REPORT ON A MEDIUM-SCALE THREE-DIMENSIONAL ARTIFICIAL SOUNDSCAPE RENDITION: RESEARCH AND DEVELOPMENT SYSTEM

Peter Lennox

1. INTRODUCTION

This is a report on the "first time out" of an experimental approach to the use of spatial sound for medium-scale use that takes advantage of recent technological developments. The intention is to utilise artificial sound in ways that were previously prohibitively expensive and logistically cumbersome. Inevitably, early stages tend to be heavily focused on technological challenges, funding issues and production concerns. I've included a section (below), reprinted with the kind permission of the editors of the Proceedings of the Maxis Event (http://www.maxis.org.uk), which outlines some of these.

This should not obscure the deeper-seated aims of the project, however, and I'd like to bring some of them to the fore here. This is not primarily a sound system, even though it has 40 speakers and amplifiers and over 800 metres of cabling. Instead, I prefer to think of it as an 'audio metaphor projection system'; it is a concrete realisation of aspects of metaphysical debate about what it is that we know (or can know) about the world through hearing, which in turn is part of the larger (and long-standing) cosmological debates of "who are we?" and "what is the universe made of?" and "what is our place in it?" These are questions that will never, in my view, be answered 'fully', yet we feel compelled to repeatedly attempt the impossible. I'm of the position that humankind's attempts to copy or recreate aspects of the world can be seen as specific examples of those more general philosophical questions, rather like a child taking apart and reassembling (or attempting to) some device in order to understand what goes on beneath the surface.

In this context, then, the reasons at the heart of attempts to artificially manipulate sound – whether musically intentioned or not- are ineluctably bound up with attempts to understand the real world-and-our-place-in-it. We manipulate aspects of the physical world not only as tool-users trying for specific physical outcomes, but in the longer term we are striving for unspecified outcomes.

My own interest is in non-linguistic sounds, and this goes deeper than just saying "music without words"; much of our understanding of the shape of the world seems in some way more primitive and fundamental than simply locating and identifying things. I would like to inquire into the nature of the primitives of auditory perception, and these seem inextricably linked to spatial matters. A significant problem here is in the need for articulating the appropriate questions; much thinking about spatial matters relies heavily on visual metaphors, and our linguistic structure reflects this. I have previously mused on the question of whether we ever might have arrived at Euclidean spatial descriptions if we as a species had never evolved sight; of course it's a moot point.

The upshot of this is that I have, with Scott Hawkins' help, assembled the physical components of an audio metaphor projections system with the aim of making it available to artists and scientists to 'play' with; the outline description is below.

2. Before the event

Following on from the Ambisonix dance night series (which used 4 main stacks and 8 peripheral speakers in dual-concentric layout) a system is being assembled for experimentation and display of artificial soundscapes. The system comprises 32 speakers (Blueroom 'Minipods') at discrete locations around the surface of a nominal sphere, or part thereof. These are controlled by a PC with 2x Sydec Soundscape sound cards and proprietary software courtesy of Bruce Wiggins and the University of Derby; a hardware 1st order decoder controls the 8 horizontally distributed Fane subs.

The radius of the sphere can be up to 20 metres, though in practice this may often be nearer to 10 metres, depending on local facilities. Generally, this virtual sphere will be used in truncated form, with the floor of the listening area intersecting the lower part, resulting in 50% to 60% of the sphere's volume being generated.

The estimated maximum comfortable capacity is likely to be 200 persons. The intention is to produce and control the salient spatial features of the sound field within this listening area, for greater than 90% of the listeners, which in practice means that spatial 'accuracy' must be discernible at any position within approximately 70% of the part-spherical sound field.

One of the research aims is to eventually closely specify what is meant by "spatial accuracy" in large-scale listening circumstances such as these, for a wide variety of expert and non-expert listeners, and to look for variations between such groups. From this, and in accord with a developmental aim, the intention is (funding permitting) to feed results back into system development in an iterative fashion, to improve control-and-display of "spatial accuracy". This is facilitated by the nature of the speaker-feed decoding software, which inherently is capable of running several different coding regimes (mono, stereo, quadraphonic, 5.1 and ambisonics, for instance), simultaneously. Comparisons between codecs, and more importantly, optimised hybrid codecs, can be auditioned and compared with minimal delay.

The other significant aim in designing this system is facilitation of collaborations with composers who have a significant interest and/or expertise in the area of "spatial sound fields" - whether as music or sound effect (such as accompaniment for televisual material). This requires audiences, of course, and so the system must be constructed with health-and-safety and aesthetic considerations properly attended to. Demonstration '3-d sound' material has been kindly contributed by composers, engineers and production companies from around the world.

Considerable interest has been expressed in this aspect of the system's capabilities, from organisers of festival and conference events; nevertheless this remains a highly specialised niche, largely consisting of professionals in use-of-sound applications. Because the system is intrinsically scaleable, down to quite small listening rooms, there is scope for flexibility in proposed usage. The system will be first demonstrated at the "Maxis" event in April 2002, at Sheffield Hallam University; a symposium of about 200 people, dedicated to the experimental uses of sound. Submissions include a variety of multi-channel audio formats [EG multiple discrete feed, multiple stereo, Quadraphonics, Dolby 5.1, Octophonics, Ambisonics 1st, 2nd and 3rd order], which the system comfortably accommodated. Discussions are in progress with several other specialised-event organisers. Nevertheless, this is a research-and-development initiative with a stated aim of pushing back boundaries beyond what is considered feasible in a business sense; this system will not be commercially available in wider markets.

Technically, the basis of the unusual flexibility which this system enjoys is the principle whereby data-channels are uncoupled from speaker-channels. This is so because "3-dimensionality" can be expressed mathematically and hence 3-d manipulations can be carried out comfortably by current generations of processors available in off-the-shelf computer technology. This mathematical system, and the technology which realises it, is called "ambisonics", and was invented by Micheal Gerzon in the 1970s; however, software realisations could not be extensively explored until quite recently, with the maturation of computer capabilities. More recently, the mathematical bases for various different approaches to spatial sound rendition has been investigated and has resulted in considerable convergence; even where technologies are based on quite different psychoacoustic principles, 'interchangeability' is becoming feasible.

2.1. 'Accidental Properties'

There are several properties of this type of system which, though not originally intended, have been observed by workers in the field. One is the issue of system headroom, or practical dynamic range of the system-as-a-whole. In a conventional public address system, where moderate to high sound pressure levels are sometimes required, many components contribute 'compression' artefacts when operating near the upper boundaries of their performance design. These components in a complex system can sum in unpredictable ways, resulting in theoretically sub-optimal system performance. The type of system under construction here, which is a development of previously realised 12- and 16-channel designs is well known to display a dynamic range which appears to be beyond the capabilities of the individual components. That is, distortion-free performance is apparent at sound pressure levels not envisaged by component designers.

Another opportunity arises out of the nature of the codec: low-frequency management can be individually attended to within the central codec, to optimise SPLs and spatial depiction. Whilst considerable dissent is evidenced in the public-sound industry as to the value of directional accuracy of low-frequency sounds, we have found strong psycho acoustical evidence in support of the notion that mechanisms for high and low frequency directional hearing should agree where possible.

A final property which has recently been observed is that, under certain circumstances, the coherent rendition of 'distance information' is better than had been predicted, or has been observed in other systems. The robustness of this type of property is under investigation.

The system also offers improved support for visual material and especially for increased viewing angles such as in experimental cinema or even 'surround vision'.

There have been several other locally-observed effects, some of which are of interest to those who have kindly given support to this project.

3. After the Event

Thanks to all who exhibited at Maxis -I heard some fantastic things.

3.1. Criticisms:

The system displayed several failings, which loosely fall into 2 categories:

- Design flaws - these could be called 'philosophical flaws'

Here I would place the notion that an 'environment' or 'place' (as I prefer) can be adequately represented in terms of the signal energy characteristics sampled at a particular location in that place - the 'energetic sphere', as it were. This incorporates the notion that, for any particular percipient at any particular time/location, the sound received is essentially in the form of plane-waves which propagate and are at right-angles with respect to the sounding object. In other words, if you face the sounding object, the sound will reach both ears at the same time and at the same amplitude, all other things being equal. This is the basis of stereo, and is also the basis of ambisonics. But it's an idealised case, and not generally wholly true in the 'real world', which contains occluding objects (even the body of the sounding object can fall into this category) and so on. This often means that the shortest signal path is quite different in length, and even angle of arrival at the perceiver, for different frequencies. The fact that our spatial hearing is so robust under such (apparently) chaotic circumstances tells us two things : 1] – that inter-aural differences theories don't tell us all there is to know about spatial hearing, and 2] that, in the conceptually complex and 'unsimplified' real world sound-fields, where signal paths are all over the place, with different frequencies arriving from different angles, at different times and at different amplitudes, nevertheless, order outweighs chaos (perceptually speaking). Rather than chaos, perhaps 'complexity' is a better word, then.

- Technical flaws i.e. sub-optimal realisation of existing design.

A number of issues:

a) The LF (low frequency was laterally reversed with respect to the mid/top– thanks to John Vaughan who spotted this, Dallas Simpson who confirmed it.

b) The 'up-down' axis was 'squashed' (due to construction difficulties) which led to a loss of some of the feeling of height; surprisingly this also caused some lateral homogenisation as well. We did manage to experiment with a workaround that improved matters considerably, but had no way of using this method during performance.

c) System latency - I know this was a problem for several, and quite severe for some (sorry, matt) - this just requires a complete redesign of PC architecture (especially PCI bottleneck issues), along with faster and more powerful machines - so it's philosophical as much as technical, at the moment!

There are several other ways to express criticism of this system, for example the lack of 'speaker-ishness' that is inherent in ambisonics is nevertheless disturbing for those who prefer more discreteness. Ambisonics provides more discrete images than there are speakers (as does stereo, but in this case all 3 dimensions) but is actually less accurate when trying to produce an image where a speaker actually *is*; there's a hybrid way round this, using a discrete feed to a target speaker and a near simultaneous (within <12ms) encoding to ambisonics.

3.2. What was achieved?

Technically, a number of valuable lessons were thrown up in this exercise, which I won't go into here. Practically, though, the fairly obvious lesson of the need for such a complex system to be transparent as possible in use was reinforced. It was apparent that, for many first-time users, there was an expectation that the user would have complete responsibility for what programme material is determined to appear at which speaker (of which there were 40 in total). Fortunately though, this system is fundamentally a great deal simpler to use than that, even with the prototype software interface that was used. Nevertheless, it is *conceptually* quite difficult to move from thinking about "channels" as speaker-feeds to thinking about "channels" as contributors to a total sound field.

In the longer term, in my view, it will become quite normal (and necessary) for the majority of users to work within a spatial representation-environment that does not have an exact and close-coupled relationship with the actual specifications of the physical replay system. I suppose an analogy would be to point out that driving a car does not entail making moment-by-moment decisions about valve-timing, firing rate, fuel mixture etc; these minute decisions are subsumed beneath an apparently simple control surface which appears to have a direct relationship with what we actually want to be in control of, in this case the acceleration (or deceleration) of the car. To achieve this apparent simplicity whilst retaining complex and minute (but vital) decision structures requires the development and fine-tuning of a comprehensive metaphor-enactment system. This is what my own work is concerned with; I would like artificial sound to begin to approach the sublime subtlety that exists in 'the real world' (which, I'm ready to admit, is overlooked most of the time by most of us).

Most of all, I was exploring the question of whether this kind of approach to spatial sound production might be an early candidate with which to address what I consider to be an increasingly pressing concern for composers and "virtual reality" developers alike: in sound terms, what makes a "place"? The question asked by this system is "how near 'place-likeness' can we get with contemporary sound field science and technology?" and this question applies equally to artificial place-making as much as it does to the artificial reproduction of real places.

Results so far are unsurprisingly inconclusive; I did get frustratingly brief glimpses (forgive the visual metaphor!) of 'place-ness', most notably in the artificial places composed by Jan Jacob Hoffman (http://www.thisplay.com/ - click on "sonic architecture link towards the left of the screen, also pictures of the system here: http://www.sonicarchitecture.de/en/events_klein_april_2002.html) in his Sonic Architecture series of 2nd-order ambisonic compositions, and in the 1st-order recordings of music in Lincoln Cathedral (a real place) kindly supplied by Dallas Simpson of Serendipity UK. For tantalising moments it was possible to hear 'past' the sound system, and listen in a manner similar to listening 'in the wild' as it were. To put this into context this virtually never happens to me in 'critical listening mode' in any system I hear; I have had such experiences more often listening to high-quality hrtf binaural material over headphones, so I've an idea what I'm listening for. Those brief moments of "protorealism" cost thousands of pounds and hundreds of man-hours, and in the overall scheme of things represent a tiny step forward when compared to the advent of wax-cylinder recordings many years ago; nevertheless, as a politician who was clearly not a botanist once remarked, "great elms from little acorns grow".

4. With thanks to:

Supporting organisations:

Funktion One Ltd,. Blast Loudspeakers, Blue Room Loudspeakers, Fane Loudspeakers, Soundfield Research Ltd., University of York, University of Derby, Sheffield Hallam University, Serendipity Ltd and Yorkshire Arts Association.

Collaborating Sound Artists:

Massimo Carlentini, Ewan Stefani, Wayne Jackson, Joseph Reinsel, Pete Kelly, James Croson, Mark Kirschenmann, Charlotte White, Kim Cascone, Phil Ogg, Scott Hawkins, Malte Steiner, Huw McGregor, Carolyn Downing, Jan Jacob Hofmann, Robert Mackay, Tae Hong Park, Brigid Burke, Scott Smallwood, Terre Thaemlitz, Team Doyobi, Smyglyssna, Kaffe Matthews, Massimo, Robert Lippok, People Like Us, Scanner, Hecker and Haswell.

Special thanks to:

Bruce Wiggins for software development, John Vaughan and Dallas Simpson for their ears and technical help, and Scott Hawkins for fundraising, administrative stamina and vision.