FROM SURROUND TO TRUE 3-D

PETER P.LENNOX[1] , DR.TONY MYATT[2] AND JOHN M.VAUGHAN MSC[3]

[1]Research Student, Department of Music, University of York, York YO10 5DD, United Kingdom
ppl100@york.ac.uk
[2]Director of Music Technology Research, Department of Music, University of York, York YO10 5DD, United Kingdom
tone@cage.york.ac.uk
[3]Research Associate, Department of Music, University of York, York YO10 5DD, United Kingdom
jmv100@york.ac.uk

To progress from surround sound to true 3-D requires an updating of the psychoacoustical theories which underlie current technologies. This paper shows how J.J.Gibson's ecological approach to perception can be applied to audio perception and used to derive 3-D audio technologies based on intelligent pattern recognition and active hypothesis-testing. These technologies are suggested as methods which can be used to generate audio environments that are believable and can be explored.

INTRODUCTION

The work described in the paper is aimed at users and designers of medium-to-large-scale multi-perceiver audio diffusion systems incorporating spatially distributed sound elements over a broad field. Experience has caused the authors to reject single perceiver-based systems for this purpose and highlighted the need to develop a new paradigm for concert and other large-scale use of surround and 3-D audio.

Capable surround sound systems, optimised for single perceivers, exist which give sufficiently accurate representations of soundfields to support visual presentations and various audio illusions. It is also possible, through a variety of processing techniques, to manipulate the surround sound signals in order to simulate movement through a soundfield. This principle underlies the processing used by video game technologies and VR systems, sometimes also using head-tracking data to inform processing. However, most existing techniques give a single perceiver audition facility and audio illusions generally break apart when the listener moves. Transferring these principles to large-area multi-perceiver presentations can often give disappointing results. Techniques have been developed to sharpen imaging and improve image depth for large-area multi-loudspeaker arrays, which are reported elsewhere. However, many surround sound systems can be disappointing, because they fail to support a sufficient depth of audio illusion over a wide range of conditions.

The room acoustic of many concert listening situations interferes substantially with the design intentions of phantom image-based systems. This represents a constraint on how far these systems can be scaled in size; they are designed for small scale use, what we will refer to as 'single-perceiver' systems.

Our most successful experiments have been based on fully optimised periphonic ambisonic large-scale diffusion systems. This may be, in part, because this format is inherently capable of being scaled to very large numbers of loudspeakers with consequent improvement of results[1]. In particular such 'scaled-up' systems exhibit a) better rendition of plane-wave characteristics, and b) less excitation of listening-room acoustics due to lowered SPLs at each loudspeaker.

Whilst genuinely impressive '3-D' illusions have been achieved using ambisonic systems[2][3], the results compare poorly with our understanding of spatial reality. Sound images created by methods such as Ambisonics, though perceptually separable, remain difficult to localise. Natural perceptual attempts to improve localisation by head turning and perceiver relocation around or toward a postulated object, always result in a decrease in spatial information apprehended; the opposite of a natural environmental situation.

Images tend to pull to the nearest loudspeaker, causing the auditory illusion to disintegrate. It is these failings that lead us to distinguish (within the large-scale listening context) between 'surround sound' and 'true 3-D'.

Phantom images may be moved by changing interaural differences. However, artificial manipulation of distance can only be achieved by amplitude manipulations, use of artificial reverberation and frequency-dependent amplitude manipulation (filtration, excitation etc.). In general, the available manipulations are unwieldy and yield perceptually unsatisfactory results when compared with natural ambisonic recordings and especially when compared with real soundfields. This is largely because of the conceptual and practical complexity of maintaining coherence with a large number of parameters.

One strand of the research at York has more recently focused upon the derivation of assessment criteria for the depth-of-illusion of sound (re)production systems. Much of this has been informed by the concurrent formulation of a spatial syntax for electroacoustic composition. Preliminary results have produced criteria that characterise the distinction between surround sound and true 3-D sound and which should, we think, inform technological development of large-area sound diffusion systems.

This work has been motivated by the desire of the authors to enable a music which actively engages our spatial perception. It is felt that this will help to create a closer, more involving relationship with the music thus presented. Music need not be distant, 'over there'. No longer a poor copy, a music *re*production may be as vital and involving as the original. It may be 'here' which, for most current reproduction systems, it is not - implying that there is something spatially lacking in our present electroacoustically (re)produced music.

1. WHY DO WE WANT 3-D AUDIO?

Compositionally, there are a number of converging interests. Composers have expressed the desire to capture, or abstract, interesting sound images from real environments to use within the immersive, 3-D musical environment they wish to create. Electroacoustic composers at York have long explored the possibilities afforded by the *de novo* synthesis of soundfields and the manipulation of 'objects' within those fields, but have found conceptual formulation and practical realisation of compositional intent difficult.

From a composer's point of view, 3-D music seems to offer the prospect of using 'space' as a musical parameter, with objects changing their aspect or location with respect to the listener, or even by the depiction of entire 'unlikely' or 'unfamiliar' environments. In practice, however, target images tend to be delineated primarily by their angular position as subtended at an ideal perceiver-position and aspect.

Overall, whilst the authors have encountered a high degree of unanimity with regard to the lack of realism available in large-scale surround systems, we have found little consensus about the concept of 'realism' itself. Of course, in this context, the term 'realism' need not be interchangeable with 'real'. That is, we need not aspire to the creation of a soundfield which will inevitably be confused with the real world. Nevertheless, we assume certain necessary attributes to our notion of 'realism'. Believability is one. For concert situations, as we have already suggested, surround sound systems fail to realise the potential for exploration, either through movement or selective attention. We might apply this as a 'minimum requirement' for true 3-D systems.

Furthermore, recent progress towards the understanding of perceptual systems leads us to believe that increased congruence of sensory information will lead to a perceptually heightened and psychologically more engaging result. Previous and current systems lead to a state of impoverished listening, for example limited by restrictions on movement within the field, where it is not possible to know more about the environment through exploration.

The development of true 3-D therefore affords the possibility of wholly controllable audio environments, in which the engineer and composer can create an illusory world. By developing a greater understanding of what the apogee of 3-D/surround implementations might be, more powerful tools become available to composers and audio engineers that will help to fuel the evolution of contemporary electroacoustic, electronic, classical and popular music. Success in this endeavour depends on understanding spatial perception.

2. SPATIAL PERCEPTION - NEW PERSPECTIVES

Philosophical discussions of the nature of spatial perception are endless. Attempts to define the terms have been recorded over a span of at least 2000 years, making it one of the longest-running debates in history.

The term itself is composed of two elements, which themselves are abstract nouns. The dangers of reifying abstract nouns are well-known[4], yet the idea of an underlying reality consisting of 'space' which we may objectively measure is so attractive that we may be forgiven for assuming that the term 'space' is synonymous with that which can be measured in units, i.e. Euclidean space. According to this method, perception mirrors reality, and is thus presumed to be similarly measurable in inviolable units.

This line of reasoning leads to the notion of five discrete senses, where the 'purest' form of perception (i.e. the most accurate reproduction of the world 'out there') occurs at the sensory organs. It becomes the brain's task to attempt to evaluate these 'pure' perceptions and make deductions about the nature of reality which are used to trigger behavioural responses. However, because of the inherently subjective nature of the brain (an egocentric organ), some of the objective nature of reality is lost. It is presumed that much philosophical reasoning is required to 'restore' the lost reality, and humans with their considerable cognitive capacity, are uniquely positioned to indulge in the deductive reasoning that this approach to perception would require; thus uncovering more of reality.

This approach fails to explain how phylogenetically lower orders of animals, with limited powers of deductive reasoning and therefore a presumed inability to deduce the metric units of Euclidean space, manage to 'understand' pure perceptions at all (never mind invoking appropriate behaviour and avoiding collisions). It is clear that this 'bottom-up' signal-processing model of perception is inadequate - it is not the approach currently used to study human perception.

2.1 The Ecological Approach to Spatial Perception

We can no longer think of sensory input processing in terms of a unidirectional stream with a hierarchical organisation of processing functions. Evidence is emerging of innumerable connections between many, seemingly unrelated, areas of the cortex, at all levels. It therefore appears that cortical networks act in parallel at different levels, in a way that allows much greater sophistication of higher-order processing[5].

Modern thinking about the nature of spatial perception owes much to the paradigm shift effected by J.J.Gibson in his *Ecological Approach to Visual Perception*[6]. Gibson proposed a move away from perception as signal processing towards perception as information processing. The significant change is that, for organisms such as ourselves, perception occurs as a result of actively seeking out appropriate information. Behaviour is not a *result* of cognitive consideration of the products of perception, but a primary method of *achieving* perception.

This requires a concept of perception as a form of hypothesis-testing[7], with predicted/actual comparison and concomitant memory storage and retrieval mechanisms. Thus, moment by moment, the utility of input from organs of sensation is limited to providing evidence against which perceptual predictions may be tested. In other words, due to physical limitations on signal speeds, both outside and inside the perceiving organism, in order to 'know' the present in sufficient detail to ensure optimal survival possibilities, it must have been *predicted*. Perception attempts to 'know' the present, whilst sensation occurs in the very recent past.

Of course, an overly simple model of perception as hypothesis-testing would fail to represent the ways in which sensory novelty might be processed. Sensation, by definition occurring in the recent past, would fail to keep up with fast-changing stimuli. This is explicable in terms of what Bregman calls 'regularities'[8], or 'perceptual assumptions' in Gregory's view[7]. Essentially, a fast-track form of processing relies on certain pre-existent, rather than calculated, assumptions which can be rapidly deployed in hypothesis-testing. These assumptions are shaped by the physical characteristics of environments we are likely to encounter, such as the presence of gravity, and the behavioural characteristics of objects we are likely to encounter.

In addition, attentional models of perception predict the perceptual systems' preference for stimulus novelty as a criterion for hypothesis-generation. We can say that, automatically, more cognitive processing is devoted to novelty and a correspondingly lower order of resolution suffices for non-novelty[9].

2.2 Signals and Information

Gibson's proposal of 'direct perception' of 'environmental affordances'[10] without the need for intervening cognitive processes fails to address the evidence for cognitive processing or why we may have evolved such apparently unnecessary and computationally expensive abilities. Nevertheless, his contention that information is *external* to the perceiver suggests that such information is, in some part, independent of the sensory modes used to apprehend it.

Millar provides a more comprehensive explanation of the interrelationships of sensory modes in her CAPIN (Convergent Active Processing in Interrelated Networks) model[11], whereby the overlap of information gleaned by different sensory modes allows cross-referencing, the complementary sensory mode-specific portion providing useful extra information. Taking these notions together, we can propose that the reason for our having more than one sense modality acting in concert (as opposed to, say, a single extremely highly-developed sense modality), is that our comprehension of the world 'out there' is correspondingly more robust. That is, when sensory evidence is extremely highly correlated, we have no need for time-consuming cognition as to the 'meaning' of signals, which has obvious survival-related advantages. Similarly, while these correlations accord with the expectations generated by our hypothesis-testing perceptual mechanisms, no burdensome computation is required.

In effect, the term 'spatial perception' refers to our apprehension of information about relationships between features of our environment at a level of detail specific to the task(s) in hand, characterised by Millar as 'task-specific' definitions[12]. Selection of information depends on the frame(s) of reference we employ[13]. It is made in accordance with 1) internal need states, and 2) external information. Thus the 'space' we may choose to perceive may be very large, but we will perceive it at a low level of resolution, or it may be small, at a correspondingly higher level of resolution.

Overall, the different (but overlapping) models of spatial perception used today have a common attribute: Their utility relates to the apprehension by *organisms* of useful information about *environmental* features. In psychology, neurology and philosophy, there are no empty abstract spaces, nor proposals as to how or why one would know them. There is no proposition of an underlying pure form of reality made of inviolable concrete units. We cannot, therefore, restrict our definition of space to the Euclidean metric. Whilst the environments we inhabit (and perceive) may well be measurable according to this frame of reference, it is by no means the only available frame of reference.

The models of perception we have outlined are based upon the active apprehension of information about our environment and allow us to state that spatial perception, that is the cognitive individuation of features of our environment, is a pattern-recognition process using a wide variety of informational and signal qualities. In the audio context, this may include (but is not restricted to) the various known interaural differences available to a perceiver.

3. SURROUND PSYCHOACOUSTICS

3.1 The 'Direction is Space' Fallacy

The origins of multi-loudspeaker surround sound systems lie in the work of Blumlein. His Patent No.394325[14] elucidates the conceptual basis as well as many of the limitations of his approach. The patent describes techniques by which audio illusions ('phantom images') can be generated by amplitude differences between two evenly-spaced loudspeakers giving rise to interaural intensity differences, and phase differences at low frequencies.

There are many references in the literature to the notion that our auditory perception of space is a result of our having two ears. A moment's reflection tells us that this is obviously not so. People with hearing in only one ear do enjoy auditory spatial perception. The reasons for the continued misrepresentation of our perceptual abilities may simply be associated with the historical paucity of research into auditory spatial perception. Indeed, as Millar points out, space perception in much of the literature is synonymous with vision[15]. Studies on audition are largely concerned with how we process *sound*. It is axiomatic to our approach that in everyday environments we listen to *objects* in *places*.

Consequently, Rayleigh's duplex theory of interaural differences has often been treated as a perceptual theory, when it is in fact a theory of sensory processes. Hence, it is often believed that our perceptual individuation of objects is achieved solely by sorting of sounds according to their directional characteristics. Partly as a result of this, surround technologies have tended to concentrate on presenting signals to the ears that accord with this crude stimulus-response model of sensory processing. The transition from the underlying theory of surround sound to that for true 3-D requires the amalgamation of existing sensory theory with findings from neuropsychology, behavioural studies, virtual reality research and artificial intelligence research to form a cohesive working perceptual theory for use in audio engineering. Modern approaches to understanding perception focus on hypothesis-testing and pattern-recognition. This implies a much fuller and more sophisticated coherent information content of signals presented to listeners in order to achieve the perceptive effects to which the systems purport to aspire. However, it will be shown that specifying the spatial parameters of audio as closely and accurately as the real world, which would lead to impossibly high signal bandwidths, is unnecessary. Furthermore, perhaps the most significant insight that modern psychoacoustics has to offer the development of 3-D audio is the realisation that audio spatial perception is primarily a time-domain process.

It appears that spatial information is derived from audio signals by the ear-brain in accordance with the temporal relationships between self-similar audio components, extracted by a process of auto- and cross-correlation of auditory signals. While a discussion of the mechanisms of neurophysiological representations of audio spatial parameters is clearly beyond the scope of this paper, it highlights a class of aural information which has been largely neglected by current technologies. Indeed, temporal differences at the ears are only system-induced using conventional techniques at low frequencies, which is informed by an outmoded view that the temporal differences associated with high frequency signals could not be represented neurologically. The more complete view tells us quite the reverse: it is in fact the temporal relationships between self-similar audio components at *high frequencies* which primarily give rise to good individuation and localisation information.

3.2 Can Surround Evolve into True 3-D?

In many fundamental ways, existing surround technologies are trapped within a conceptual and contextual framework that has limited development. For instance, the development of surround technologies has to a large extent been guided by requirements of the film and TV entertainment worlds, and here supporting a narrow visual field with a 360° soundfield has necessitated techniques which result in a soundtrack whose depth of illusion is deliberately poor. Where the focus has been realism, technologies have often relied on psychoacoustic principles which are weak and simplistic, or on abstracted mathematical concepts which falsely assume that the world is perceived in geometric terms. Although the contextual limitations on the development of surround technologies are objective-orientated, and thus we would only expect to see development to the degree necessary for specific applications, the conceptual limitations are profound and unnecessary.

The preceding discussion of spatial perception should now highlight the limitations of phantom imaging systems to maintain audio illusions, and explains why these break down so readily by deviating even slightly from the optimum listening position. Most current surround sound systems employ substantially the same concepts, although there have been extensions in order to firm up image location at one part of the soundfield, namely the centre-front. The limitations to the conceptual basis of current systems, such as the constraint of a single specifically-located perceiver, are derived from the crudeness of the psychoacoustical basis. Even sophisticated surround systems such as Ambisonics have only a slightly broader psychoacoustic base, in that a wider range of psychoacoustic cues are attended to with a degree of frequency-dependent optimisation, and sweet-spot restrictions over a reasonable range of frequencies are limited.

In rejecting the requirement that a perceiver should, ideally, be placed equidistant from the speakers, we are conveniently rejecting the causes of one of the inherent problems with phantom imaging systems, namely the tendency for image components to perceptually ‘pull’ to the nearest speaker. In ambisonic systems, this phenomenon sometimes manifests itself as complete 180° reversals across the horizontal plane. It is possible to substantially enhance the quality of ambisonic soundfields, and research at York has resulted in trials in which the usable area is extended. However, each perceiver receives broadly the same perspective, with image distortions increasing with distance from the centre. While further development of this system could perhaps lead to a first stage ‘minimum level’, the surround sound element being only part of an eventual 3-D system, the authors, nevertheless, feel that the conceptual basis for this type of system precludes its qualification as a large scale ‘true 3-D’ system.

4. EXPLORING PERCEPTUAL SPACE

4.1 An Object-Centred Approach

The systems we have hitherto encountered could be described as being based upon a ‘perceiver-centred’ philosophy. We would advocate a complementary approach, based on an ‘object-centred’ philosophy. In this approach, we are concerned with the information-yielding properties of real objects in real environments, what Gibson calls ‘affordances’[10].

In this approach, the object, the environment and the perceiver are assumed to be integral. The separation of the ‘what’ from the ‘where’, either as a result of conceptual abstraction as in physics, or ‘representational abstraction’ as in sound recording and reproduction, inevitably leads to some informational damage. Restoration or, more accurately, reconstruction, of the perceptually important information, is the aim of this approach.

The first artificial abstraction to consider is the conceptual separation of 'audio-space' from the spatial representations engendered by holistic deployment of the variety of sensory modes. We must accept that a proposed 'audio-only' spatial representation is inevitably inferior to the type of space rendered by the perceptual mechanisms which normally act in concert. That is, our hearing did not evolve in isolation. Modern studies of perception emphasise the interconnections between the sensory modes at even quite peripheral levels. Quite simply, in 'normal' environmental conditions, our audio perception incorporates information from the visual world, the haptic world, the olfactory world, and so on. The distinctions between these 'worlds' are quite artificial, and should be treated with caution.

Although we accept the artificial nature of such a conceptual division, it is nevertheless instructive to consider the differences between the auditory and 'other' worlds in terms of the information-yielding properties of environmental features. Take the visual world, for example. The overwhelming factor is that our visual perception occurs as a result of the existence of ambient light. Very few organisms, objects or surfaces are self-lit. They reflect, refract and occlude sunlight. This has profound implications with respect to the way in which we perceive those elements which are most important to us. We rely not on the energy *output* of aspects of the environment to distinguish features of the environment, but on very fine discriminations between the ways in which ambient energies are affected. Thus, in vision, our environment is 'sorted' according to colour, optical texture, shape, size and movement. The key to individuating objects is our ability to discern 'edges' - the discontinuity between the optical characteristics of adjacent surfaces. This ability allows us to recognise 'shape'.

By contrast, in the auditory domain, we have no need, or mechanism, to delineate 'edges' in order to detect organisms. We make quite different discernments about objects in the auditory world, whose perceptual edges are temporal rather than spatial. Auditory sensation depends on the existence of energetic processes in a way that vision does not - one may see an environment in which nothing is moving, one would not hear it.

4.2 What and Where

It is useful to separate spatial perceptions into 'what' and 'where'. There is considerable neurobiological evidence that behaviourally-flexible higher primates employ neurological representation of multiple 'what' and 'where' processing streams in the visual system[16]. How they are functionally integrated is under investigation, and the authors are not aware of similarly detailed studies of audio spatial perception, but assume that equivalent questions of integration must be addressed.

Nevertheless, we can say that the 'what' representation yielded in the audio mode is different from, complementary to, and as richly detailed as that derived from the visual mode. Whilst form is not as well rendered, material composition, structure, type-of 'what' and size-of 'what' often are. On first consideration, our 'where' representation seems less detailed in the auditory mode, but this is to entertain the outdated misconception of sensation being equivalent to perception and thinking of senses as competitive/hierarchical votive systems. In the complementary and overlapping model, audition is as likely to occasion foveation as is vision. Furthermore, our audition is able to render 'where' information unavailable to vision. For visually occluded objects, either behind the perceiver or in another, presumably adjacent, space, information is readily available to audio perception as to rate of movement, change of movement and even reason for movement. This type of information may often be significant in determining call to action. It is available completely independently of changes in interaural differences. We have termed this type of 'where' information *ambient labelling* information.

Although we will leave a detailed description of *ambient labelling* and its implications with respect to defining 3-D audio environments to another paper, this type of object-centred information is vital both to an understanding of what information we might wish to present in 3-D audio, and to ways in which we might wish to build up a notional physical space or environment. In audio spatial perception, the corporeal presence has as much to do with temporal relationships with local surfaces as traditional concepts of spatial relationships, that is, an object is known in part by its effect on the local environment through reflections, reverberations and so on.

4.3 Perceptual Significance

We have concluded that, for our large-scale listening context, the desire to produce 3-D music is synonymous with that to produce an information environment. Such informational environments are conceptually understandable in terms of ‘what’ and ‘where’. By contrast, surround sound, because of its heritage from ‘stereo’, is basically a ‘what’ system, the ‘where’ being limited to the angular separation of ‘whats’, with limited depth of field and more importantly no real depth of significance. Within the listening environment, the chief sources of *ambient labelling* information are the static loudspeakers. This restricts the possible range of *perceptual significance*. The problem is exacerbated by the use of too few loudspeakers.

Initial difficulties with redefining the psychoacoustical basis are concerned with ‘units of measurement’. If we are to forego the elegant simplicity of a geometric model with its absolutely defined metric units, what units should replace them? The problem is that the units of perceptual space are not absolute, objectively measurable units, independent of a perceiver’s viewpoint. They are, instead, ‘relative units’, their values varying according to the perceiver’s assessment of the importance of various features of an environment.

In our perception of environments, the relationship between signal properties, sensory data and perception is necessarily loosely-coupled[17]. Gross changes in signal properties are not necessarily more perceptible than minute changes. They do not yield more information. The authors employ the term *perceptual significance* to emphasis the ways in which cognitive functions select for attention those information-yielding properties which have the potential for facilitating the most useful predictions.

The important conclusion is that not all spatial relationships of similar metric measurements are of equal significance to a perceiver. A distance of, say, ten metres between the perceiver and a spatial position which contains some reward is quite different to a, metrically similar, vertical one, especially a downward one. So the units of measurement of such an environment are entirely mediated by the significance to the perceiver of that which is measured.

An important aspect of *perceptual significance* is that many of an organism’s judgements of distances are made with reference to the time taken to close them. We are familiar with the situation where, on asking ‘how far’ something is away from us, common answers are given in units of time, ‘5-minutes’ walk’, or ‘one hour’s drive’, for instance. Judgements of this type are predictive of the energy requirement of a proposed course of action. For more proximal situations, energetic judgements may be made more directly. One may think of something as so-many paces away. Temporal judgements may be split-second and made relative to predictions about the behavioural potential of a ‘what’. In this context, the maximum *speed* capabilities (of a ‘what’) are far less relevant to the situation than the *accelerative* capabilities. Auditory perception contributes the ability to holistic perception of continuously monitoring for novelty across the whole of the local environment. It is well-suited to detecting accelerative changes in the locomotory behaviour of proximate objects. As we have stated elsewhere, these detections may be explicable independently of interaural differences, or even pinna effects.

Information as to the spatiotemporal behaviour of ‘what’ features of our environment may be directly available as recognisable patterns through time. The significance of this is that, unlike vision, for an early-stage assessment, no cross-referencing with any other features of the environment (save the perceiving organism) need be postulated. Perceptions of this sort are fast, simple egocentric reference-frame based. They comprise high perceptual significance components and thus the behaviour may be wasteful of energy.

It is therefore valuable for a survival-related perception to have an audio space comprising several ‘nested’ frames of reference, the subdivisions determined by the type of behaviour required in accordance with *perceptual significance*. Energy need only be expended in accordance with necessity. Our audio perceptual space may thus be divisible as follows:

- here and urgent;
- here but not urgent;
- near and coming;
- near and could come;
- near and going;
- adjacent space and coming;
- adjacent space and moving (could come);
- distant and coming;
- distant and mildly interesting.

There are probably other subdivisions of interest, for instance, ‘here and empty’, in which the perceiving organism is the only sound-producing ‘what’.

It is also probably possible to further abstract the domains of perceptual space into simpler classes which are of more use in determining the information content of components of synthesised soundfields, and while conventional musical parameters such as timbre are clearly going to have an important, but relatively easily predictable, impact on *perceptual significance*, the authors seek to highlight the way in which the development of human spatial perception has afforded particular prominence to the proximity and behaviour of objects. In particular, the process of abstraction itself is of crucial importance because it demonstrates how an understanding of *perceptual significance* can be used to implement the massive informational content that a deeper apprehension of perceptually interesting audio environments requires.

4.4 Cartoonification

The assumption that 3-D audio is built on providing the senses with signals whose spatial relationships are as closely and metrically defined as their physical relationships places onerous burdens on technological development. As indicated, one of the most important advances that must be made in order to develop from surround is the appreciation that the world is not comprehended in Euclidean terms. Classifying the spatial characteristics of sounds in terms of the way in which our perceptual systems grade stimuli for action not only allows us to give very usable psychoacoustic cues, but also allows for a drastic reduction in the signal bandwidth. For instance, an audio source which is distant enough and of such character as to afford little attention does not need to, and should not, be accurately defined in terms of localisation, proximity or indeed signal quality. This process of abstraction we call *cartoonification*. It is a way of increasing information bandwidth without necessarily increasing the signal bandwidth significantly.

Cartoonification is, however, not in itself a signal processing technique, but a philosophical tool which may be employed to develop a variety of signal processing regimes. For example, a significant information-yielding property of many audible objects is that of asymmetry of audio dispersion, what we call *facingness*. Although *facingness* is not as sharply depicted in the audio as the visual domain, nevertheless for many medium-sized objects there is substantial asymmetry of output, especially of the higher frequencies, due to occlusion by the sound object's body. In many environments, the shortest signal path to a perceiver of these components is via one or more reflecting surfaces. In an artificial soundfield of sufficient competence, the spatiotemporal characteristics of these reflections may be simulated to provide *ambient labelling* information about the *facingness*, and change-of-*facingness*, of an object in our simulated environment.

Another useful example is in simulating and emphasising aspects of the behaviour of objects such as movement through an environment. Often changes in subtended angle are used (sometimes alone) to denote movement, but, especially for distant objects, this is unlikely to be the main perceptual cue to this aspect of an object's behaviour, and, as we have demonstrated, subtended-angle approaches would in any case need to attend to the auditory effects of the change of an object's location with respect to its surroundings. However, using a variety of time-varying rolling comb-filters applied to the upper portion of the spectrum of an object's signal, it is possible to simulate audio patterns which are perceptually interpreted as movement. When this is used in conjunction with treatments for facingness (as this usually correlates with the direction of travel and therefore provides useful information), the authors have found that accurate simulation of object trajectory by closely defining interaural difference information becomes unnecessary.

5. CONCLUSIONS

If Gerzon's theories[18] regarding the basis of Ambisonics and the paramount importance of the interaural differences so generated were correct, then artificial manipulations (of interaural differences) ought to have resulted in specific and realistic perceptions. Unfortunately, they did not. There is widespread agreement amongst composers that soundfields produced exclusively by electroacoustic means lack a 'sense of space'. Clearly, auditory spatial perception is of a somewhat different character to the classical concept of geometric space. The judgements we usually wish to make are not abstract ones. In environmental audition, distances between objects are less important than distances of objects from ourselves. The latter are measured not in metric units, but in terms of the time and energy it may take to get here (or there). We have proposed true 3-D as a 'space' which is not the classical physical space but an informational environment which we term *perceptual space*. In our auditory *perceptual space* we have a unique class of information about the 'what' and 'where' which we call *ambient labelling* information. Application of the concept of *perceptual significance* allows a process of *cartoonification* which may 'accurately' and efficiently portray a true 3-D soundfield which can be explored and is believable.

We have also questioned the appropriateness of the principle of system scaling for large-scale listening situations. Our approach indicates that increasing the discreteness of 'what' and 'where' information processing, perhaps even maintaining separation through to the eventual diffusion, may represent a more fruitful approach. Full audio bandwidth may be reserved for 'what' information, whilst 'where' information may be stored and transmitted as control data for utilisation by local signal processors in accordance with local diffusion and listening conditions. The authors' experiments with hybrid systems comprising a variety of techniques, including monaural, stereo, multiple-discrete and interlocking and/or adjacent ambisonic fields, lead us to believe that such an approach does not entail loss of fidelity. We look forward to composers designing pieces of music whose listening area extends beyond a single space and whose enjoyment can be enhanced by physical and conceptual exploration - this, in our view, would epitomise the notion of true-3D music.

REFERENCES

[1] Bamford, J.S. 1995. *An Analysis of Ambisonic Sound Systems of First and Second Order.* MSc Thesis, University of Waterloo, Ontario, p.128.

[2] Malham, D.G. 1992. "Experience with Large Area 3-D Ambisonic Sound Systems". *Proceedings of the Institute of Acoustics,* 14, 5, pp.209-215.

[3] Vennonen, K. 1994. "A Practical System for Three-Dimensional Sound Projection" In: *Synaesthetica '94, Proceedings of the Symposium on Computer Animation and Computer Music.* Australian Centre for the Arts and Technology, Canberra, 1-3 July 1994.

[4] Millar, S. 1994. *Understanding and Representing Space: Theories and Evidence from Studies with Blind and Sighted Children.* Clarendon Press, Oxford, p.7.

[5] *Ibid.* p.55.

[6] Gibson, J.J. 1979. *The Ecological Approach to Visual Perception.* Houghton Mifflin, Boston.

[7] Gregory, R.L. 1973. *Visual Perception.* Oxford University Press, Oxford, p.6.

[8] Bregman, A.L 1990. *Auditory Scene Analysis. The Perceptual Organisation of Sound.* MIT Press, Cambridge, Mass., pp.10-36.

[9] Dixon, N.F. 1981. *Preconscious Processing.* John Wiley & Sons, Chichester, p.224.

[10] Gibson, J.J. 1979. *The Ecological Approach to Visual Perception.* Houghton Mifflin, Boston, pp.130-143.

[11] Millar, S. 1994. *Understanding and Representing Space: Theories and Evidence from Studies with Blind and Sighted Children.* Clarendon Press, Oxford, p.15, p.262.

[12] *Ibid.* p.11.

[13] Brewer, B. and Pears, J. 1993 In: Eilan, N., McCarthy, R. and Brewer, B. (eds.) *Spatial Representation. Problems in Philosophy and Psychology.* Blackwell, Oxford, pp.25-31.

[14] Blumlein, A.D. 1931. British Patent No.394325.

[15] Millar, S. 1994. *Understanding and Representing Space: Theories and Evidence from Studies with Blind and Sighted Children.* Clarendon Press, Oxford, p.1.

[16] McCarthy, R. 1993 In: Eilan, N., McCarthy, R. and Brewer, B. (eds.) *Spatial Representation. Problems in Philosophy and Psychology.* Blackwell, Oxford, pp.314-324.

[17] Dixon, N.F. 1981. *Preconscious Processing.* John Wiley & Sons, Chichester, p.2.

[18] Gerzon, M.A. 1975. "Pan Pot and Sound Field Controls" *NRDC Ambisonic Technology Report No.3.* National Research Development Corporation, London.

BIBLIOGRAPHY

Gibson, J.J. 1950 (repr.1974). *The Perception of the Visual World.* Greenwood Press, Westport, Connecticut.

Gregory, R.L. 1968. "The Evolution of Eyes and Brains - a Hen-and-Egg Problem" In: Freedman, S.J. (ed.) *The Neuropsychology of Spatially Oriented*

Behaviour. The Dorsey Press, Homewood, Illinois. pp.7-17.

Gregory, R.L. 1994. *Even Odder Perceptions*, Routledge, London.

McAdams, S. and Bigand, E. 1993. *Thinking in Sound: The Cognitive Psychology of Human Audition*. Oxford Science Publications, Oxford.

Moray, N. 1969. *Listening and Attention*. Penguin Books, Harmondsworth.

Nielsen, S.H. 1993. "Auditory Distance Perception in Different Rooms", *Journal of the Audio Engineering Society*, 41, 10, pp.755-770.