Concerning the Perception of Sound in 3-Space

David Worrall

3-space

With the increasing sophistication of tools for capturing, generating and manipulating sound, has developed a growing interest in the use of 3-space[1] in sonic design of all sorts including electroacoustic music and soundscapes. For example, Karlheinz Stockhausen recently said:

> I think that all music will become space music and that space becomes as important as pitch in the traditional music, as durations and rhythm and metre and there is a very new development of harmony of space and I mean space chords, space melodies and that doesn't mean pitches, it means movement on several levels around the listener: above, below, in all directions.
> (Stockhausen 1977)

My own interest in 3-space as an aspect of musical composition began in 1981 with *Mixtures and Re-collections*, my first composition by direct computer synthesis. *Mixtures* is a quadraphonic composition which relies on the fact that 3-space acts as a strong stream segregator; sounds located at the same physical place at the same time are associated. (See Bregman 1990: 75-6). Three layers, identical except that they are transposed in pitch, are time compressed, and revolve around the auditorium at different velocities. What one hears is the spatial accretion of the layers into multiple musical gestures in different locations in the 3-space. A stereo mix of the work even further de-emphasises the original layering.

Later we developed a 16 channel composition/performance environment (Worrall 1989) which used a 'brute-force' approach to moving sounds in 3-space. Sounds for each of the loudspeakers, equally distributed over the surface of a hemisphere, were generated independently under MIDI control. These experiments have led to the construction of an ambisonic distribution system (Vennonen 1994) which is still being further developed (Vennonen 1995) and refined to include real-time encoding and decoding of multiple audio signals.

Developing a parametric relationship between 3-space and A-space

A primary concern during the development of a usable 3-space was, and still is, how to compose for it; how to integrate 3-space into different composing methodologies. I began, perhaps naturally enough, to think of 3-space as a parameter of the sound. Using the parametric approach for developing compositional materials as proposed by Pierre Boulez, Karlheinz Stockhausen and others, music is composed in A-space[2] as the evolution of ideas in the multidimensional parameter space of pitch, duration, loudness and timbre. 3-space is simply added as another parameter.

> There remains a fifth dimension, which is not, strictly speaking, an intrinsic function of the sound phenomenon, but rather its index of distribution: I refer to space. Unfortunately it was almost always reduced to altogether anecdotal or decorative proportions, which have largely falsified its use and distorted its true functions.
> (Boulez 1971: 66)

Boulez, speaking about ensemble writing, says:

> ... spatial distribution seems to me to merit a type of writing just as refined as the other sorts of distribution already encountered. It ought not only to distribute spaced-out ensembles according to simple geometric figures, ... it must order the micro-structure of these ensembles.
>
> It is obvious that the index of distribution, space, acts not only on the durations, but also on pitch, dynamics and timbre within the time-span; static or mobile distribution can be considered as maintaining certain relationships with all these interacting characteristics. These relationships are of greater subtlety than those of simple speeds--angular or lateral according to the spatial layout in use, even leaving out of account the *local* acoustical conditions, which seriously complicate the problem. ... It seems to me that the real interest in distribution lies in the creation of 'Brownian movements' within a mass, or volume of sound, so to speak; hence it is a question of elaborating a strongly differentiated typology of relationships, to be set up between the phenomenon itself, whether individual or collective, and its actual, absolute place in real space.
> (Boulez 1971: 66-7)

---

[1] Concrete space, or real space (hereafter called 3-space) defined by the OED as:
> *1. Continuous extension viewed with or without reference to the existence of objects within it;*
> *2. Interval between points or objects viewed as having one, two or three dimensions.*

[2] Abstract space, defined by the Penguin Dictionary of Mathematics as:
> *A formal system characterized by a set of entities, together with a set of axioms for operations on and relationships between these entities (eg metric spaces, topological spaces, and vector spaces).*

As Trevor Wishart has eloquently elaborated (Wishart 1996), this is essentially Platonic Idealism and thus fraught with difficulties, not the least of which is the way it imposes limitations (mainly derived from the limitations of notation) on the composer, especially in making works using transformation techniques on the sonic material directly. It has thus become a necessity to find more appropriate ways to explore the nature of 3-space, in much the same way as it has become important to find more appropriate ways to explore the nature of timbre space.

Similarities and intersections of timbre space and 3-space

Timbre is a difficult term and often defined by negation, as Begult suggests:

> The spectral content (spectrum) of a sound source, along with the manner that the content changes over time, is largely responsible for the perceptual quality of **timbre.** Sometimes referred to simply as 'tone color,' timbre is actually difficult to define. A 'negative' definition that is commonly cited for timbre is "the quality of sound that distinguishes it from other sounds of the same pitch and loudness."
> (Begault 1994:35)

Interestingly, he goes on to suggest:

> This could be extended to spatial hearing, in that two sounds with the same pitch, loudness, and timbre will sound different by virtue of placement at different spatial locations.
> (Begault 1994:35)

Both timbre and spatial cues depend on the morphology of the spectral modification (emphasising some frequency bands whilst attenuating others) of the sound source. Theile (Theile 1986) and others have suggested that it is not the stimulation of the eardrum that determines timbre, but rather the overall sense of hearing, a higher-level brain function that "identifies the timbre and location of the sound source". Clearly the two interact; modifications of timbre (including all its sub-dependencies such as fundamental frequency, amplitude, spectra etc.) affect our perception of location as well as proximity (distance), angle (lateral (left-right, front-back)) and azimuth (height) and, for moving sources, velocity.

In his overview article on music representation systems, Roger Dannenberg suggests that 3-space is an aspect of timbre.

> With timbre we are still learning what to represent. My explanation is that, starting with sound, we picked out the two things we understood, pitch and amplitude, and called everything else timbre. So timbre is by definition that which we cannot explain. As aspects of timbre are isolated and understood, such as spatial location and reverberation, these components come to be regarded separately, leaving timbre as impenetrable as ever.
>
> Taking a less cynical view, real progress has been made towards timbre representation. The classic studies by David Wessel (1979) and John Grey (1975) used multidimensional scaling and refined the notation of the timbre space. Although these studies represented timbre in terms of perceptual dimensions, others have represented timbre in terms of control dimensions, such as the set of parameters needed for a particular synthesis algorithm.
>
> The current practice is usually to represent timbre by name and number. Real timbre space has so many dimensions that it is often preferable to locate and name interesting points in the space. MIDI program change messages and the concept of 'instrument' found in many software synthesis systems are examples of this approach.
> (Dannenberg 1993: 23)

In-keeping with Wishart's warnings on the way notational concerns impact on our perception of the importance of different aspects of the multidimensionality of sound,

> It is notatability which determines the importance of pitch, rhythm and duration and not vice versa .... much can be learned by looki
> ng at musical cultures without a system of notation.
> (Wishart 1996: 6-7)

we need to be cautious in suggesting any hierarchical relationships between 3-space and timbre. In fact timbral modifications often suggest spatial modulations and this could suggest that spatial perception happens even later in perceptual processing than timbre, as it radically depends upon it. At best we can say, like most abstracted parameters of multidimensional spaces, they are inextricably linked.

Bearing in mind Wishart's appropriate cautioning about timbre:

> A catch-all term for all those aspects of a sound not included in pitch and duration. Of no value to the sound composer!
> (Wishart 1994: 135)

The instrument concept is useful because it affords 'toolness', and in our current development of an ambisonic performance environment, referred to earlier, we are exploring the MIDI sampler model as a way of representing 3-spatial movement by 'continuous controller patterns' which are stored (in the ambisonic spatialiser) as 'patches' in breakpoint tables (time-offset, angle, proximity). These tables can then be interpolated, compressed or expanded permitting spatial 'patches' with global time independence - i.e. the are available as spatial control parameters in real-time.

Functional approaches to hearing - the limitations

Most models of aural perception, firmly rooted in Platonic dualism/idealism, rely on a functional analysis of the physiology and psychophysiology of hearing. In contributing to the understanding of *how* we hear, psychophysical studies continues to be relevant to the sound designer, especially for tool-building. However, the role of functional studies such as those by Fletcher and Munson who mapped curves of equal loudness, is not, as I discuss later, the whole story.

In functional studies of the ability of humans to locate sounds in 3-space, the role of the ears' pinnæ have been found to be primary. These studies, mostly using broad-band noises and sine tones, are used to describe what are known as binaural head-related transfer functions (HRTF's). HRTF's can be thought of as maps of frequency-dependent amplitude and time differences that result primarily from the complex shaping of the ears' pinnæ. Each individual has a unique binaural HRTF although there are some generalizable characteristics. Begault's recent book (Begault 1994) contains a thorough summary of this research which has received added impetus from the growing interest in the role of audio in multimedia and virtual reality, as it's title attests.

Hearing compared with listening

Given the role of post-mechanical analysis in timbre and space perception, the <u>what</u> must at least be as important as the <u>how</u>. Whilst psychophysical studies are useful in describing how one hears these types of test noises in immerse environments, they are of limited value outside of them. More particularly, they tell us nothing of to *what* we are *listening*. Stephen Handel puts it like this:

> Listening is not the same as hearing. The physical pressure wave enables perception but does not force it. Listening is active; it allows age, experience, expectation, and expertise to influence perception. It is often helpful to illustrate how the ear is like a microphone or how the eye is like a camera. It is a mistake, however, to equate the ear with listening or the eye with looking, or to equate the faithful recording of sound energy or light energy with hearing or seeing. We hear and see things and events that are important to us as individuals, not sound waves or light rays. ... The study of listening must take place within the context of the environment in which listening *evolved,* since it is the product and reflection of that environment. (Handel 1989: 3)

Also, despite each ear receiving different auditory signals, the auditory 'image' is one of objects that appear coherent in the outside world and not in the ears or head. This is also true despite head and body movements, so the observer must dynamically compensate for the position and orientation of the head. It is clear, then, that a general theory of spatial cognition that can be used to integrate data from diverse disciplines must address the functions of the relevant perceptual and cognitive structures and processes.

The phenomenal or naturalistic perception of 3-space

The evidence used by phenomenal or naturalistic approaches to perception consist of one's conscious experiences. 'Naturalistic' means that the evidence concerns responses to whatever stimuli occur naturally within the environment; there is no attempt to modify these stimuli or create artificial ones. Probably the most radical and to a certain extent least formal work in this vein is that of James Gibson. Developed from Aristotelian empiricism, his approach (Gibson 1966) deals in considerable detail with certain aspects of 3-spatial orientation such as depth perception and what he calls an 'ecological approach' which focuses on the process of adaptation between organism and environment. Thomas Lombardo explores the evolution of this idea of reciprocity in detail. For Gibson,

> The structures and capacities of animals were described relative to their ways of life within an environment; in turn, the environment was described relative to the ways of life of animals. An explanation of perception involved a dynamic interdependency of animal and environment. Gibson's epistemology is direct realism, as was Aristotle's--the "object" of perception is the real world, viz., the environment. Perhaps the most difficult and unique point to grasp is treating perception as an ecological phenomenon rather than a mental or physiological event, yet Gibson's direct realism only follows if perception is defined ecologically. Perception does not reside in the brain or the mind any more than life resides in cells or in some inexplicable living spirit. Neither mentalism nor physicalism is correct. Perception, as well as life, is ecological; perception exists at the reciprocal interface of animal and environment within an ecosystem. (Lombardo 1987:5-6)

Gibson's early work was in the study of depth perception and is critically reviewed by Bruce Goldstein (Goldstein 1984). His ideas were developed with particular reference to visual perception and a thorough translation of them into the aural domain is needed. Early work is based on three main ideas; the ground theory of space perception, invariants, and direct perception.

According to the **ground theory**, information contained on the ground (usually horizontal) plane is a texture gradient. The elements that make up a textured surface appear to be packed closer and closer together as the surface stretches into the distance; there is more texture detail the closer the object is to the observer. This gradient results in an impression of depth, and the spacing of the gradient's elements provides information about the distance at any point on the gradient.

For sound , Gibson's texture roughly equates with reverberation, which causes the texture of a sound to be more indistinct the further away from the auditor it is. Along with reverberation, texture gradients share other depth cues such relative loudness (more distant elements of the gradients get softer) and frequency spectrum. A texture gradient extends over a large volume of 3-space and no matter where you move on the gradient, the elements of the gradient provide information that enables you to determine the distance between different locations anywhere else on the gradient. This is in-keeping with our known ability to perceive the difference between a soft sound emitted close to the listener and a loud sound emitted at a distance, even when they have the same amplitude at the ear. One consequence of this theory is that if we present sounds in different virtual reverberant spaces closely in time, our ability to locate their virtual proximity could be considerably impaired. However if different sounds are presented within a single virtual reverberant space, they will, because they will have different textures consistent with the overall texture gradient of the space, be more likely to be perceived as different sounds. This accords with the findings of John Chowning (Chowning 1971) who found the specification of two types of reverberation, both local and global, resulted in superior location detection of computer generated sounds.

When an observer moves relative to a texture gradient, the texture will be in constant flux to the ears, i.e. the contours that define the textures of the gradient change, but some information - the texture of the gradient - remains, in Gibson's terminology, **invariant** and thus the scale of depth in the scene remains constant. According to Gibson, it is this invariance information that we use in everyday perception as we move through the environment. Another example of invariance is provided by the way movement of an observer causes the textures of sounds in the environment to flow. This flow of the environment illustrates how, when the environment speeds past an observer who is travelling, there is flow of other sounds in the environment which is everywhere different to the way it flows from the location towards which the observer is moving. Since this point is at the centre of the flow pattern, it does not change (at least as much) i.e. it stays relatively invariant. Aurally, this change in relative invariance is caused by the front-back asymmetry of the body (including but not exclusive to the head) as mapped in an HRTF. Thus, to 'stay on course' while moving toward a sonic source, one only needs to keep the invariant centre of the flow pattern centred on the source.

A consequence of this theory is that it may be possible to create the illusion of a sound moving towards a listener by keeping the sound-ground invariant and changing the texture of the sound. To create the illusion of the listener moving towards a sound, a change in the texture of the ground is needed as well. I have yet to test this postulate - perhaps it is possible to create the illusion of moving backwards or sideways or even perhaps to dance!

Perhaps the most controversial of Gibson's ideas is his explanation of how an observer uses this invariant information. His answer is that we pick up invariants by **direct perception** and just use them. We are pretty good at perceiving the correct sizes of objects in the natural environment, even if these objects are located at different distances from us. Gibson feels that the information provided by invariants is present in a form that an object's size in our field of view *and* its distance from us can be perceived directly and immediately from an unconscious mental process of estimating how many units of a texture gradient are covered by the object without needing to be being processed in any way. Gibson's explanation works well when we can see both object and background. However, some researchers have noted that Gibson cannot explain how we can perceive movement when movement relative to the background is not visible, as in the case when a spot of light is seen in the dark or when an object is seen against a completely textureless background. In these cases, the background cannot provide the information we need to decide whether the object is moving or we are moving.

Summary and temporary conclusions

Timbre and 3-space have a number of commonalties. A functional approach to composition with both has some uses but suffers from its dependency, partly through notation, on static conceptualisations of the what are in reality extremely, multidimensional sonic spaces. This is especially the case for electroacoustic composition which frequently uses morphological techniques. For musical grammars to be flexible enough for electroacoustic composition, they need to be able to break, or at least loosen these dependancies so that they can account for the differences between hearing and listening.

Throughout the evolutionary history of any species, species-typical developments in spatial competencies have become adapted to changes in the ecological contexts encountered by the species over the course of the life-span; that is, organisms develop more sophisticated spatial orientation competencies to meet increasing needs for spatial activity. Naturalistic theories of perceptual psychology, based on observation in complex environments offer insights not offered by other methods.

3-space can be considered to be defined as much by its contents and what they afford us as vice versa. Gibson's ecological approach to analysis of what the environment is, combined with stream analysis, is likely to provide significant insights into our understanding of 3-space and it's use in sonic design. Perhaps it is fitting then to end this temporary conclusion with him.

> I am also asking the reader to suppose that the concept of space has nothing to do with perception. Geometrical space is a pure abstraction. Outer space can be visualised but cannot be seen. The cues for depth refer only to paintings, nothing more. The visual third dimension is a misapplication of Decartes's notion of 3 axes for a coordinate system.
>
> .... Space is a myth, a ghost, a fiction for geometers... For if you agree to abandon the dogma that "percepts without concepts are blind,"as Kant put it, a deep theoretical mess, a genuine quagmire, will dry up.
>
> (Gibson 1979: 3)

References

Begault, D.R.(1994) *3-D sound for virtual reality and multimedia*. Cambridge, MA: AP Professional.

Bregman, A.S. 1994 .*Auditory Scene Analysis. The perceptual organisation of sound*. London: MIT Press.

Boulez, P. (1971) *Boulez on Music Today*. London: Faber and Faber. Originally published as *Penser la Musique Aujourd'hui.* , Paris 1663. Trans. Susan Bradshaw and Richard Rodney Bennett.

Chowning, J.M. 1971. "The simulation of moving sound sources". In *Journal of the Audio Engineering Society* 19,1. pp2-6.

Dannenberg, R.B. (1993). "Music Representation Issues, Techniques, and Systems". *Computer Music Journal* 17,3 pp20-30.

Gibson, J. J. (1966). *The senses considered as perceptual systems*. Boston: Houghton Mifflin.

Gibson, J. J, (1979). *The ecological approach to visual perception*. Boston: Houghton Mifflin.

Goldstein, B.E. (1984). *Sensation and Perception* (2nd Ed.) Belmont, California: Wadsworth Publishing Company.

Grey, J.M. (1975). *An Exploration of Musical Timbre.* Center for Computer Research in Music and Acoustics, Department of Music Report No.STAN-M-2, Stanford University, February.

Handel, S. (1989). *Listening: an introduction to the perception of auditory events*. Cambridge, Massachusetts: MIT Press.

Lombardo, T.J. (1987). The Reciprocity of Perceiver and Environment. The Evolution of James J. Gibson's Ecological Psychology. Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Stockhausen, K. (1977). Interview with Paul Bronowsky in Access All Areas, broadcast 4 May, ABC Television, Australia.

Theile, G. (1986). "On the standardisation of the frequency response of high-quality studio headphones." *Journal of the Audio Engineering Society*, 34. pp 953-969.

Venonen, K. (1994). "A Practical System for Three-Dimensional Sound Projection." In *Proceedings of the Symposium on Computer Animation and Computer Music.* Canberra, Australia: Australian Centre for the Arts and Technology.

Vennonen, K. (1995). *An Ambisonic Channel Card*. Grad. Dip. Thesis. Canberra Australia: Australian Centre for the Arts and Technology.

Wessel, D.L. (1979). "Timbre space as a musical control structure". *Computer Music Journal, 3,* No. 2, 45-52.

Wishart, T. (1994) *Audible Design. A Plain and Easy Introduction to Practical Sound Composition*. York, UK: Orpheus the Pantomime Ltd.

Wishart, T. (1996) *On Sonic Art.* Ed. Simon Emmerson. UK: Harwood Academic Publishers.

Worrall, D. (1989). "A Music and Image Composition System for a Portable Multichannel Performance Space: A Technical Overview". *Chroma* 1/3. Journal of the Australian Computer Music Association. December. pp3-6.