Multi-channel sound in spatially rich acousmatic composition

Dr Robert J. Dow

University of Edinburgh, School of Arts, Culture and Environment

Department of Music, Alison House, 12 Nicolson Square, EDINBURGH, EH8 9DF, UK

Telephone: [+44] (0)131 650 9352 Fax: [+44] (0)131 650 2425

E-Mail: R.J.Dow@ed.ac.uk

ABSTRACT

The acousmatic composer, whose music is fixed onto a recording medium and composed with the understanding that it's final reception will be achieved *via* playback over loudspeakers, is very naturally concerned both with the perception of (musically significant) space within his or her music, and the way in which his or her music is conveyed, and thus perceived, within a particular listening space. It is tempting to believe in sound reproduction's potential for fidelity: the possibility of literal sound recreation. For many, the ideal reproduction of an acousmatic work would be one which creates a transparency between the intentional sound and the experienced sound. In reality of course, in a group listening situation at least, this can never be fully achieved.

The desire for more spatial control has, however, led many acousmatic composers away from the normative practice of using stereo reproduction, to a renewed interest in a whole gamut of multi-channel sound reproduction formats. These include the deployment of multiple loudspeakers as multi-mono and/or multi-stereo arrays, standard formats such as Dolby 5.1, and more experimental formats such as ambisonics and wave field synthesis. Multi-channel formats may appear to give, *prima facie*, more scope for the composer to realise spatial detail and differentiation. Yet, as the disappointing development in, for example, much of cinema sound design illustrates, their naive use can produce the opposite effect: a lack of spatial clarity.

This paper aims to compare the use of a variety of common multi-channel formats in acousmatic composition, and examine their various merits, particularly with reference to current acousmatic performance praxis. The specific problems of the use of spatially rich material in this domain, for example environmental sound, and the difficulties such material presents in final reproduction will be discussed. Finally, a possible solution to these problems, using simultaneous multiple techniques, will be put forward.

1. PREAMBLE

It is, perhaps, unsurprising that for some composers involved with the creation of work using recorded sound there is a desire to 'fix' the performance of such work too. The process of composing acousmatic music in a studio environment is naturally reflexive, giving rise to both an impression of working directly with sound, and a feeling that 'what you hear is what you get'. Furthermore, the final product of working in the studio is something that is 'fixed' onto a medium. In the case if digital media, the sound is represented by an array of numbers which may be copied exactly again and again: the composer's intent has been formed into something absolute, a carefully crafted original capable of being cloned, or so it would appear.

Sound reproduction has a *prima facie* potential for fidelity: literal sound recreation within a specific performance space. The ideal reproduction of an acousmatic work is, for many, one which establishes a transparency between the intentional sound (as presumably experienced by the composer in his or her studio) and the actual sound experienced by the audience. The current renewed interest in multi-channel reproduction is driven partially by a desire to create this transparency.

However, as James Lastra has pointed out, every sound is:

> '[...] spatio-temporally *specific* or in a broad sense of the term, *historical*. Given that a sound is inseparable from the time and space of its production, each sound becomes an essentially unrepeatable *event*—an event distinguishable from all others.' [1]

Indeed, the well known existence of an audience 'sweet spot' within a particular performance space implies the co-existence of usually many more less ideal listening positions. This in turn indicates that during any performance of a specific piece of acousmatic music, there will be a limitless range of different 'versions' simultaneously broadcast to the audience, which of these they hear dependent on their listening position within the auditorium.

Moreover, the fixed representation of the composition on, for example, tape or disk, is mediated during its performance by the particular occurrence of both its reproduction technology and its reproduction space. Various characteristics of the technology used for reproduction will be imparted on the perceived sound, and indeed equipment is often selected because of its particular characteristics, to conform to a particular taste or set of normative practices. Of course, this is most striking where the equipment is not of sufficient quality to allow it to reproduce important, structural events within the composition, but far more subtle effects can change the audience's perception of the music.

The acoustic of the auditorium can have an immense effect on the reception of reproduced sound, particularly with regard to recorded spatial detail. Importantly, there is often a tension between the juxtaposition of virtual (intended) space and real (performance) space and the movement of sound objects within them. This is frequently most noticeable when a source with a high degree of recorded reverberation (either real or artificial) is reproduced in a highly reverberant space, which is one good reason against the surprisingly common practice of bathing the whole of a completed acousmatic composition in some degree of artificial reverberation to create an impression of cohesive virtual space. However, at the other extreme, a sound event recorded close to its sound source in a dry space, will sound distant and reverberant if respectively reproduced and auditioned at opposite ends of a large cathedral.

2. PERFORMANCE

The performance (diffusion) of an acousmatic composition is the act of its realisation. There is no original: the aspiration of diffusion is to recreate the acousmatic work for a specific spatio-temporal instance rather than reproduce a copy of it and therefore, as Smalley has pointed out, 'this final act becomes the most crucial of all.'[2] Viewed purely technically, diffusion offers the possibility of continuously changing the dynamic, volume (many versus fewer speakers), sound colouration and spectral range (through the use of different sorts of loudspeakers), spatial placement and spatial articulation within the performance space in real-time. How well any of these can be modified depends, of course, on the diffusion system being used.

It is worth re-emphasising at this point that diffusion is not simply about forming the best presentation of something absolute and fixed. Diffusion is performance, and as such does not merely come down to a glib desire to enact the composer's intentions, or a bringing into relief the music's apparent musical structures. Performance, including acousmatic performance, of course has to be analytical in the sense of divining and reacting to these intentions and structures, but importantly must also be intuitive, and to an extent, spontaneous: a creative force in its own right. As Nicholas Cook expresses it, the playing off of these instinctive and analytical approaches to the music is 'almost like creating a spark through the juxtaposition of two opposed electrical poles' [3].

The diffusion artist can take risks, playing with the audience's expectations and indeed may even choose to ignore suggestions of diffusion tactics tagged in the music by the composer. For instance, short, energetic, fast moving gestures such as appear at the beginning of Pete Stollery's *Shortstuff* (1993), would automatically suggest to most acousmatic performers rapid, energetic articulation of these gestures within the three dimensional performance space, emphasizing their natural shape and trajectory to create a feeling of excitement and energy. As an experienced diffusion artist himself, Stollery could almost have an expectation that this would occur, and it is worth noting that *Shortstuff* is unquestionably a work that can seem very flat in the studio and very alive if well performed in the concert hall: it is a work that requires performance. Yet, other approaches to a performance of *Shortstuff* may simply ignore these expectations and may decide to, on occasion, even fight the obvious structure of the work. These interpretations can reveal other, perhaps less obvious, facets of the music, bringing a deeper understanding of it through the performer's intervention.

3. FORMATS

In contemplating different playback formats, it should be remembered that the needs of the sound recordist are not necessarily those of the acousmatic composer or performer. It is not common for acousmatic music to merely try to reproduce a known sonic experience; although indeed this may be part of artistic intention. This is not to belittle the problem of producing good recordings: it is often takes more skill to create a sense of reality than unreality.

There are numerous playback formats available to us today, all with their own merits and demerits and it is not the intention of this paper to explore them all. The formats discussed below are those in common use by acousmatic composers today.

3.1 Loudspeaker Two-Channel Stereophony

Even with the emergence and re-emergence of a plethora of multi-channel sound reproduction formats, vanilla 2-0 stereophony (or simply, 'stereo') is for many the format of choice for the composition of acousmatic music. This is no doubt partly to do with the fact that high quality stereo reproduction facilities are relatively easy to set-up and cost effective compared with their multi-channel equivalents. However, it is also important to note that although there have been a number of experiments conducted regarding the performance of multi-channel acousmatic works, current acousmatic performance practice is still firmly based on stereo reproduction.

Good stereo diffusion systems under the hands of good performers can circumvent some of the well known problems of non-diffused stereo: for example, it offers the possibility of more audience cover (larger 'sweet spot') with fewer 'spatial holes', there is more scope for real changes of spatial perspective (the sound is not limited to one simulation plane) and the sound may envelope and be moved around the audience.

However, it is worth reiterating that the intention of stereo diffusion is not to create a more perfect or an enhanced stereo, but is to essentially 'recompose the acousmatic work for a particular spatio-temporal situation.' [4] Stereo diffusion is most effective where a work has been conceived with stereo diffusion in mind, and where the use of space is approached in an 'organic' rather than an 'architectonic' manner, to use Jonty Harrison's distinction. [5] The absolute placement of sound objects within their associated composed spaces is not important here. Rather, it is the dynamic relationships and interactions both between individual sound objects and between sound objects and their *perceived* spatial context which is significant, this perceived spatial context comprising the superimposition of both the composed space and the actual listening space (the virtual and real space respectively).

For the acousmatic composer, therefore, the limitation of stereo, even diffused stereo, should not be seen so much in terms of inadequate spatial fidelity, as it so often is. What stereo cannot offer is the articulation of particular sorts of relationship, both between discrete sound objects and discrete sound spaces and between sound object and sound space. For example, controlled spatial separation of contrapuntal material which is fairly similar in nature can be difficult or impossible to achieve. One of the many 'tricks' of composing acousmatic music for diffusion, is to create spectral and/or dynamic separation of sounds to facilitate their spatial separation during diffusion.

In spatially rich acousmatic music, the use of only two speakers (particularly in a typical performance space) is often not enough to convey the spatial sumptuousness of the sonic image, yet the distribution of this image over multiple loudspeakers does not necessarily lead to the creation of a coherent, but simply larger image. Instead, the multiple images have a tendency to distort and muddy the overall image, creating a perceptual quagmire.

3.2 3-2 Stereo (5.1-Channel Surround)

The increased interest and availability of 5.1 systems for home use has naturally led to some interest in this format by acousmatic composers. The hope is, it is assumed, that such a surround sound system will offer greater flexibility for spatial articulation in situations where previously one might expect only stereo to be available. Important too, is the existence of standard versions of 3-2 stereo, which in theory can give the composer specific expectations of any properly set-up array.

5.1, designed mainly for cinema and home cinema use and therefore with a primarily frontal sound stage in mind, uses conventional 3-0 stereo (Left, Centre and Right) with the addition of two rear/side channels which were originally intended to support ambience, 'room impression' and effects. [6] The centre channel allows dialogue to be properly focused behind the cinema screen and does so without the timbral distortion that a centre image in 2-0 stereo can produce. In theory, it could also allow the creation of a wider front sound stage, plugging the 'hole in the middle'. However, standard 5.1 surround has the Left and Right speakers located at ±30° for compatibility with two-channel stereo and thus does not make use of this possibility.

The angle of the 'surround' speakers (±110°) makes their projection more characteristically side as apposed to rear in nature. The angle of the speaker is set thus to give a compromise between the surround sound being both lateral enough to envelope the listener effectively, and being rear enough to allow effects to be panned behind him or her. It is worth noting that even when additional, paralleled surround speakers are used, as allowed in the ITU standard (to provide sufficient surround cover in larger spaces and distributed evenly between ±60° and ±150° when used), in standard 5.1 surround there are no loudspeakers set directly behind the audience.

For the acousmatic composer, 5.1-channel surround has a number of serious flaws, particularly where spatially rich source material is being utilised. The 5.1 standard was not conceived with accurate 360° imagining in mind, and the uneven distribution of the speakers (and again, the provision in the standard for supplementary, paralleled, surround speakers) makes this difficult to achieve. There are, of course, various novel microphone techniques that have been proposed for making recordings in a format more or less suitable for 5.1-channel surround reproduction, although unless an ambisonic approach using a Soundfield microphone is used, it should be remembered that none of these is particularly portable, making environmental and other location type recordings almost impossible. [7] Generally however, imagining in standard 5.1 arrays is unavoidably going to be at its best between the front loudspeakers, fairly poor at the rear and decidedly variable at the sides. [8]

Such techniques may be suitable for the reproduction of relatively stable sound images (recordings of music, film sound, *etc.*), particularly where most of the relevant sound information is frontal, but for the acousmatic composer who both wishes to process these recordings and manipulate their spatial qualities, multi-channel recordings intended for 5.1 playback are exceptionally difficult to work with. Instead, what many composers working in 5.1 rely on is source material based on trusty two-channel stereo or even mono recordings, which are then spatially manipulated with varying levels of success in the 5.1 arena using panning laws and artificial reverberation.

If two-channel stereo signals are used, the centre speaker is frequently seen as a problem. In this case, in terms of imagining, there is no actual recorded signal to drive the speaker, so the centre speaker either becomes redundant or its signal has to be derived in some way from the left and right channels (although this may generate good results if executed properly, yielding, for example, a more stable central image [9]). Furthermore, this speaker's central position necessitates its careful use: it can become easily too prominent, and as much as it can enhance the frontal stereo image, it can also destroy it, particularly if driven by a decorrelated signal.

If the centre speaker is essentially ignored, the composer is left with a rather non-uniform quadraphonic set-up, which is unfortunately how the 5.1 format is used by some composers. The irregular speaker array shape and wide side speaker angles, makes uniform stereo panning basically impossible, and this may tempt composers to treat sound objects monaurally, leading to their perceived movement as being very much on the peripheries of the 5.1-system, their sound often locating itself, as it were, 'in' the speaker.

In terms of performance, 5.1-channel surround presents similar problems to other multi-channel set-ups (for example, octaphonic). Since there is usually no real, coherent 5.1 sound image being created, it cannot be sensibly manipulated in the same manner as a two-channel stereo image in a multiple stereo array. Composers who are both interested in 5.1 and sound projection are inclined to diffuse the frontal left/right image as normal, leaving the centre and side/rear speakers fixed for those 'special effects' which cannot be easily conveyed through diffusion (for example, circular or random motion and genuine front/rear separation), enhanced surround ambiance, and creation of a central source discrete from the rest of the diffusion system where this makes sense (for example, recorded text and other vocal material).

Finally, it is worth noting that the LFE (Low Frequency Effects) channel, which is band-limited to 120Hz, has really no place in acousmatic music composition, and certainly not in its performance. As the five other channels are all full bandwidth, there seems little point in utilising this channel at all, particularly where a proper bass-management system is used, which is *de rigueur* for professional performance systems.

3.3 Octaphonic

Recently, there has been a renewed interest in 8-channel playback systems, doubtless stemming at least in part from the availability of cheap 8-channel+ sound cards, and the immediacy that software such as MAX/MSP, pd and supercollider bring to the creation of multi-channel signal and spatial processing. The most common speaker configuration is that of a uniform circle around the audience, the speakers spaced evenly at 45° intervals, either with both centre and rear-centre speakers (*i.e.* placed at 0° ±45°, ±90°, ±135°, and 180°: sometimes known as the 'double diamond' configuration) or the same configuration rotated by 22.5° to give a flat front (*i.e.* placed at ±22.5°, ±67.5°, ±112.5° and ±157.5°: sometimes know as the 'four pairs' configuration).

The current acousmatic repertoire is not limited to this uniform format, however, and indeed the variety of extant 8-channel format variations does require the composer to be quite specific about speaker layout and playback expectations in his or her performance instructions! Jonty Harrison's *Streams* (1999), for example, uses four, two-channel stereo loudspeaker pairs set in different spatial planes ('main', 'wide', 'rear' and 'distant'). This enables the opportunity for the sound image to flow between different stereo perspectives and allows easy integration with existing two-channel diffusion rigs, thus creating some performance potential. The key objectives of the regular 8-channel array, on the other hand, are to permit both straightforward, evenly balanced panning of sound objects around the audience and uniform sound distribution, allowing the audience to be fully enveloped.

Octaphonic playback systems do offer a certain amount of flexibility and their use can be approached in various ways. The double diamond configuration can be considered to consist of multiple 3-0 stereo planes, the centre speaker of each compensating for the relatively large angle of the 'Left' and 'Right' speakers. In the simplest scenario, this could comprise a front and rear 3-0 stereo systems with side fill. However, there is also the potential at least, to enable both the projection of manifold, simultaneous stereo images, and relatively smooth three-channel stereo panning.

The four pairs layout appears, to an extent, more like a standard two-channel diffusion system, with front and back pairs and two side pairs. Two-channel stereo panning can be easily achieved, although the image width is really rather too narrow too produce good stereo imagining. However, 2-0 stereo source material may be treated more as it would be in a two-channel diffusion system, particularly if the regular array is warped somewhat, adjusting the angle of the front and rear pairs to give a better image (tending more towards ±30° and ±150°), and adjusting the front and rear side pairs so they act more like 'wide' and 'side' speakers. A wider frontal image can therefore be achieved by combining the 'main' and 'wide' speakers, and image movement can be realised from front to rear *via*. the 'side' speakers.

For the performer of acousmatic music, the diffusion of 8-channel pieces, unless they are designed with some sort of non-standard playback in mind (for example, again, *Streams*) is difficult in ways similar to those described for 5.1-channel surround, although at least there are not those difficulties associated with the irregular shape of the 5.1 layout. Experiments using multiple octaphonic arrays (as an extrapolation, it is supposed, from the use of multiple two-channel speaker pairs in many diffusion systems) have been inclined to concentrate on the movement within the listening space of framed, circular, and to an extent, more varied sound trajectories: there is no real perception of a sound image, with its associated, composed space, being articulated in the listening space. Indeed, it is very hard to create an illusion aural solidity which successful diffusion requires, particularly if the music investigates spatially rich material, using this type of system.

3.4 Ambisonics

The approach of ambisonic recording and playback systems is to sample and reconstruct sound wavefronts at a point. The significance of this is that ambisonic reproduction is both homogeneous (no direction is treated preferentially) and to an extent coherent (the image remains stable if the listener changes position within it). It can produce a true 3-dimensional sound image given adequate recording and reproduction means. There are a number of microphone techniques available at least for first-order ambisonic recordings. The simplest of these is to use a specialist Soundfield microphone, which produces a three dimensional signal in what is known as B-format, which comprises four components (labeled W, X, Y and Z). The advantage of the Soundfield microphone is that a portable version is available, which makes field recordings of environmental and other spatially rich sounds which are generally impossible to bring into the studio, completely feasible, although of course a portable 4-channel recorder is also necessary to record the B-format signal (assuming that all 3-dimensions are required).

To reproduce the wavefront, the B-Format signal is decoded using a matrix specific to the speaker array in use. This is normally a symmetrical array of 4 or more speakers in pairs, although it is in theory possible to decode the signal for non-symmetrical arrays, such as for a 5.1-channel surround system. If the height component is to be decoded, then speakers are also required above and below the audience. What is particularly useful is that the composer can produce a work in B-format, which can then be decoded on any particular ambisonic set-up, and of course this includes the ever prevalent 8-channel array, be it in double diamond or four pairs configuration. This is particularly exciting for composers using spatially rich material, as not only can ambisonics recreate the complete experience of being in a particular place, it can do this in a manner which is not completely dependent on a particular reproduction set-up.

Unfortunately, ambisonics does have a number of serious limitations, particularly in terms of acousmatic music performance. Since it recreates the wavefield at a particular point, the sweet spot is particularly small, although various methods can be used to make this less problematic, and indeed the off-axis experience is not very disturbing, although not exactly ideal. The image is fixed, that is to say, there is possibility with existing systems at least, to move images around in real time and so to diffuse an ambisonic work.

In terms of the first problem, higher order ambisonic systems, which have additional directional components, can create improved directional encoding covering an increased listening area on reproduction. However, there is no microphone designed as yet that can actually record in even second order ambisonic format, which is a difficulty for those interested in using spatially rich recordings. Composers using second order ambisonics (such as Natasha Barrett) synthesize these signals artificially from either B-format or stereo recordings. For example, just as in wavefield synthesis (see below), virtual stereo sound sources may be created within the ambisonic sources (imagine using a soundfield microphone to record a stereo sound emitting from two loudspeakers in a particular space).

3.5 WaveField Synthesis (WFS)

WFS is a relatively new contender in the world of multi-channel sound formats, and is causing much excitement at the moment, particularly in the realm of cinema sound. Like ambisonics, WFS also attempts to recreate a particular wavefield, but it does so over an area rather than a point. Its reproduction system utilises an array (which can be linear or circular) of small loudspeakers which act as point sources and in combination recreate the wavefield. WFS does not suffer from the same problem of limited coverage that ambisonics does. However, it is essentially impossible to record the wavefield for a given space in the same easy manner that is for ambisonics using the soundfield microphone, given the size of sampling required.

Instead, current approaches to WFS: a) record the direct sound of each sound source (*i.e.* treating them as a point source using a directional microphone), b) record the early reflections of each source and c) record the reverberation of the sounding space. These are then 'combined' at reproduction to give the impression of the sound in a particular space. Sounds can be localised within the reproduction space extremely well, although there can be problems if the sound source is actually located more or less exactly where and audience member is sitting: in this case the sound feels as though it is actually within the middle of ones head, which can be very disturbing. There are systems being developed that allow source sounds to be moved in real-time within the space, or over programmed trajectories.

The main problem for the acousmatic composer, who is of course extremely concerned with the intrinsic spatial character of his or her source material, is that this spatial character cannot really be recorded, it can only be synthesised. It should also be realised that there needs to be as many recording channels as there are individual point sounds to be manipulated in space, although again as with ambisonic recordings, it is possible to create virtual stereo (or indeed 5.1 or other multi-channel formats) within the wavefield. Thus is would be possible to have stereo recordings being emitted by, as it were, pairs of virtual loudspeakers.

4. CONCLUSION

The complex nature of sound within spaces, and the limitations of current technology, does not permit the ideal recording and reproduction of sound. Sound reproduction is a comprise, and because of this, the acousmatic composer cannot expect full spatial control once his or her work is no longer heard from presumably the ideal listening position, his or her seat in the original production studio. Diffusion, even in these heady days of refined multi-channel loudspeaker systems, still has an important role as mediator between artistic inception and reception.

Of the formats investigated in this paper, stereo still appears to offer the best compromise between being able to represent spatial information over a relatively large area, while allowing real performance opportunities. Ambisonics and WFS are still not quite able to offer this, but in future may offer much more.

5. ACKNOWLEDGEMENTS

The author would like to acknowledge the Arts and Humanities Research Board in the funding of this research. He is also indebted to Dr Joseph Anderson, University of Hull, for his help and advice.

6. REFERENCES

1. Lastra, J (1992) 'Reading, Writing and Representing Sound', In Altman R (ed.) *Sound Theory, Sound Practice*, 65-86. New York: Routledge
2. Smalley, D (1986) 'Spectro-Morphology', In Emmerson, S (ed.) *The Language of Electroacoustic Music*, 61-93. London: Macmillan
3. Cook, N (1999) 'Analysing Performance and Performance Analysis', In Cook, N and Everist, M (ed.) *Rethinking Music*, 239-261. OUP: Oxford
4. Dow, RJ (2003) 'Sound Diffusion and the Sonic Image', *Diffusion* September 2003, 2-6
5. Harrison, J (1999) 'Imaginary Space—Spaces in the Imagination', In *Proceedings of the Australasian Computer Music Conference, 1999; Victoria University of Wellington*, 7-15
6. For a more detailed description of 3-2 stereo, see for example, Rumsey, F (2003) *Spatial Audio*, 86-94 Focal Press: Oxford
7. See for example, Theile G (2001) 'Multichannel Natural Music Recording Based on Psychoacoustic Principles', In *Proceedings of the AES 19th International Conference June 2001*, 201-229
8. Rumsey, F (2003) *op. cit.* at 189
9. See for example, Gerzon, M (1992) 'Optimum Reproduction Matrices for Multispeaker Stereo', *J. Audio Eng. Soc* **40**:7/8 (July/August), 571-589