound) is audible depends both on the sound material (the amount of high frequency content), as well as on the source position and the listener position. Some methods to mitigate the effects of spatial aliasing have been proposed (e.g. Wittek 2002).

In most current WFS implementations the speakers are placed in a horizontal circle or square around the

---

[2] www.sonicarchitecture.de.

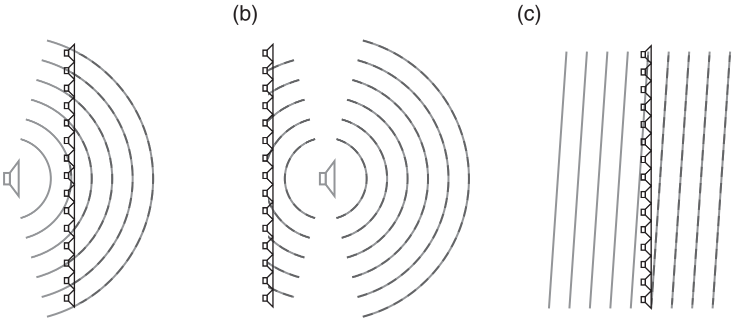

Figure 3. Common source types in Wave Field Synthesis implementations: (a) Point source behind the loudspeakers; (b) Point source in front of the loudspeakers; (c) Plane wave.

listening area so sound sources can only be placed in the horizontal plane. They can be placed in this plane at any place behind, and even in front of, the speakers. Another option is to create a plane wave source, which will seem to come from the same direction (independent of the listener's position in the listening area), rather than from a specific location (see figure 3).

Wave Field Synthesis offers new possibilities for electroacoustic music (Baalman 2007, 2008), in that it allows more accurate control over movement of sound sources, enables other types of movement than those often seen with circular octophonic setups (such as movements which use depth and movements through the listening space), and provides an interesting possibility for contrasts between the effect of point sources and of plane waves: a source with a fixed position versus a source which moves along when the listener walks through the sound field. Room reflections can be added by positioning mirror image sources as virtual sources, and reverberation can be reproduced by a number of plane waves from different directions to create realistic sounding (virtual) acoustics.

3.2. Reflections and reverberation

In room acoustics the room response of a space is usually divided into early reflections (occurring up to around 80ms after the direct sound) and reverberation. The early reflections come from the walls, floor, ceiling and other surfaces close to the sound source. Later reflections (in the reverberation tail) have reflected multiple times and create a colouration of the sound.

Reflections can be mimicked by delaying and attenuating – and optionally playing back from a different virtual position or direction – a sound source and adding it to the original sound. The delay times and attenuation factors can be determined by calculation of a mirror-image source model.

To simulate reverberation, algorithms using a number of filters, delay lines and feedback loops have been developed, for example the Schroeder reverberator (Schroeder 1962).

It is also possible to record or measure the impulse response of a space and use convolution to add this room response to any sound. An impulse response can also be calculated by acoustic modeling software.

Room acoustics can also be described with perceptual parameters such as presence and envelopment, which determine the balance between the direct sound and reverberated sound.

3.3. Decorrelation

Decorrelation of audio signals creates timbral colouration and combing due to constructive and destructive interference, which produces a diffuse sound field and does not suffer from image shift and the precedence effect (Kendall 1995; Vaggione 2001). This method does not aim to create a physically correct image, but rather uses psycho-acoustic effects to achieve a diffuse, broad sound image. This effect is related to the inter-aural cross-correlation coefficient (IACC), which has an influence on the apparent source width (ASW) (e.g. Baalman 2002). Like reverberation, decorrelation can be used to fill the space with sound, or to create a broad enveloping sound image.

3.4. Combinations and hybrid use

Combinations of technologies are of course possible. Reflections and reverberation can be used with any of the mentioned reproduction technologies, as can decorrelation.

Various reproduction technologies can also be combined with each other: for example, you can project a stereophonic setup onto an ambisonics or WFS setup by using the stereophonic speakers as virtual sources. This can be advantageous in certain cases as virtual sources can be placed at any angle and

Figure 4. *They Watch*, 2009. Installation created by Thomas Soetens and Kora van den Bulcke. Produced by Workspace Unlimited. Photograph courtesy of Workspace Unlimited, used with permission.

distance and are not limited by physical constraints, such as walls or other objects in the room. However, it is not without potential drawbacks as the sound source will be reproduced by several loudspeakers rather than just one. This can result in adverse effects of localisation blur, phase cancellation, and precedence (where the sound arrives earlier at the listener's ears from one speaker, creating a false spatial impression). It depends both on the original sound material as well as the reproduction technology whether such a virtualisation of the loudspeakers retains the same quality as the original.

Most audio reproduction technologies have been developed in an audio engineering context with the aim to reproduce or simulate a real acoustic event. In the composition of electroacoustic music we are not so much concerned with a realistic sound image, rather we are interested in creating tensions between familiar, natural and unfamiliar, virtual sound images. While audio engineers are searching for a methodology for a 'perfect' spatial audio reproduction, composers are free to use any technology to realise their compositional concepts, and choose the most suitable (re)production method – based upon its strengths or its weaknesses – in order to create the spatial tensions they desire to present to the listener. For example, to create temporal developments the composer can shift from using one spatial reproduction technology to another, thereby giving the listener conflicting spatial cues (Barrett 2002). As such, the audio technology itself becomes a parameter of composition.

4. APPLICATION IN *THEY WATCH*

As an example I would like to discuss the spatial composition I created for the interactive installation *They Watch*, which is a production of Workspace Unlimited.[3]

> *They Watch* is an immersive art installation with virtual characters literally watching visitors. Several duplicates of the virtual characters – one man, one woman, and both portraits of the artists – surround and interact with visitors, who are tracked as they move about the physical space, and even projected into the virtual space. … The subtle collaboration of the real and virtual agents and environments conflate to engender a hybrid space where the observer becomes the observed.[4]

In the installation the visitor enters a 360° projection space and moves around in a virtual 3D space (see figure 4). The visual projection is a 3D environment created in a customised *Quake* game engine. The audio reproduction is achieved by a circular 16-channel setup in a double ring (two speakers vertically above each other, reproducing the same channel). The visitor's movements in the space influence his or her position in the virtual space, as the 3D projection changes with the movement, and the soundscape dynamically moves and shifts in relationship to the visitor.

The soundscape consists of two distinct sound environments within which the visitor can move, and each of these environments is built up from several layers of sound containing sound sources with a particular sonic quality (called sound objects in the following discussion).

---

[3] The research for this project was a part of Workspace Unlimited's research project *Gameorama*, led by Thomas Soetens and Kora van den Bulcke, funded by the Institute for Science and Technology (IWT) Vlaanderen, Belgium. Besides the author of this paper, Matt McChesney and Hans Samuelson were also involved in the realisation of the sound server for *They Watch*. The sound server and the composition are implemented in SuperCollider (http://supercollider.sourceforge.net).

[4] www.workspace-unlimited.org.

The first is an environmental soundscape, when the visitor is in the centre area of the space. At this moment the virtual characters the visitor sees on the screen are all at a distance. The soundscape is a layering of sound objects moving around the space, as well as a field recording, which is played back using ambisonic reproduction. Some of the sound objects are linked to the virtual characters on the screen and have vocal character. The other sound objects have no visual counterpart and have a more musical or noise-like character.

The second soundscape is an intimate environment as the visitor approaches the screen. As the visitor gets closer to the virtual characters, they increase in size. Here, the soundscape is representing the intimate 'thought-space' of the virtual characters. The soundscape is much quieter, as only a few sound objects are moving around the space (some piano-based sounds and a high-pitched noisy sound). Voices reciting a poem are projected (or focused) onto the visitor's physical location using WFS-technology to create the effect of hearing the sound inside the head.

Towards the end of the composition, a climax is created. As the intensity and density of the soundscape is increased, the visitor loses control over the environment, and the two different soundscapes merge, the voices of the virtual characters being projected in a more chaotic, confused fashion onto the visitor's physical position.

In this composition, a number of technologies were used for the realisation of various compositional ideas:

- Ambisonics was used to reproduce a stereophonic field recording of an ambient space. Ambisonics was chosen as this gives a particularly spatial and diffuse impression of the recording, where the loudspeakers and the sound sources cannot be localised per se.
- Circular amplitude panning technologies placed or moved sound sources around the space, using an attenuation of the volume and a low pass filtering to simulate distance. A slightly larger panning width was used, between three speakers, rather than two, to create a slightly wider sound-source image.
- Wave field synthesis technology was used to project sound inside the listening area to create a very intimate effect of creating sound inside the listener's head.

Even though the distance between the loudspeakers was considerably larger than WFS theory prescribes, for the particular effect that we wanted to achieve the WFS algorithms were very effective. Initially, we also experimented with using WFS for matching vocal sounds to the positions of the virtual characters on screen in the first sound environment, but it was found that this did not match up well between the visual location on screen and the localisation of the audio. (This issue of source location mismatch between a 2D image and 3D sound projection is discussed in detail in De Bruijn and Boone 2003.)

The choice for using specific technologies was motivated by both the concepts for the piece (specific atmospheres for a one-person experience), as well as its technological and performance context (the use of a circular speaker setup in conjunction with a circular visual projection). To realise the compositional ideas, we did not restrict ourselves to one given technology, but rather used the technologies that were most suitable for each idea.

5. SPATIAL AUDIO ATTRIBUTE STANDARDISATION

Currently, there is a progression from the traditional track/channel-oriented way of working, where an audio track corresponds to what is played on a certain speaker channel and the composition is mixed down to the speaker signals that will be used for the reproduction, to an 'object-oriented' way of working, where the audio track is treated as a sound-source object, and the spatial characteristics that are needed for the reproduction of the source are stored as meta data, so that the reproduction method is independent of the encoding: for example, MPEG4 (Plogsties, Baum and Grill 2003), and XML3DAudio (Potard 2006). This has as the advantage that spatial compositions are interchangeable between systems as well as reducing the number of stored audio channels for systems that use many loudspeakers. More recently, SpatDIF5 has been proposed as a standard for encoding the spatial data (Kendall, Peters and Geier 2008).[5] In a sense, these formats try to separate the compositional techniques from the reproduction technology.

Each spatialisation method has its own advantages and limitations, and artistic concepts may work better using one technology rather than another. While a common standard and control protocol for describing spatial attributes will facilitate transferring a spatial composition from one system to another and from one technology to another, it should be kept in mind that whether or not the same effects can be achieved with another technology is dependent on the techniques used with a particular technology. Even within the same technology implementation details may vary by system and/or software, resulting in a different reproduction of the composition.

Furthermore, it is questionable how well compositions made for small systems (and thus usually also for a smaller physical space) scale up to large systems, and vice versa. It is unlikely that there will be a straightforward way of scaling up compositions to a larger space (both virtual and physical): a composer may have composed a piece for a small setup and

---

[5] www.spatdif.org.

have audio material on different tracks that are related to each other, which when they are scaled up in terms of single tracks – and in doing so moved further apart from each other – may lose their common sound impression. When working with a larger system it may make sense to have a larger number of separate sound sources, as more sound sources will be distinguishable by the listener (as was done by composer Hans Tutschku, when he adapted his composition *Rituale*, originally written for the Ilmenauer WFS system for the WFS system of the TU Berlin, which was much larger), whereas in downscaling several tracks could be combined. An issue here is also the question of intimacy; a piece composed for a small room tends to be much more intimate than a piece for a large room. There may be cases where a change in scale does not make compositional sense.

Thus, while a common standard will ease the reproduction of compositions on systems other than those for which they were composed, it has to be noted that a transfer to another system is a different interpretation of the piece and that – while making the transfer – the compositional techniques used in the work should be taken into account in order to stay faithful to the compositional concept.

6. DISCUSSION AND CONCLUSION

As stated throughout this article, it is important to make a distinction between compositional techniques and the technologies used to realise them. Although the technologies may motivate the choice of compositional technique to use – as certain technologies are better at realising certain concepts than others – it is critical to be aware where the compositional choice is influenced by the technology used, in order to discern whether or not it would be worthwhile to try another technology to realise the concept if one technology is not affording the desired results. Barrett, in her discussion, treats working with 3D audio reproduction as a different compositional technique (Barrett 2002), whereas I would like to emphasise that this is a different technology to realise a compositional concept.

Additionally, the tool implementing a technology can restrict, or open up, the possibilities of using various compositional techniques. The availability of 3D sound reproduction technologies requires the development of useful tools for composers to use these technologies, as they are essentially a bridge between the compositional ideas and the realisation with the various technologies. The 3D technologies point towards an emphasis on 'object-oriented' working with sound, rather than fixed rendering to channels which will be played back over a fixed speaker array. While for some composers this is not a new concept in working with audio, for others it may require a change in their thinking about space in composition, as they are, for example, very much used to working with stereophonic technology and base their compositional ideas on the possibilities within that technology. Tools that are flexible with regard to the spatial audio reproduction technology applied are useful for the comparison of various spatialisation methods during the compositional process, but are also essential to enable a hybrid use of technologies within one composition.

Spatial compositional concepts and techniques on one hand, and spatial audio technologies and tools on the other hand are in a continuous discourse with each other, where concepts and techniques can set demands for technologies and tools, and the latter can help develop, but also limit thinking of compositional concepts and techniques. Awareness of and reflection on where the boundaries are between the techniques and technologies we use are essential to further new concepts for spatial composition and the development of innovative technologies and tools for their realisation.

REFERENCES

Adriaensen, F. 2007. A Tetrahedral Microphone Processor for Ambisonic Recording. *Proceedings of the Linux Audio Conference 2007*, Technische Universität Berlin, 22–25 March.

Baalman, M.A.J. 2002. *Spaciousness in Concert Halls: Perceptibility of IACC-differences*. Master's thesis, Technische Universität Delft.

Baalman, M.A.J. 2007. On Wave Field Synthesis and Electro-Acoustic Music: State of the Art 2007. *Proceedings of the International Computer Music Conference 2007*, Copenhagen, Denmark, 27–31 August.

Baalman, M.A.J. 2008. *On Wave Field Synthesis and Electro-Acoustic Music, with a Particular Focus on the Reproduction of Arbitrarily Shaped Sound Sources*. PhD thesis, Technische Universität Berlin.

Barrett, N. 2002. Spatio-Musical Composition Strategies. *Organised Sound* **7**(3): 313–23.

Bartetzki, A. 2007. A Software Based Mixing Desk for Acousmatic Sound Diffusion. *Proceedings of the Linux Audio Conference 2007*, Technische Universität Berlin.

Berkhout, A.J. 1988. A Holographic Approach to Acoustic Control. *Journal of the Audio Engineering Society* **36**(12): 977–95.

Blauert, J. 1997. *Spatial Hearing*. Cambridge, MA: The MIT Press.

Böhmer, K. 1961. Raum-Former. In *Das böse Ohr: Texte zur Musik 1961–1991*, ed. Burkhardt Söll. Cologne: DuMont, 1993.

Daniel, J., Nicol, R. and Moreau, S. 2003. Further Investigations of High Order Ambisonics and Wavefield Synthesis for Holophonic Sound Imaging. *Proceedings of the 114th Convention of the Audio Engineering Society*, Amsterdam, 22–25 March.

De Bruijn, W.P.J. and Boone, M.M. 2003. Application of Wave Field Synthesis in Life-Size Videoconferencing. *Proceedings of the 114th Convention of the Audio Engineering Society*, Amsterdam, 22–25 March.

Gertich, F., Gerlach, J. and Föllmer, G. 1996. *Musik …, verwandelt. Das Elektronische Studio der TU Berlin 1953–1995*. Hofheim: Wolke Verlag.

Huygens, C. 1690. *Traité de la lumière; où sont expliquées les causes de ce qui luy arrive dans la réflexion et dans la refraction et particulièrement dans l'étrange refraction du cristal d'Islande; avec un discours de la cause de la pesanteur*. Leiden: P. van der Aa.

Kendall, G.S. 1995. The Decorrelation of Audio Signals and its Impact on Spatial Imagery. *Computer Music Journal* **19**(4): 71–87.

Kendall, G.S. 2007. The Artistic Play of Spatial Organization: Spatial Attributes, Scene Analysis and Auditory Spatial Schemata. *Proceedings of the International Computer Music Conference 2007*, Copenhagen, 27–31 August, vol. 1: 63–8.

Kendall, G.S., Peters, N. and Geier, M. 2008. Towards an Interchange Format for Spatial Audio Scenes. *Proceedings of the International Computer Music Conference (ICMC)*, Belfast, August.

Kopp, S. 2005. *Programmierung eines Plugins zur Dekodierung quadrofon matrizierter Audiosignale*. Master's thesis, Technische Universität Berlin.

Lossius, T., Balthazar, P. and De la Hogue, Th. 2009. DBAP: Distance Based Amplitude Panning. *Proceedings of the International Computer Music Conference (ICMC) 2009*, Montreal, QC: 489–92.

Malham, D.G. and Myatt, A. 1995. 3-D Sound Spatialization using Ambisonic Techniques. *Computer Music Journal* **19**(4): 58–70.

Momeni, A. and Wessel, D. 2003. Characterizing and Controlling Musical Material Intuitively with Geometric Models. *Proceedings of the 2003 International Conference on New Interfaces for Musical Expression*, McGill University, Montreal, QC, 22–24 May: 54–61.

Moore, B.C.J. 2004. *An Introduction to the Psychology of Hearing*. London: Elsevier Academic Press.

Plogsties, J., Baum, O. and Grill, B. 2003. Conveying Spatial Sound using MPEG–4. *Proceedings of the AES 24th Conference*, Banff, 26–28 June.

Potard, G. 2006. *3D-Audio Object Oriented Coding*. PhD thesis, University of Wollongong.

Pulkki, V. 2001. *Spatial Sound Generation and Perception by Amplitude Panning Techniques*. PhD thesis, Helsinki University of Technology.

Ramakrishnan, C., Goßmann, J. and Brümmer, L. 2006. The ZKM Klangdom. *Proceedings of the New Interfaces for Musical Expression – NIME 06*. Paris: IRCAM – Centre Pompidou in collaboration with MINT/OMF, Sorbonne University, 140–3.

Rutz, H.H. 2004. *Meloncillo – eine graphische Benutzeroberfläche zur musikalischen Raumklangsteuerung mit Implementierung einer OSC-Schnittstelle zur Klangsynthese*. Master's thesis, Technische Universität Berlin.

Schroeder, M.R. 1962. Natural Sounding Artificial Reverberation. *Journal of the Audio Engineering Society* **10**(3): 219–23.

Treib, M. 1996. *Space Measured in Seconds*. Princeton, NJ: Princeton University Press.

Vaggione, H. 2001. Composing Musical Spaces by Means of Decorrelation of Audio Signals. *Addendum of the COST G–6 Conference on Digital Audio Effects (DAFX–01)*, Limerick, 6–8 December: Addendum 1–8.

Varèse, E. 1936. New Instruments and New Music. In C. Cox and D. Warner (eds), *Audio Culture: Readings in Modern Music*. New York: Continuum, 2006.

Wittek, H. 2002. *OPSI: Optimised Phantom Source Imaging of the High Frequency Content of Virtual Sources in Wave Field Synthesis*. Munich: Institut für Rundfunktechnik.