3D AUDIO AS AN INFORMATION-ENVIRONMENT: MANIPULATING PERCEPTUAL SIGNIFICANCE FOR DIFFERENTIATION AND PRE-SELECTION.

Peter P Lennox[1], John M Vaughan MSc[2] and Dr Tony Myatt[3]

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

ABSTRACT

Contemporary use of sound as artificial information display is rudimentary, with little 'depth of significance' to facilitate users' selective attention. We believe that this is due to conceptual neglect of 'context' or perceptual background information. This paper describes a systematic approach to developing 3D audio information environments that utilise known cognitive characteristics, in order to promote rapidity and ease of use. The key concepts are *perceptual space*, *perceptual significance*, *ambience labelling information* and *cartoonification*.

1. INTRODUCTION

This paper, part of a series, describes a model of human auditory spatial perception that is quite different from the theories that underlie the use and design of most current audio displays incorporating some form of surround-sound technology.

Techniques for audio display incorporating surround sound are varied. In general these are optimised for single perceivers, particularly in support of visual presentations. However, they often fail to support a sufficient depth of audio illusion over a wide range of conditions. In particular, the results compare poorly with our experience of spatial reality. The available 'perceptual' manipulations are unwieldy and yield unsatisfactory results when compared with real sound environments. 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 the University of York has focused upon the derivation of assessment criteria for the depth-of-illusion of sound (re)production systems. It has also lead to the formulation of concepts that may assist in the development of technologies for providing information via audio displays, such as where visual senses are otherwise engaged or impaired.

Conceptually, contemporary use of sound as artificial information display is rudimentary when compared to developments in vision displays. This 'video chauvinism'[1] may be deeply rooted in the structure of the human brain, in the way we 'think' about the world about us. Surround sound appears to offer greater capacity for simultaneous information-transmission, and, from a sound designer's point of view', offers the prospect of using 'space' as a rich informational 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 subtended angle. Crucially, perception is concerned with the behaviour of real objects in real environments. The representation of sounding objects in sound fields as reconstructed plane waves from point sources in abstract empty space is always going to place a limitation on believability. 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.

One of the most significant problems is the failure of surround sound systems to realise the potential for exploration, for example information sorting through selective attention. Certainly, we would characterise these systems as being capable of rendering very limited 'depth of significance'. Typically, the perceptual foreground 'fills up' rather rapidly which leaves little scope for background or subtle information, crippling attentional processes. For this reason we have argued[2] that surround sound is not synonymous with 3-D audio.

A common feature of technical representations of audio spatial attributes is the concentration on accurately depicting 'objective' attributes of a specified sound *field*, as in the reconstruction of plane-wave characteristics of sounds for a specified listening position. However, the authors suggest qualitative differences between notions of sound fields and sound environments. The former are simplified, quantifiable representations of the latter, and are necessarily 'task-specific'. Thus, examples of perception in sound fields are special cases. It follows that there may be classes of information (or information-yielding properties) available in sound *environments* which may not be incorporated in particular sound *fields*. It further follows that theories of spatial perception cannot adequately be founded on theories of sound *field* perception.

2. DEVELOPING MODELS OF SPATIAL PERCEPTION

It was, until recently, believed that our perceptual individuation of objects is achieved solely by sorting of sounds according to their directional characteristics, based simply on the notion that our auditory perception of space derives from interaural difference information. Partly as a result of this, surround technologies have tended to concentrate on presenting signals to the ears that accord with this basic stimulus-response model of sensory processing. Such systems could be described as being based upon a 'perceiver-centred' philosophy. We would advocate a complementary approach, based on an 'object-centred', 'sound environment'-depiction philosophy. In this approach, we are concerned initially with the information-yielding properties of real objects in real environments.

We can no longer think of perception in terms of a unidirectional sense-data stream within a hierarchical organisation of processing functions. From the neurosciences there is strong evidence of innumerable parallel connections between many areas of the cortex at all levels[3], and *from* higher levels *to* more peripheral ones[4]. This provides for multiple ‘what’ and ‘where’ processing streams, at least in the visual system[5], and it is reasonable to suppose that analogous processes are available in the auditory domain.

While the differentiation is not particularly strong and there is considerable overlap, we have characterised 'what' processing as determining 'perceptual significance', while information about 'wheres' is provided to our perceptual systems by 'ambience labelling information', or information yielded by the interaction of sounding objects with their environment. This is of course an oversimplification, because we also contend that perceptual significance is concerned with the apprehension of behaviour in a qualitatively different and much more direct way and which occurs in priority to the construction of 'what' and 'where' perceptions.

This strand of thinking about the nature of spatial perception owes much to J.J.Gibson's Ecological Approach to Visual Perception[6]. Gibson proposed a move away from perception as signal processing towards perception as information processing, and further, that 'direct perception' of 'environmental affordances' was available without the need for intervening cognitive processes. That is, perception is a holistic apprehension of a qualitative feature of the world, namely ‘information’ which in turn is extrinsic to the percipient. Cognitive mapping onto a three-dimensional representation follows, imperfectly at best, so that we do not receive an inchoate sensory array, subject it to three-dimensional analysis, then further sort into the objects and background. It is quite the reverse: we apply selective attention to objects, and selective *in*attention to background, before we are able to complete a particular spatial perception. Therefore attributes of the available information, what we have termed audio 'texture' and which arise particularly from the behaviour of organisms, map directly to interpretations such as 'danger', and if it is possible to identify, extract and re-present the perceptually most important elements of this texture, treatments are available in audio displays that engender selective attention through a process that we have called 'cartoonification'.

It is clearly important in this respect to know what the 'parameters of thingness' are as it potentially tells us much about perceptual significance and cartoonification. We have elsewhere[7] identified and defined a number of important characteristics, related primarily to the central characteristics of behaviour and 'body'. For instance, one of the reasons why auditory patterns of behaviour are so perceptually significant is undoubtedly due to the fact that any sound-producing body is potentially an organism. However, we believe that the detection of organisms relies on much subtler parameters, and that patterns or textures can be discerned that afford detection or interpretation according to degrees of urgency, such as locomotory sounds or accelerative changes. However, the authors hope to make this the subject of a future paper.

An outcome of these psychological perspectives is the need to reassess whether the accurate representation of geometric spatial parameters should be the overriding goal of auditory display system design. We (elsewhere[2]) coined the term ‘perceptual space’ to acknowledge the substantive distinction between abstract physical space and that which we perceive. The primary contention here is that Euclidean theories of space are not comprehensive in their ability to describe the environments we inhabit, and that theories of perception based on geometric space will inevitably be of limited predictive utility. ‘Ambience labelling information’, an intrinsic property of environments, and ‘perceptual significance’, which is embodied in the structure of human percipients, are, we believe, the key features of perceptual space.

3. AMBIENCE LABELLING INFORMATION

Ambience labelling information concerns the relationship between sounding objects and their environment. Although it will not be discussed in detail in this paper, the term refers to a real class of information available in real environments, for which we postulate reciprocal dedicated unconscious cognitive processes. We observe that sounding objects are *generally* asymmetrical in their audio output, and that the immediately proximal features of the environment are also *generally* asymmetrical. Thus, as the sound output of an object radiates outward, colliding and reiterating, it acquires a rich ‘history’, in the form of textural inhomogeneity.

Whilst this type of information is *exterocentric*, from the point of view of perception, 'ambience labelling information' provides context: it is the sound of the environment in which objects find themselves. So, in a given sound environment, we do not just hear the sounding *objects* in that environment. Indeed, often the majority of the sound in an environment is reflected sound, but a rich pattern of reflected sound, which, being inhomogeneous, serves to anchor the objects 'out there', irrespective of specific percipient positions. Without this background, the 'foreground' objects of perception do not actually make sense: we might regard this background as fundamentally necessary for spatial perception and certainly, as we emphasise later, for the individuation of foreground features.

The significance of this can be exemplified by the fact that, from the point of view of human auditory spatial perception, the sound of an object moving is determined to a much greater extent by the change over time of the object's audio relationship with its environment and other sound objects, than by the change over time of subtended angle to the listener, and indeed the character of movement is apprehensible quite independently of changes in interaural differences of any sort. That is to say, we can perceive movement directly, independently of our understanding of local topography. Surround sound audio displays that rely simply on mapping geometric relationships between sound objects and the listener, however accurately, will ultimately fail to produce believable audio environments, as the crucial information channels required for the cognitive understanding of the environment are missing. It is this factor that, we believe, characterises the distinction between sound fields and sound environments.

4. PERCEPTUAL SIGNIFICANCE

The authors employ the term 'perceptual significance' to emphasise the ways in which cognitive functions select for attention those information-yielding properties that have the potential for facilitating the most useful predictions. This highlights the finding that spatial perception is shaped by the adaptive need to rapidly codify features of the environment in accordance with relevance hierarchies through the apprehension and interpretation of, primarily, behaviour. In contrast to the exterocentric nature of ambience labelling information, we would characterise perceptual significance as perceiver-centred processing of ambience labelling information.

The central principle of perceptual significance is that information-features in real environments fall naturally into ‘relevance hierarchies’ that pertain to individual survival. It is well known[4] that perceptual systems employ a process that we have characterised as 'cartoonification' in order to achieve fast efficient representation of the world. We postulate multiple such ‘audio feature detection systems’, each specialising in a narrow range of features, and each contributing to multiple, nested or parallel hierarchies of urgency. An investigation into how such hierarchies are neurologically enacted and integrated to form the apparently holistic perception of the world about us is of course beyond the scope of this paper, however, there is an established and growing body of evidence for such hierarchies (e.g.[5]).

In a previous paper[2] the authors proposed a central relevance hierarchy composed of perhaps eight discrete tiers. Within that, we recognise the overwhelming propensity for the apprehension of behaviour, and characteristics of organisms in particular, whose potential for threat/reward permits further sorting. For the purposes of ascertaining the most useful categories for the presentation of information, we can assume that the tiers are generally a sub-class of perceptual systems devoted to the apprehension of potential danger, but that the most perceptually significant events give rise to responses by virtue of dedicated unconscious processes. That is to say that *immediate* danger precedes What and Where identification and is available as a direct and immediate product of behaviour detection. The apprehension of *potential* danger is clearly derived from What and Where, and we can propose that behavioural, intentional and contextual analysis plays a part in this, but precedes any detailed view. Crucially, we seek to highlight the presence of a number of primitive processes that precede and initiate selective attention.

Certainly, the mystery of 'selective attention', that is, the apparent paradox whereby complex ‘intelligent’ attentional processes are occasioned by much simpler, more ‘stupid’ mechanisms, is resolved in understanding that primitive feature-detectors do not, of themselves, ‘judge’ *importance*, merely *presence*. If feature detection systems are arranged in appropriate hierarchies, the apprehension of urgent information is an automatic and emergent property of the system, without one having to postulate that a complex analysis must be performed *before* ‘urgency’ can be recognised.

As a caveat here, we should observe that there is no evidence to suggest that such hierarchies should exist entirely discretely for, say, the auditory domain and the visual domain, and so it would be unjustifiable to assume that all the pertinent information required for rounded *auditory* spatial perception should in itself *be* audible.

It is a basic tenet in our approach that such hierarchies as we describe do not merely facilitate a ‘bottom-up’ unidirectional information flow from the periphery to the higher functions, but that ‘top down’ and ‘sideways’ information flows are quite as important. In this way, primitive mechanisms that have, in themselves, only limited discriminatory capabilities are ‘guided’ by more central processes, thus demonstrating functional flexibility appropriate to diverse contextual requirements.

5. CONTEXT

There are extensive ramifications of the subject of ‘context’, all relevant to the present discussion, but too many to cover in this paper. The importance of a ‘perceptual context’ that corresponds to ‘environment context’ cannot be overstated, as it is, we believe, what facilitates ‘selective attention’.

Underpinning the objects, features, events and relationships of the world about us are layers of ubiquitous, stable 'contexts': temporal, spatial, historical and so on. These contexts are real, physical facts *in* the world. The relevance to the present discussion is that, whilst these contexts are important (vital even) to perception, they are not to what we pay selective attention. It follows then, that cognitive representation of this important class of information must be characterised in some other way: we use the term ‘selective *in*attention’.

From the perspective of considering perceptual significance, ‘context’ would seem to be of a very low order of urgency, and we would expect ‘context information’ to be almost beneath notice, a sort of ‘perceptual background’, *unless* it is somehow ‘wrong’. On the other hand, the most perceptually significant occurrence would be a drastic, wholesale and unexpected (i.e. out-of-context) *change* of context. Rather than an ability to pay increased selective attention to objects and features, for which there is no evidence of human superiority when compared to other species, it seems plausible that the major difference between humans and other percipients is in this, our ability to cognitively represent complex and multiple contexts. We characterise ambience labelling information as ‘context information’.

6. CARTOONIFICATION

Perceptual space can be considered in the light of a 'foreground/background axis', from maximum significance to completely ignored: the gradations between are presently innumerable. By proposing 'cartoonification' strata, we hope to reduce the problem to definable qualities. We have qualified a structure of relative significances that is potentially quantifiable. In so doing we are merely formalising what many users of artificial environments (e.g. composers and sound designers) intuitively grasp. This perceptual space can also be considered in the light of 'what' and 'where', together with 'what' in 'where': each category contains simplified graduations of relevance which we can program for.

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. However, 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 information-stimuli for action not only allows us to extract the most important psychoacoustic cues, but also allows for a drastic reduction in the signal bandwidth. For instance, an audio source that 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 while at the same time reducing the signal bandwidth.

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 sound fields, and while conventional audio 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, potential for threat (and/or reward) and behaviour of objects. Furthermore, 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.

In developing cartoonification techniques, we find that classes of treatments fall into three nominally discrete domains: The cartoonification of wheres, the cartoonification of whats, the cartoonification of whats into wheres, and the cartoonification of events.

6.1. 'Where' Cartoonification

Many existing surround sound systems (and even single-speaker displays) are potentially capable of successfully representing space in a cartoonified way. In a previous paper we considered the 'shape' of perceptual space and concluded that

> *"'space' can be loosely decomposed into 'my space', 'adjacent space(s)' and 'distant'. In sound terms, each of these can be further decomposed into 'things' (that instantiate sound) and 'place-features' (that do not 'make' sound). The things are the perceptual foreground items, to which we pay selective attention, whilst the place-features are heard, but not attended to. Considering the spaces purely in terms of place-features, for a moment, 'my space' should command greater perceptual processing because a) the things in it are potentially more urgent (according to perceptual significance … and b) there is more textural inhomogeneity available in the ambience labelling information."*[7]

'M*y space', 'adjacent space' and 'distant space'* are hierarchical perceptual abstractions that attach themselves to specific *places.* A place is that which contains physical features that reflect, refract, absorb and occlude sound such that its 'shape' is discernable (to a greater or lesser extent) and can be mapped onto a neurological representation of three-dimensionality *if required*.

'My space' is self-explanatory, and has a shape that extends around the percipient. Objects in my space are interesting because of their proximity, and the implications for threat or reward. Organisms in this space are disproportionately so, due to their inherent capacity for locomotion.
Adjacent space does not extend to and (generally) around the percipient. Sounds occurring in adjacent spaces are rarely confused with those in my space. Events and things in adjacent space are less interesting than those in my space, but nevertheless inherently possess the potential for impinging on my space
'Distant' space is inherently very non-threatening, generally, and is almost never confused with either of the other two.

Two crucial results arise from this in considering the presentation of information via auditory displays. First, except for binaural presentations and to a very limited extent in multispeaker systems, the available depth of field lies firmly beyond the speakers. This places a limitation on the simulation of proximity and thus on the degree to which urgency can be presented. In fact, the authors' experience suggests that the special class of perceptual significance that we have characterised as 'immediate danger' is probably completely unavailable in most auditory displays, except through quite gross out-of-contextness.

The second result is the importance of 'ground effect'. We have elsewhere[7] highlighted the importance of the first six or seven early reflections in establishing the character of the ambience label, that is, the perceptually significant aspects of the interaction of a sounding object with its environment. In most simulated spaces ground reflections will form a useful part of the ambience label and give strong cues as to proximity and location. We know that, to the extent that proximity affords urgency, the temporal and spatial relationship of the direct sound and the ground reflection provide considerable scope for prescribing importance.

What this evidences is a more general result that in order to achieve any kind of depth of perceptual significance at all the context for the audio presentation must be clearly established. The authors have found that the essential elements of this context are coherence and consistency, more important even than complexity or accuracy. It is necessary to present a 'believable' background to the auditory display that reacts consistently to sounding objects within it in order to allow perceptual significance and hence selective attention to foreground objects, or rather it is the coherence over time of the background material that enables selective *in*attention that in turn allows objects within the audio space to be perceived as features with relevance and perhaps importance or even urgency. By presenting a reasonably well-defined background context, the perceptual space is opened-up and makes room for a number of potentially information-yielding audio events that can be individuated and to which selective attention can be applied. Although we have not attempted to define the maximum number of such simultaneous information channels, it is clear that the urgency hierarchy is but one axis and that localisation and other differentiation mechanisms such as spectral content are available such that the number available is potentially quite large. This is in stark contrast to 'dry' single-speaker displays in which the presentation of multiple simultaneous information channels would very quickly become confusing.

There is of course no need to 'fill up' the perceptual space except where the information content is very dense, and there is an argument for allowing as much perceptual space as possible between channels in order to facilitate individuation and delineation of prescribed relative importance. One technique that the authors have found to be highly successful is to make use of the domain of perceptual space that we have characterised as 'adjacent space'. The cartoon characteristics of this are the ambience labelling characteristics of that space, ideally a quite different space acoustically, and projection into the 'proximate zone' ('my space') which is highly directional. Placing audio streams into an adjacent space maps onto the lower-middle tiers of the importance/urgency hierarchy, but of course these could move into 'my space' if relevance was reassigned.

In terms of selecting appropriate spaces and manufacturing context, there are as yet few definite groundrules. It would be common sense to portray a notional space that was simple and thus both perceptually easy to comprehend and simulate, but in which audio did not suffer significance distortion, such as through flutter echoes, strong nodes or long reverb times. Subject to these constraints, spaces could and should themselves be cartoonified, and the authors envisage software 'presets' available according to the depth of perceptual space required and the characteristics both of the information presented and of the perceiver environment in which the presentation occurs**.**

A further example of where cartoonification is in the use of convolution to capture 'the acoustic' of a place.

6.2. 'What' Cartoonification

An entire class of treatments for the cartoonification of 'whats' is also available. Whats can be broken down simply into organisms, objects, and features: the distinguishing characteristics lie in their 'behaviour'. An 'object' is just that which makes a sound (in a given 'context'). An 'organism' is that which makes a sound that betrays sentient behaviour, either vocal or locomotory.

Clearly, dry, close-mic recordings of sounding objects are cartoonified representations of whats, and experience tells us that the selection of one or two sets of signal outputs, with sufficient accuracy of spectral representation, more than satisfies the key criterion of believability. Although this is not an area to which the authors have given much experimental attention, we are aware of the massive potential for conveying information despite considerable spectral compression, for instance, especially where the information conveyed is in semiotic form as in the spoken word.

6.3. 'What-into-Where' Cartoonification

It is important to note, however, that 'things' (the object of 'what' analysis) are not 'point sources' but have very definite shape and size. This has particular implications for our perceptual ability to locate them in any environment, real or artificial.

The cartoonification of what into where, which we have already begun to exemplify above, affords further rich opportunities for differentiation and preselecting according to prescribed importance. These are primarily behavioural parameters, and the importance of these aspects of sounding objects is highlighted by the very strong adaptive bias afforded the apprehension of behaviour.

Comingness is, in our opinion possibly the most urgent class of information in our environment, and we are especially well adapted to detecting and gauging this from quite simply decomposable elements[8]. Treatments for comingness do not rely simply on specifying the temporal and spatial relationships between direct and early reflected sound, and it is possible that generalised treatments for comingness are available that do not rely on change-of-location at all. Change in volume is one, movingness is another, and change in air absorption of high frequencies is another. However, comingness (or 'looming') has a more general purpose in promoting auditory information up the relevance hierarchy. Of course, this would need to be accompanied by concomitant demotion of other elements in order not to overcrowd the available perceptual space (and system headroom).

Movingness is a very clearly apprehensible phenomenon, both in terms of locomotory sounds, and in changes in the timbral quality of non-locomotory sounds: changes in early reflection patterns cause small but coherent 'comb-filter' effects and we suppose quite primitive mechanisms for rapid detection and analysis of these effects, a movement detection system which may well liaise very closely with a 'change of interaural difference (IAD) detection system', in conjunction with the 'thing' analysis system, but is not dependant on coherent IAD information.

Hence movingness is itself a parameter for which good audio cartoons are available. Often changes in subtended angle are used (sometimes alone) to denote movement, but, especially for objects in distant spaces, 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), the authors have found that accurate simulation of object trajectory by closely defining IAD information becomes unnecessary.

Facingness, a significant property of many audible objects, is afforded by the asymmetry of audio dispersion. In an artificial sound field of sufficient competence, the spatiotemporal characteristics of early reflections may be simulated to provide ambience labelling information about the facingness, and change-of-facingness, of *what* in *where*. This facilitates selective attentional processes, with the percipient choosing (and indeed *able* to choose) information streams. Facingness is quite discernible, for many sounding objects, although to what degree of accuracy is presently unclear. Change of facingness is rapidly noticeable, and may offer an early warning of 'intention' of an organism.

To paraphrase, facingness may provide an importance/urgency 'switch', and can be fairly readily cartoonified provided that the sound output is reasonably directional and not made up of too many component parts. Often only two signal samples are required, with appropriate treatments for ambience labelling, and indeed for certain sources such as humans, voice recording or good synthesis together with spectral treatment for body occlusion can suffice.

In my space, position of things is very acutely perceived. IAD information is important, as is IAD for early reflections and a general sense of envelopment engendered by room reverberations (where applicable). Change of position of a sounding thing is generally quickly obvious. Position in an adjacent space is generally less perceptible (though by no means imperceptible), though change of position may be noticeable because of movingness. In distant spaces, position is diffused to a general area (yielding stable IAD information) and movingness is almost certainly the primary source of information about change of position.

6.4. Cartoonification of Events

Whilst 'events' obviously concern 'things' in 'places', there are occasions when things do not move, but change their auditory output. We think the key candidates for cartoonifying are: *startingness*, *ongoingness*, *changingness*, and *finishing*.
Exaggerating the abruptness of starting and stopping clarifies that aspect of the event, without undue use of high volumes. Similarly, suppressing dynamic range and abruptness can help to move something into the 'background' to make space for a new event.
One generalised cartoonification process that the authors have found to be particularly successful is that ambience labelling information is very effectively conveyed using high-frequency information alone. This has permitted the development of hybrid systems where low-frequency capabilities are very efficiently used and only the high frequencies are rendered with spatial accuracy.

However, signal processing for cartoonification is barely developed, and we expect empirical derivation of new treatments to evolve substantially and rapidly, in exactly the same way that the variety of currently-available signal processing techniques has come about.

7. CONCLUSIONS

Clearly, auditory spatial perception is not limited to the classical concept of geometric space. The judgements we usually wish to make are not abstract ones. We have proposed specific classes of information that arise in real environments, and which have potentially measurable physical characteristics. We have termed these classes of information 'ambience labelling information'. We have further proposed the existence of neurological substrates specifically dedicated to processing for these classes of information, according to the concept of perceptual significance. In so doing, we are aware that the model we propose bears some comparison to ‘modular theories’ of perception, such as Jackendoff’s 'computational theory'[9]. In this, we see some convergence of apparently disparate approaches, including behaviourism, computational theory, Gestalt theory, and most especially as elaborated by J.J.Gibson in his 'Ecological Approach'. The resultant 'information-environment' we have termed 'perceptual space', which incorporates classical physical space, but is enriched by the inclusion of supposedly ‘subjective’ considerations. We then propose that the hypothesised neurological processing substrates may be addressed in artificial environment-rendering using simplified representations of the higher-order information-types to which they are adapted, in a process we call 'cartoonification'.

The relevance to techniques for auditory display arises from the opportunities provided for differentiation and pre-selection. We suggest that by treating for perceptual significance according to the psychoacoustic model that we have described allows for about eight discrete categories of relevance. Principally, perceptual significance is denoted by the apparent behavioural characteristics, or at least cartoonified abstractions of these that lock into preconscious processes for various calls to action. We have also described how differentiation is not primarily related to angular separation, which is poorly perceived and reproduced, but heavily dependent on a description, even a very simplistic one, of the interaction of the sound source with its environment. Crucially, this demonstrates that relatively much greater attention needs to be paid to the space, perceptual and physical, in which sounding objects find themselves, in order to achieve true depth of significance.

We anticipate that the schemata here outlined can be quantified and realised separately and collectively, in hardware/software combinations resulting in a globally manipulatable 'audio information-environment' that intuitively facilitates selective attention.

8. REFERENCES

[1] L. C. Larijani, *The Virtual Reality Primer.* McGraw-Hill. 1993.

[2] P. P. Lennox, T. Myatt and J. M. Vaughan, "From Surround To True 3-D" in *Proceedings of the AES 16th International Conference on Spatial Sound Reproduction*, Rovaniemi, Finland, 10-12 April 1999, pp.126-135.

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

[4] V. S. Ramachandran, and S. Blakeslee, *Phantoms In The Brain: Human Nature and the Architecture of the Mind.* Fourth Estate, London, 1998.

[5] J. Atkinson, *A Neurobiological Approach to the Development of 'Where' and 'What' Systems for Spatial Representation in Human Infants.* In: N. Eilan, R. McCarthy, and B. Brewer (eds.), *Spatial Representation. Problems in Philosophy and Psychology.* Blackwell, Oxford, pp.325-339. 1993.

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

[7] P. P. Lennox, J. M. Vaughan and T. Myatt, "3D Audio as an Information Environment." To be presented at the *AES 19th International Conference on Surround Sound Techniques, Technology and Perception,* Schloss Elmau, Germany, 21-24 June 2001.

[8] J. G. Neuhoff, "An Adaptive Bias in the Perception of Looming Auditory Motion." Unpubl. Manuscript. 2001.

[9] R. S. Jackendoff, *Consciousness and the Computational Mind.* MIT Press, Cambridge, Mass. 1987.