View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by Scholarly Materials And Research @ Georgia Tech

CONCEPTS OF PERCEPTUAL SIGNIFICANCE FOR COMPOSITION AND REPRODUCTION OF EXPLORABLE SURROUND SOUND FIELDS.

Peter Lennox[1] and Tony Myatt[2]

1 Signal Processing Applications Research Group, University of Derby, Derby, UK
p.lennox@derby.ac.uk
2 Music Research Centre, Department of Music, University of York, York, UK
am12@york.ac.uk

ABSTRACT

Recent work in audio and visual perception suggests that, over and above sensory acuities, exploration of an environment is a most powerful perceptual strategy. For some uses, the plausibility of artificial sound environments might be dramatically improved if exploratory perception is accommodated.

The composition and reproduction of spatially explorable sound fields involves a different set of problems from the conventional surround sound paradigm, developed to display music and sound effects to an essentially passive audience.

This paper is based upon contemporary models of perception and presents proposals for additional spatial characteristics beyond classical concepts of three-dimensional positioning of virtual objects.

1. INTRODUCTION

Many authors (e.g. [1]) have proposed that adding parameters of spatiality to artificial sound can improve information transfer, allowing the percipient to choose the objects of attention in conditions of multiple concurrent auditory streams. Especially where multiple sources are essentially similar (same-sex voices, for instance) spatially separating them can improve listener performance [2]. Most simply, this is achieved by separating sources angularly. In a more sophisticated way, items could be separated by angle and range, the latter depicted by manipulating a few parameters such as amplitude, direct/indirect sound ratio and equalization. However, "spatiality" (i.e. that which appeals to human spatial perception) is subtler than this, so there are many more potential parameters that could be used to broaden the information bandwidth in artificial spatial sound.

A notable distinction between real-world and artificial auditory environments lies with the types of information transaction they afford. In the former, perceivers actually explore the environment to extract more detailed information from features of particular significance. By contrast, artificial (auditory) environments are 'informationally shallow' as they do not cater to perceptual exploration strategies. Phantom images (illusions of sounding objects between speakers) are notoriously unstable to a moving perceiver (in terms of location and size, or "apparent source width" [3]), resulting in a perceived loss of realism. Range (distance from image to perceiver) is similarly exposed as illusory when one moves. Exploration reveals more about the audio system and the unreality of the audio content.

Although a motionless percipient should be thought of as a *special case*, surround sound applications tend to de-emphasise the ambulant capabilities of perceivers and are predicated on assumptions of a static (and passive) listener. This may be quite appropriate for music listening since contemporary music must be predicated on the technical limitations of available display systems.

Users of personal computers are usually immobile, normally seated. Domestic surround sound must be heard from a quite specific position and orientation for the spatial illusion to appear as intended [1][5][6][7].

In artificial environments of these sorts, the opportunities for physical, ambulant exploration are sharply circumscribed, are presented to the percipient in simplified form which is already partially sorted, so that the composer of the environment governs listeners' attention. This intrinsically emphasises the reception element of perception.

The presentation to a static listener of binaural signals conveying an impression of spatiality is therefore a restricted kind of spatiality, which does not afford the full range of cognitive spatial abilities that might be deployed in real environments.

2. EXPLORATION: INFORMATION SORTING IN REAL ENVIRONMENTS

Real environments are incredibly complex and the appropriate sets of *meanings* (including spatial relationships) must be extracted economically and accurately, and in timely fashion; vast quantities of sense-data must be sorted.

This sorting can be discussed in terms of cognitive and behavioural exploration. Cognitive exploration encompasses attention /inattention [8] and cognitive sorting mechanisms are exemplified in Auditory Scene Analysis [9].

Rather than assume that a percipient's perceptual processes must involve the accurate recording and subsequent analysis of all available sense data prior to understanding the key environmental meanings, exploration implies the active pursuit of meanings.

Indeed, many commentators have stressed the need to consider human perception in a wider paradigm than laboratory circumstances, wherein control of extraneous factors is gained at the expense of generalisability of the conclusions drawn. Jarvilehto [10] considers that environment and percipient comprise a holistic system and that sensible conclusions about perception cannot be drawn from the study of perceivers in isolation. James Gibson stressed that vision can be considered at the scale of environmental interaction:

> *We are told that vision depends on the eye, which is connected to the brain. I shall suggest that natural vision depends on the eyes in the head on a body supported by the ground, the brain being only the central organ of a complete visual system. When no constraints are put on the visual system, we look around, walk up to something interesting and move around it so as to see it from all sides, and go from one vista to another. That is natural vision…* [11]

Here, Gibson also stresses the *ambulant* nature of visual perception, which touches on what is meant here by exploration.

Behavioural exploration in the form of head-turning and locomotion, are examples of the percipient choosing *to what* sensory apparatus should be exposed.

Active interrogation of the environment is achieved by interfering with what is to be explored, by prodding, picking up and moving things; the environment is stimulated into yielding information.

2.1. Cognitive mapping and reference frames

Exploratory behaviour that goes beyond simple random activity requires hypotheses about what *might* occur next. This implies ongoing, high-level representation of the *perceptually significant* [12] features of the local environment, including cognizance of spatial relationships. This has been characterized as *cognitive mapping* [13], and the concept can be used in conjunction with that of *cognitive reference frames* (e.g. [14]).
We have stated [15]:

> *A spatial frame of reference is a set of spatial relations between features, or landmarks, that can serve as background context for spatial action, perception and conception.*
> *The frame defines a place within which can be defined positions, or locations. Frames of reference provide contexts in which spatial knowledge is organised. A perceiver might simultaneously use several frames of reference in a situation and use multiple frames sequentially. Different situations might require different combinations of reference frames.*

Reference frames can be coarsely classified as *egocentric* (perceiver-centred) and *exterocentric* (non perceiver-centred). In the former, the position of items is referenced from the perceiver, whereas in the latter, external landmarks can be used as reference, so that an environment can have *shape* irrespective of the position of the particular perceiver. This second type of frame, that supports perception of an externalized world, is vital for perceiver mobility. For a fuller discussion, see Campbell [16].

2.2. Intuitive Physics

Some authors [21] find evidential support for well-developed *intuitive physics* in pre-verbal infants, which amounts to the existence of *a priori* assumptions about the physics of events, objects and features. Such assumptions are imposed on the incoming sense-data to facilitate rapid sorting of the voluble flow of information that sensation makes available. However, this does not imply that the distinctions made in *intuitive physics* are arbitrary matters of opinion; they can rest on physical distinctions *in* the world that are available to sensation.

The implication is that phylogenetic and ontogenetic developments shape the knowledge structures with which we decode the meanings in our everyday environments. In this view, there is no initial *objective* rendition of space at the sites of sensation, since those knowledge structures must be meaningfully connected with items in the external environment.
Notably (in the present context) items of perceptual interest *rarely* appeal to just one sense-mode, other than momentarily. Although the distinct nature of sense modes is intuitively obvious, that of *perceptual modes* is less so. In real world situations, the senses act in concert, and the concomitant *perception* is of a unified, multi-modal world.
This has two implications:

1. That sense modes are unlikely to be truly equivalent (and hence redundant) in some modes of perception; rather they are specialized to extract more robust information from the same scene
2. They are unlikely to be insulated from each other; therefore they can probably contribute to, *and draw from* the same cognitive spatial representations.

2.3. An ecological approach to cognitive categories in spatial perception.

This notion is that spatial representation is non-unitary, deploying multiple concurrent neural representations of events, objects, features and relationships in order to address different aspects of necessary tasks.

Some representations can be rather similar to classical notions of three-dimensional space, replete with *landmark features*, and these can economically represent the static, *background* elements of the environment. Others can specialize in *movement* of discrete entities such as *objects* and *organisms*, and more abstractly, *behaviours* and *intentions* (although more *physically* abstract, they are *perceptually significant* items that perception has evolved to understand). Movement (of entities in the vicinity) is clearly potentially urgent; one would expect well-developed systems that devote considerable cognitive resources to movement problems. Since some types of movement are likely to be more urgent than others, one would further assume a 'resource hierarchy' that reflects this (we have previously used the term perceptual significance [15] to denote this). In other words, there is little reason to assume that all spatial attributes in a given environment should be treated even-handedly in cognition.

2.4. Movement

In common sense terms, a human sized organism approaching rapidly (or even accelerating) will probably command full attention at the expense of other items in the vicinity. Therefore, movement can be categorised according to its urgency, as: coming, going, passing, fast, slow, accelerating, changing direction and so on.

2.5. Location

An analogous situation exists for location. Beyond the positioning along X, Y and Z co-ordinates (centred, say, on the perceiver), there are distinctions of *near* and *far*. But *perceptual* and *physical* "near" may differ. For example, a physically-near item in an adjacent room is perceptually further than items in *this* room. An item that can move fast is perceptually closer than a slow item. In this instance, we could say that such an item is *causally* near.

Location is further complicated because it is not simply a matter of direction and distance; the relationships between a sounding object and its surroundings are clearly audible. One can hear when an object is near a wall, passes behind an occluding feature or moves past successive reflective and absorptive features. For objects having anisotropic output (because of body-occlusion) one can hear that item's orientation with respect to features and oneself (what we have termed 'facingness' [15]) The overall term we have used to encapsulate the audible consequences of sounding objects' physical relationships with environment features and the listener is ambience labeling [15].

The matter is similar for background environmental features: a very wide, gently curving road can seem very narrow and twisty when one is traveling at high speed. An escape route such as a doorway is very far away if there are intervening obstacles and threats.

2.6. Perceptual categories: cognitive cartoons

The principle underlying the categories of perception (in real environments) is the *potential for interaction*, which is as physically real as an object, a process, a position or a classical spatial relationship.

It has been proposed that there are dedicated neural sub-systems that specialize in understanding 'things' (the *what* systems) and others for understanding the location of those things (the *where* systems) [16] These categories maybe too simplistic. Some things (such as ambulant organisms) clearly challenge a perceiver in a qualitatively differently way than do inanimate objects. Nevertheless, evidence for specialized systems is substantial; and the question at hand is one of identifying what categories these sub-systems actually specialize in, in a more detailed way.

It is impossible for our perception to process all available information in a given environment. Perception achieves the remarkable feat of selecting appropriate information for before examining that information closely.

Our proposition is that potential interactions are cognitively processed in a "cartoon" form: pared-down representations that capture the salient features quickly enough to facilitate real-time interaction.

3. EXPLORABLE SOUND FIELDS

We can assume that the taxonomy of spatial attributes in an artificial sound field differs from that in real environments (artificial environments should not be intrinsically dangerous, for example). Still, many of the perceptual systems that have evolved in real world examples might also be available (albeit in modified form) in artificial fields. This may be useful in engineering environment simulations, for training, investigating human behaviour, interacting in complex ways with real environments, or simply for entertainment.

Although metrically precise *images* are particular virtues of *visual* perception, the exploration of artificial spatial sound can be conducted through an investigation of the *special* virtues of audition, with respect to the appreciation of space.

Visually occluded objects (behind the perceiver or behind another item) yield audio information about the rate and change of movement, and even the *reason* for movement. These environmental conditions can signify some call to action.

More importantly and unlike vision, this kind of information is available in parallel with location information and object-type information. Counter-intuitively, one can detect movement prior to the detection of position. This is more obvious when one reflects that a monophonic recording can convey movement (such as *coming, passing* and *departing*) even when precise location in terms of the egocentric appreciation of direction (via interaural differences) is unavailable. Clearly, there is physical information concerning the spatial relationships between source and proximate features that does not rely upon direction-perception.

3.1. Degrees of explorability and cartoonification in artificial sound fields

It is philosophically a moot point whether it will ever be physically possible to produce an artificial environment that is perceptually indistinguishable from a real one. It certainly is not currently possible. The effectiveness of an artificial environment rests on an appropriate simplification of elements, with respect to the task in hand. Artificial auditory cartoons [17] should reciprocate relevant cognitive cartoons so that they are 'intuitively graspable' [18] to facilitate information reduction in complex interactive artificial environments. Logically, one can *signify* a spatial attribute rather than fully specify it.

This is, of course, precisely what happens in computer games, which are pared-down environment simulators. As the percipient explores a situation virtually, passing by illusory sources, the effect is simulated by simplistic shaping of audio amplitude, with perhaps some rudimentary panning.

Different levels of explorability may be acceptable in different presentation environments, headphone listening, concert hall presentation and ambulant presentation situations.

Sound fields that are explorable should:

- contain perceptually significant sound events
- provide more information to the listener through exploration/movement than that provided to passive/static listeners
- provide audio information relating to the environment that surrounds the listener and the items in the illusion
- provide information about the relationships between sounding objects
- contain information relating to the interaction of objects and environments
- consistently display objects, features and relationships in a stable, non-egocentric reference frame

3.2. Plausibility controls

Tools are required which manage perceptual significance in artificial sound fields. The immediate objective should be plausibility rather than absolute physical accuracy. The principle of cartoonification is useful here. An example of an existing cartoon treatment is to be found in the nature of sound fields themselves. A sound field is a cartoon of the audible part of an environment (for a discussion of the differences between a sound field and sound environment, see: [19] Cartoons can be nested, one inside the other, so that within a sound field, individual elements can be cartoonified. Movement can be signified by manipulating inter aural differences and pinnae effects in a conventional manner. A facsimile of an object moving through an actual environment can be added to this. A crude implementation, e.g using phasing, can imbue a virtual object with some of the characteristics of movement. Several simulated early reflections, dynamically panned (and with ongoing amplitude fluctuations to simulate environmental irregularities) can be added. Some control of simulated orientation of objects might also be desirable (what we have previously referred to as *facing-ness*, since many real objects turn to 'face' their direction of travel). If change of range is to be simulated (bearing in mind that real objects circling a perceiver at constant range would be unusual), then controls for *coming*, *passing* and *departing* need to be available.

Controlling all these parameters coherently is currently problematic, though we argue that the nature of the problem lies in the conceptualization of interface structures rather than the underlying processes. Many of the individual signal processing techniques are available. The principle of cartoonification might help here, too. It might be that fast moving objects can be treated less subtly than slow ones. Strictly speaking, moving objects do not have location, they move through locations. The perceptual understanding of movement trajectories cannot consist of a detailed analysis of location if a fast moving object has departed from the location before an analysis can be completed; it may not be appropriate to try to *accurately* to render location at all.

Finally, it must be observed that simulating movement of the perceiver *through* an artificial environment presents a different set of problems. Whilst many of the cartoon treatments may be similar (panning and change of range for instance) their coherent combination must differ from the movement of sources for a static listener.

4. IMPLICATIONS

It is theoretically possible to manage plausibility by simulating the perceiver-environment relationships found in real situations. This is done by catering to cognitive exploration and ambulant exploration so that interaction with the environment extracts increased information beyond the static, passive case. However, engineering the binaural sound field for a single position will be insufficient, since ambulant exploration exposes the percipient to successive different sound fields. This exposes shortcomings in the single sound field approach, since the spatial cues for an ambulant listener may differ substantially. The auditory equivalent of parallax, changes in relative amplitude during movement, and different transformations of the ratio and angle of incidence for direct and reflected portions of audio signals may all contribute to the robust perception of an externalised environment. Managing these with conventional sound field methods appears impracticable, and an alternative approach is to model individual sources, physically, and environmental features as components, in a fashion conceptually similar to Wave Field Synthesis [23].

Explorability has significant implications for the design and configuration of sound reproduction systems and also the design of creative tools for the spatial sound production.

4.1. Extent and scale of explorability

For an artificial environment that does not have a 'centre', built from components which take the form of small, local fields, there are questions of the overall dimensions in which spatial exploration can take place (and what happens at the outer boundaries of the sound field) and the range of fine detail available – e.g. how minutely it can be explored. These are potentially challenging.

The realisation of explorable sound fields is likely to utilise multiple sound reproduction techniques and speaker configurations. Spatialisation tools for creative applications will be required to enact this, potentially without requiring users to engage with the mechanisms of sound reproduction. This suggests an interface protocol and sound representation system capable of sophisticated communication with spalialisation "engines", where the mechanism of spatial reproduction is implemented according to the perceptual significance of the content and the chosen reproduction techniques/loudspeaker arrays.

5. CONCLUSIONS

Perceptually explorable artificial sound fields will require significant changes in approach to those currently used for the creation and implementation of spatial sound fields, beyond presenting the impression of three-dimensionality to a single listener position. Radical changes are necessary with respect to both content generation and conception of sound designs, and in engineering techniques to implement explorable fields. An incidental benefit lies in the creative possibilities. For instance, a composer could create a piece with no 'listening centre' that can only be fully appreciated by ambulant listeners. We are currently exploring the aesthetic potential of this approach.

6. REFERENCES

[1] Begault, D. "3-D Sound for Virtual Reality and Multimedia" retrieved Jan 2006 from http://human-factors.arc.nasa.gov/ihh/spatial/papers/pdfs_db/Begault_2000_3d_Sound_Multimedia.pdf

[2] Garadat, S and Litovsky, R. "Speech intelligibility in free field: Spatial unmasking in preschool children" retrieved Jan 2006from: http://www.waisman.wisc.edu/~litovsky/publications/garadat_litovsky_2007_jasa.pdf

[3] Potard G., and Burnett I., (2003) "Study on sound source apparent shape and wideness." In: *Proceedings of the 2003*

International Conference on Auditory Display, Boston, MA, USA, 6-9 July 2003

[4] Gerzon, M. A. “What’s wrong with Quadraphonics”. Originally published in Studio Sound, (May 1974) Retrieved: July, 2004 from http://www.audiosignal.co.uk/What's%20wrong%20with%20quadraphonics.html 1974 .

[5] Theile, G. “Multi-channel Natural Music Recording Based on Psycho-acoustic Principles”. Extended version of the paper presented at the AES 19th International Conference. Schloss Elmau, Germany. Retrieved: May, 2003, from http://www.irt.de/IRT/FuE/as/multi-mr-ext.pdf. 2001.

[6] Glasgal, R. (1999) "The Blumlein Conspiracy". Retrieved: August 2000, from http://www.ambiophonics.org/blumlein_conspiracy.htm.

[7] Blumlein, A. "Improvements in and relating to Sound-transmission, Sound-recording and Sound-reproducing Systems", British Patent Application 394325. (1939) Retrieved: Aug, 2004, from http://www.doramusic.com/patents/394325.htm .

[8] Lennox, P.P., Myatt, A., Vaughan, J., “3D Audio as an Information Environment” in *Proceedings of Audio Engineering Society 19th International Conference on Surround Sound Techniques, Technology and Perception*, Schloss Elmau, Germany, 2001.

[9] Bregman, Albert S., *Auditory Scene Analysis: The Perceptual Organization of sound*. Cambridge, Massachusetts: The MIT Press. 1990

[10] Jarvilehto T. “The theory of the organism-environment system: I. Description of the theory”. First pub. 1998 in *Integrative Physiological and Behavioral Science*, 33, 321-334. Retrieved August 2004 from: http://wwwedu.oulu.fi/homepage/tjarvile/orgenv1.pdf 1998.

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

[12] Lennox, P.P., Myatt, A., Vaughan, J., "3d Audio As An Information-Environment: Manipulating Perceptual Significance For Differentiation And Pre-Selection", ICAD *Proceedings of the 7th International Conference on Auditory Display*, 2001. 2001.

[13] Tolman, Edward C.. “Cognitive Maps in Rats and Men” First published in *The Psychological Review*, 55(4), 189-208. Online: Classics in the History of Psychology retrieved Aug. 2004, from: http://psychclassics.yorku.ca/Tolman/Maps/maps.htm 1948

[14] Eilan, N., McCarthy, R.A. & Brewer, W. (Eds.) *Spatial Representation: Problems in Philosophy and Psychology*. Oxford. Blackwell. 1993.

[15] Lennox, P.P. "Spatial Music and Spatial Perception in Artificial Environments" PhD Thesis, University of York, UK. 2004

[16] Atkinson, J., (1993) “A neurobiological approach to the development of ‘where’ and ‘what’ systems for spatial representation in human infants” in (Eds.) Eilan, N., McCarthy, R., and Brewer, B. Spatial Representation Oxford. Oxford University Press.

[17] Gaver W.W. (1993a) “What in the world do we hear? An ecological approach to auditory source perception” in: *Ecological Psychology* 5, pp. 1-29.

[18] Kahneman, D. Tversky, Frederick, Krantz, (2002) “Maps of Bounded Rationality” Nobel Prize Lecture December 8, 2002

[19] Lennox *et al* (2000) opinions of “sound fields” in archives of the Auditory discussion list last retrieved Jan 2006 from: http://www.auditory.org/mhonarc/2000/msg00505.html

[20] Campbell, J. “The role of physical objects in spatial thinking” in: *Spatial Representation* Eds., Eilan, N., McCarthy, R., and Brewer, B. Oxford. Oxford University Press. 1993.

[21] Spelke, E.S., & Van de Walle, G.A. "Perceiving and reasoning about objects: Insights from infants". In M. Eilan, R. McCarthy, & B. Brewer (Eds.), *Spatial Representation* (pp. 132-161). Oxford: Blackwell. 1993.

[22] Lennox, P. P. Myatt T. and Vaughan, J. M. "From Surround To True 3-D" in Proc.of Audio Engineering Society *Proc. Of 16th International Conference on Spatial Sound Reproduction*, Rovaniemi, Finland, 1999.

[23] Berkhout, A.J. Ridder, J. van der Wal, L.F. "Acoustic Imaging by Wavefield Extrapolation", in: *Acoustical Imaging*, vol. 10. New York. Plenum Press. 1982.